Neurocomputing 000 (2017) 1–9

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Method of predicting human mobility patterns using deep learning

Dong Yup Kim, Ha Yoon Song*

Department of Computer Engineering, Hongik University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 29 January 2017
Revised 2 July 2017
Accepted 31 July 2017
Available online xxx

Keywords:
Human mobility pattern
Mobility pattern prediction
Deep neural network
Deep belief network
Positioning data
Location data

ABSTRACT

As human location and movement data are becoming easily accessible, owing to the prevalence of mobile devices, the use of such mobility data is gaining an increasing amount of interest. In our work, we establish a relationship between human mobility and personality, and attempt to model and predict movement patterns. Deep-neural-network and deep-belief-network models of deep learning are used in conjunction, for training the neural network. Both passive positioning information and active location information are used for the mobility information dataset, and the big five factors are used for the personality data. Both mobility information and personality information are split into training and verification groups, and are subsequently used to train and verify the neural network. The results are expressed in terms of hit ratios according to model factors comparing the predicted and observed values, and the parameters for the neural networks for the highest accuracy are identified. We use this optimal neural network to show the correlation between human personality and mobility patterns. Actual prediction is attempted, and is found meaningful in some conditions.

1. Introduction

Recently, artificial intelligence (AI) is becoming one of the viable solutions for future sophisticated services. Deep learning, in particular, is one of the worthy techniques, as demonstrated by Alpha Go [1]. The applications of deep learning are now widely available, owing to the improvements in computer performance and the discovery of new algorithms.

In this paper, we will utilize deep learning to reveal the relationship between a humans mobility information and personality information. Human mobility analysis has been conducted in various fields, and the results can be applied to industry and services. Once the relationship between personality and human mobility is revealed, it will be one of the bases for human mobility analysis and for the prediction of human movement patterns. One of the preceding studies shows the relationship, through regression analysis using individually collected mobility information and the big five factors (BFF), which are collected through polls [2]. Instead of using regression, deep learning in this study will be one of the methods used, as the amount of data collected is larger. Among the various deep learning models, the deep belief network (DBN) and deep neural network (DNN) will be utilized to identify the correlation between human mobility and human personality.

The collected mobility information will be used to construct artificial networks. The neural network will be trained using the personality data and mobility information, and then, it will be verified. Using DBN, the neural network will be created and the initial weight values will be decided, since it is possible to pretrain a DBN. In the case of DNN, which is capable of classification, the neural network will be trained once more, and then, the results will be acquired according to the inputs. The optimal parameters for neural networks will be detected from the results. The optimal neural network can then be used so that the effect of personality on mobility patterns and the correlations between personality and mobility patterns can be identified using synthetic personality data.

The remainder of this paper is organized as follows. Section 2 introduces the related works. Section 3 discusses the data collection and processing methodologies; two types of datasets are used: location dataset and personality dataset. In Section 4, the DNN and DBN used for the training and verification of data are discussed. Section 5 shows the experimental methods, and Section 6 presents the results from the trained neural network. Finally, Section 7 mentions the conclusions and the future research directions.

2. Related works

The two major backgrounds for our research are deep learning as a methodology, and human mobility and personality as data of major interest.

* Corresponding author.
E-mail address: hayoon@hongik.ac.kr (H.Y. Song).

https://doi.org/10.1016/j.neucom.2017.07.069

2.1. Human mobility and human personality

Human mobility has been analyzed in various fields. A study was previously conducted to derive the mobility model for each individual from location information [3]. The individual mobility model for a person can be built as follows. With positioning devices such as GPS receivers and smartphones, the raw positioning data of human mobile traces can be obtained. The positioning data can be obtained as frequently as once in every second. The positioning data contain the latitude, longitude, and time of collection. Each positioning datum comprises a set of positioning data sequences. The positioning datasets for each person can be transformed into individual mobility models. Considering the spatiotemporal density of positioning data, location clusters can be established, which have concentrated positioning data. In other words, the positioning data within a specified time range and distance range can be aggregated and formed into a location cluster. For actual implementation, the location clusters can be found using a clustering method called expectation maximization [4]. Then, the transition probability between location clusters can be calculated. From the concept of Markov chain, the location clusters can be mapped onto states of the Markov chain, and the transition probability between location clusters can be mapped onto the transition probability between the states of the Markov chain. Then, the individual mobility model can be seamlessly represented in the form of a Markov chain. With this procedure, location clusters and transitions between location clusters can be formally integrated and represented using the Markov chain. The generated Markov chain is one representation of the human mobility model.

Based on this study, the group mobility model can also be derived from the raw positioning data [5]. Group mobility models can also be extracted, and represented in the form of a Markov chain.

From the aspect of psychology, it has been explained that human mobility patterns can be derived from psychological factors. In other words, human psychological factors can affect the human's destination, route, and so on [6]. Individual personalities can influence the human mobility models, and human habits can also influence the human mobility ways [7,8].

One of the very first studies analyzed each human movement and psychological characteristic, by utilizing GPS collecting devices [9]. Another study used regression analysis between location information and human personality [10]. However, the direct application of regression analysis to a raw dataset could contain effects from the skewness of the dataset collected by volunteers. Another study used the back propagation network to minimize the effect of data skewness [2]. Apart from the above studies, our research utilizes two types of data for human mobility patterns: mobility patterns established from the raw positioning dataset, and those established from a volunteer's actually visited location information, which is collected actively by the volunteer. In addition, the BFF values for each volunteer will be utilized to fulfill our purpose.

2.2. Neural network

Deep learning can be defined as an artificial neural network with a number of hidden layers. The deep learning model can provide solutions for complex problems such as pattern cognition, risk factor detection, and so on. Studies regarding deep learning have been steadily continuing from the past, and have become noticeable because some of the limitations were solved, for example, the overfitting problem, the computing power required to train a number of hidden layers, etc. [11,12].

The well-known deep learning models are the DNN, convolutional neural networks (CNN), recurrent neural networks (RNN), and DBN. The DNN is the base model for other neural networks, and consists of an input layer, output layer, and hidden layer [13]. The DNN can model complex nonlinear relations, for example, the DBN is for object identification modeling [14]. The advantage of the DNN is the ability to model complex data with a relatively small number of units [15]. The CNN has a structure with one or more convolutional layers, in addition to an upper artificial neural network, and it shows good performance in video and voice application fields. Computer vision and acoustic modeling fields have also shown improved performances with the application of CNN [16,17]. RNN is a neural network with bidirectional links for each unit. Unlike other neural networks, memories inside this neural network can be utilized, and it is effective in the language modeling and handwriting recognition fields, providing high recognition rates [18,19]. The DBN has links between each layer, but no links between units in the same layer. It has the characteristic that the neural network is constructed from the bottom layer. Therefore, pretraining is enabled, and it is possible to reset the weight of each hidden layer effectively using a small number of training data [20].

We will use the DNN and DBN in this study. The details of the neural networks used in our actual experiments will be discussed in Section 4.

3. Location and personality information

Two types of data must be prepared for the experiment: position information and location information and personality information. The positioning data form a passively collected dataset, and in this dataset, we will focus on the transition patterns between locations. The term “passive” indicates that volunteer participants do not care about data collection after the initiation of the positioning device. The positioning device will collect the positioning data without any manipulation from the volunteers. The location data form an actively collected dataset, meaning that the participants actively check in at certain locations; we focus on the location preferences of persons in this dataset.

3.1. Mobility information

For the mobility information, two types of data are prepared. The first is the positioning data, which are collected by the participants' positioning devices. Participants carry GPS receivers or smartphones with positioning apps, and without any active action, positioning data are collected passively. For example, data can be collected by using mobile applications such as Sports Tracker [21] or GPS receivers such as Garmin Edge 810 [22]. From the collected positioning data, the core part of $<$ latitude, longitude, time $>$ are extracted and used. Table 1 shows the number of positioning information, the duration of collection, and the number of location clusters processed from the raw positioning data collected by the participants. Seventeen participants provided their positioning data. From the raw positioning dataset, the location clusters can be extracted as shown in [3].

From the location clusters, the transition probabilities between clusters can be extracted. Among the various location clusters, we focus on three location clusters: Home (H), Work (W), and etc. (E). Therefore, nine types of transitions can be made, and the transition probability can be expressed in a 3×3 matrix, as shown in Eq. (1).

$$\begin{array}{cc} & \begin{array}{ccc} H & W & E \end{array} \\ \begin{array}{c} H \\ W \\ E \end{array} & \begin{pmatrix} y_{1,1} & y_{1,2} & y_{1,3} \\ y_{2,1} & y_{2,2} & y_{2,3} \\ y_{3,1} & y_{3,2} & y_{3,3} \end{pmatrix} \end{array} \quad (1)$$

Then, the elements of the matrix can be calculated using the time progression as follows. Once the transition is made from one loca-

Please cite this article as: D.Y. Kim, H.Y. Song, Method of predicting human mobility patterns using deep learning, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.07.069

Table 1
Details of positioning data collection for positioning data analysis.

Participants	Number of positions	Collecting period(y/m/d) From	To	Number of location clusters
Park	153,683	2015/03/02	2015/06/20	4
bigjam	124,135	2015/03/10	2015/06/19	6
bjh	3,747,296	2014/01/16	2015/07/10	70
bobtong	269,128	2015/03/09	2015/06/03	6
bongtabgE	51,804	2015/03/14	2015/04/22	8
byeolmyeong	297,270	2015/03/12	2015/06/19	5
cdy	810,680	2013/05/15	2016/03/10	38
cuda	1,247,892	2011/09/23	2015/05/23	16
jsh	224,315	2014/09/02	2015/07/16	57
judong	7,063	2015/03/04	2015/04/07	2
kdy	3,034,065	2013/05/20	2016/03/11	123
ksy	234,863	2012/09/21	2015/05/31	41
leb	439,470	2013/06/18	2015/05/04	46
ljs	2,507,542	2013/05/31	2015/07/18	147
samhwa	106,240	2015/03/05	2015/06/21	9
shy	1,349,325	2012/10/11	2015/07/12	78
ydb	291,412	2014/09/12	2015/11/23	36

Table 2
Number of check-ins by each participant for location data analysis.

Participants	asy	bigjam	bjh	bobtong	bongtangE
number of check-in	464	25	772	205	128
Participants	byeolmyeong	cdy	cuda	ihn	jsh
Number of check-in	19	1408	135	1121	285
Participants	judong	kdi	kdy	khi	kjw
Number of check-in	27	527	1171	165	474
Participants	knh	ksy	leb	lhn	ljs
Number of check-in	81	285	881	95	1533
Participants	lkh	lyr	park	pebble	psy
Number of check-in	242	210	121	2084	198
Participants	rhm	sdi	shy	sol	ydb
Number of check-in	237	132	6506	2043	1308
Participants	yik				
Number of check-in	205				

tion i to another location j,

$$C_{i,j} = C_{i,j} + 1, \quad i \neq j \tag{2}$$

where $C_{i,j}$ stands for the count of transitions from i to j; i and j correspond to locations, one among H, W, and E; and $t(i)$ stands for the time at location i.

Otherwise, no transition is decided when enough time is spent at a certain location i:

$$C_{i,i} = C_{i,i} + 1, \quad i = j \text{ and } t(j) - t(i) > \text{threshold} \tag{3}$$

where $C_{i,i}$ stands for the count of stays at i.

The probability value can be expressed as

$$y_{i,j} = \frac{C_{i,j}}{\sum_i \sum_j C_{i,j}} \tag{4}$$

where $y_{i,j}$ stands for the transition probability from $label_i$ to $label_j$.

The vector $< y_{1,1}, y_{1,2}, y_{1,3}, y_{2,1}, y_{2,2}, y_{2,3}, y_{3,1}, y_{3,2}, y_{3,3} >$ is transformed from matrix ($y_{i,j}$), and it will be input to the output layer of the DNN for training.

The second type of data related to the location information form the location dataset. The location dataset is also collected by the participants. The participants carry smartphones and actively check-in at their favorite places using an app called “Swarm” [23]. The location data have parts such as location name, category, number of check-ins, and so on. Thirty-one participants provided their location data. Table 2 shows the number of location data collected by each participant.

For experiments, the location categories are classified as home, work, restaurant, etc., and can be calculated with probability values as shown in Eq. (5):

$$y_i = \frac{L_i}{\sum_i L_i} \tag{5}$$

where L_i is the count and i is one among home, work, restaurant, etc. y_i is calculated for each participant. The vector $< y_1, y_2, y_3, y_4 >$ will be input to the output layer of the DNN for training purposes.

Table 3
BFF values of 17 participants for positioning dataset analysis.

Participant's nickname	O	C	E	A	N
Park	3.400	3.556	3.375	2.778	2.875
bigjam	3.100	3.556	3.500	2.778	2.500
bjh	2.700	3.222	3.250	2.667	2.750
bobtong	2.000	2.889	3.875	2.778	2.875
bongtabgE	3.000	3.333	3.250	2.556	2.875
byeolmyeong	3.800	3.556	2.750	3.111	3.375
cdy	4.000	3.667	4.000	3.889	2.750
cuda	3.400	3.222	3.625	3.222	3.250
jsh	3.400	3.556	3.625	2.889	2.500
judong	3.100	3.444	3.000	3.333	2.875
kdy	4.200	4.333	3.500	3.556	2.625
ksy	3.500	3.000	2.125	3.556	3.000
leb	3.300	3.889	3.250	3.667	2.625
ljs	3.600	3.333	2.750	3.222	2.750
samhwa	3.300	3.556	3.125	2.778	3.125
shy	4.300	4.111	3.500	3.111	2.625
ydb	3.500	3.778	3.375	3.222	3.000

3.2. Personality information

BFF is the most suited for our purpose to represent human personality [24–26]. BFF is composed of five factors: openness (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N). Each factor has a value between 0 and 5, and is positively proportional to the corresponding personality. This numerical nature of BFF enables the value of each factor to be directly applied to the DNN. Tables 3 and 4 show the BFF of each participant, separately, for the positioning dataset and location dataset, respectively.

4. Deep learning model

We will use the DNN and DBN in this research. The DBN pretrains the neural network with the initial weight value, i.e., it constructs the neural network. The DNN is used for classification. It requires a training stage and utilizes the neural network constructed by the DBN. These two types of artificial neural networks will be used to obtain the mobility information as the result, using the inputs of BFF of each volunteer, and to train the weights for each layer.

The DBN can be generated by layering multiple restricted Boltzmann machines (RBM) over one input layer [27]. An RBM is constructed of one visible layer and a hidden layer. Each layer is constructed of units (nodes). Unlike the Boltzmann machine (BM), the RBM has no link between the units within a layer; it has links to all units in other layers, and each link has its own weight. Owing to this characteristic, the RBM requires a shorter training time compared to the BM, as the number of hidden layers is more. In order to train the RBM efficiently, a method called contrastive divergence (CD) is used [28]. The training of the RBM progresses in a greedy-layer-wise manner [29]. The greedy-layer-wise training has two steps: pretraining and fine tuning. In the pretraining step, the inputs to the visible layer of the RBM will adjust the weights of the units in each layer, and the results of an activation function will be transmitted to the inputs of the visible layer of the next RBM. In

Table 4
BFF values of 31 participants for location dataset analysis.

Participant's nickname	O	C	E	A	N
asy	3.400	3.222	3.375	3.333	3.125
bigjam	3.100	3.556	3.500	2.778	2.500
bjh	2.700	3.222	3.250	2.667	2.750
bobtong	2.000	2.889	3.875	2.778	2.875
bongtangE	3.000	3.333	3.250	2.556	2.875
byeolmyeong	3.800	3.556	2.750	3.111	3.375
cdy	4.000	3.667	4.000	3.889	2.750
cuda	3.400	3.222	3.625	3.222	3.250
ihn	3.100	3.111	3.500	2.667	2.875
jsh	3.400	3.556	3.625	2.889	2.500
judong	3.100	3.444	3.000	3.333	2.875
kdi	2.200	3.444	3.000	3.111	2.625
kdy	4.200	4.333	3.500	3.556	2.625
khi	3.100	3.000	3.250	3.333	3.250
kjw	2.600	2.778	3.375	3.111	2.625
knh	3.400	3.000	3.000	2.889	2.500
ksy	4.333	3.125	2.250	3.200	2.889
leb	3.300	3.889	3.250	3.667	2.625
lhn	3.000	3.444	3.250	2.444	3.000
ljs	3.600	3.333	2.750	3.222	2.750
lkh	2.400	3.444	3.500	3.333	3.125
lyr	3.100	3.667	3.375	3.222	3.500
park	3.400	3.556	3.375	2.778	2.875
pebble	3.100	3.444	3.500	2.444	2.875
psy	3.200	3.222	3.500	3.333	2.875
rhm	3.300	2.889	3.125	3.111	3.250
sdi	3.100	3.444	3.000	3.333	2.875
shy	3.800	4.000	3.125	3.778	2.250
sol	3.800	3.556	3.750	2.889	2.750
ydb	3.500	3.778	3.375	3.222	3.000
yik	2.000	2.889	3.875	2.778	2.875

Table 5
Meaning of output for positioning data analysis.

Output layer	Specification
Unit 1	Probability to stay at home
Unit 2	Probability to move from home to work
Unit 3	Probability to move from home to etc.
Unit 4	Probability to move from work to home
Unit 5	Probability to stay at work
Unit 6	Probability to move from work to etc.
Unit 7	Probability to move from etc. to home
Unit 8	Probability to move from etc. to work
Unit 9	Probability to stay at etc.

our study, we chose a sigmoid function as the activation function. The pretraining will be repeated for all RBMs until the uppermost RBM is reached, and then, the fine tuning will begin. Fine tuning basically adjusts the weight so as to minimize errors. Based on the result of the uppermost hidden layer, the inputs to the adjacent lower hidden layers are assumed using the error back propagation algorithm [30].

Based on the input data and the DBN, the neural network will be generated and the weights will be initialized. Since there is no output layer in the DBN, similar training will be conducted for the DNN with the same neural network generated by the DBN. Now, the DNN for classification requires fine tuning. The input and output values are fed to the DBN, and the result values are predicted. The difference between the predicted result and the observed result value are found using an error function, and the weight values in the hidden layers are calibrated using the error back propagation algorithm [31]. Fig. 1 shows the outline structures of the DBN and DNN used in this research. Except for the output layer, the DBN and DNN possess the same neural network structure.

For the neural network structure used in this study, the number of units in the visible layer is five, which is the same as the number of BFFs. In vector form, the input vector is in the form of $<O, C, E, A, N>$. There are different numbers of units in the output layer. For positioning data analysis, nine units are used in the output layer. Table 5 shows the mapping between the output layer and the transition probability, for positioning data analysis. For location data analysis, four units are used in the output layer. Table 6 shows the mapping between the output layer and the staying probability, for location data analysis. Each hid-

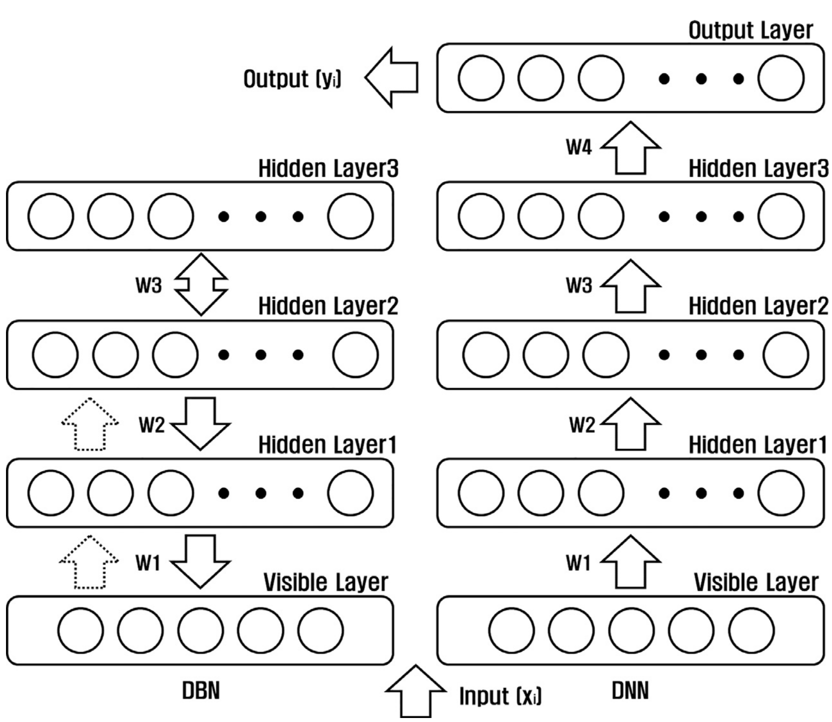

Fig. 1. An outlook of artificial neural network.

Table 6
Meaning of output for location data analysis.

Output layer	Specification
Unit 1	Probability to visit home
Unit 2	Probability to visit work
Unit 3	Probability to visit restaurant
Unit 4	Probability to visit etc.

den layer has 500 units [32]. In summary, the neural networks for experiments have five units in the input layer, for both positioning data analysis and location data analysis. Neural networks have nine units in the output layer for positioning data analysis and four units in the output layer for location data analysis. Weights between each layer, including the multiple hidden layers, will be determined by the training process, for both DBNs and DNNs. DBN is used for pretraining. The weights between layers will be initialized and trained. Then, the DNN, which has the same layer structure as the DBN with the pretrained weights, will be trained using the input and output vectors. The DBN and DNN training procedures are the same for both positioning data analysis and location data analysis.

5. Experiment method

Using the prepared DBN and DNN, a set of experiments are conducted on the mobility information set and the personality information set. In order to train the neural network, the datasets are divided into two groups: training data group and verification data group.

Each participant's data are selected as the verification data while all other participant's data are regarded as the training data group. Positioning data and location data are processed separately with different neural networks. Therefore, 17 verification data groups exist for a positioning dataset with 17 participants, and 31 verification data groups exist for a location dataset with 31 participants. For verification, the input value to the training neural network is the participant's personality factor, and the mobility probability value will be obtained as the result. The results predicted by the neural network will be compared with the observed values, and therefore, the error rates can be calculated as in Eq. (6).

$$\text{error rate} = \frac{\sum |y_i - o_i|}{2}, \qquad y_i = \begin{cases} 1, & p_i \geq \text{threshold} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where o_i is the observed value and y_i is the predicted value. The threshold value is 0.4, considering probability, in this research. Instead of the error rate, we are interested in the hit rate, as shown in Eq. (7).

$$\text{hit rate} = 1 - \frac{\sum \text{error rate}}{S} \tag{7}$$

where S is the number of verification data. The hit rates of all participants will be obtained for each participant, and the average of the hit rates is used to find the optimal neural network with the highest hit rate. The followings process is used to find an optimal neural network.

(i) A neural network has learning rates and a number of hidden layers. These parameters should be calibrated such that a high hit rate is obtained.
(ii) With the optimal neural network, the neural network is trained using the location clusters from the positioning dataset, and the result is analyzed.
(iii) With the optimal neural network, the neural network is trained using the location dataset, and the result is analyzed.

Table 7
Hit rates with various number of units on hidden layer.

Participants	Number of units in hidden layer			
	3	45	150	500
Park	0.50000	0.68918	0.0180	0.13063
bigjam	0.34550	0.34550	0.50000	0.44382
bjh	0.31135	0.31135	0.31135	0.32244
bobtong	0.59288	0.59288	0.59288	0.50000
bongtangE	0.50000	0.50000	0.50000	0.28000
byeolmyeong	0.42070	0.42070	0.42070	0.42070
cdy	0.33168	0.33168	0.27820	0.33168
cuda	0.34014	0.34014	0.13104	0.34014
jsh	0.06921	0.06921	0.54344	0.54344
judong	0.50000	0.916666	0.50000	0.05555
kdy	0.26983	0.26983	0.31826	0.26983
ksy	0.53959	0.50000	0.53959	0.13323
leb	0.16642	0.16642	0.16642	0.70582
ljs	0.29543	0.50000	0.29543	0.43190
samhwa	0.50000	0.50000	0.06607	0.06607
shy	0.08516	0.08516	0.08516	0.82967
ydb	0.42947	0.50000	0.42947	0.11834

We found that 2 and 3 are suitable values for the number of hidden layers [33]. Finally, the neural network is trained with all participants data. For the experiment, the inputs to the neural network were prepared as synthetic personality data, as shown in Table 11. Such synthetic personality data will show the effect of each personality factor on the mobility pattern.

6. Results

The analysis of hit rate and the prediction of mobility patterns are the main issues addressed in this section. The details of finding an optimal neural network will be addressed, and the results from the positioning dataset and location dataset will be analyzed separately. Synthetic personality data are used to identify the relationship between personality and mobility.

6.1. Optimal network

We focused on two parameters of neural networks, to find the optimal neural network: the number of units in each hidden layer and the number of hidden layers. We verified by setting the value of the number of units as 3, 45, 150, and 500. Table 7 shows the hit rate values according to the number of units in the hidden layer, for the positioning data analysis. With a small number of units in the hidden layer, a hit rate of 0.5 is mostly observed, and it is identified that classification does not work well in the neural network. Therefore, the number of units for an optimal neural network is decided as 500. Then, the number of hidden layers is decided. Among the various possibilities, 2 and 3 are chosen, and the learning rate is set from 0.025 to 0.5. The learning rate is not a part of the neural network structure, but is a parameter for training the neural network. With these learning rates and numbers of hidden layers, experiments for comparison are made.

Figs. 2 and 3 show the experimental results of hit rate. Fig. 2 is obtained from the positioning dataset, when two and three hidden layers are used. Fig. 3 is obtained from the location dataset, when two and three hidden layers are used.

For each graph, the x-axis shows the learning rate of the neural network and the y-axis shows the hit rate between the predicted value and observed value. For both cases, a higher average hit rate is found for the case with two hidden layers. The learning rate for obtaining the highest hit rate is found to be 0.04. With a very small learning rate, the hit rate is mostly 0.5, implying imperfect training of the neural network. For example, Table 8 shows

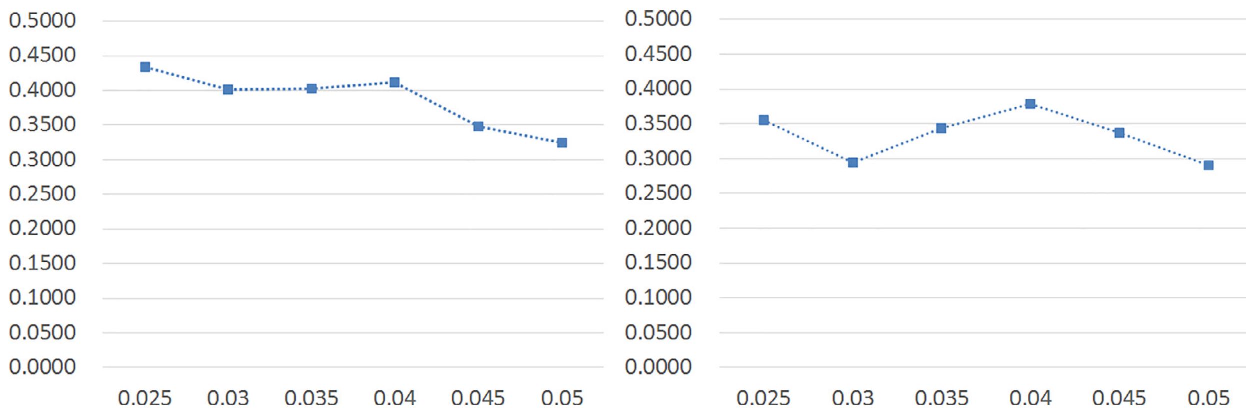

Fig. 2. Hit rate with 2 hidden layers (left) and 3 hidden layers (right) using positioning dataset.

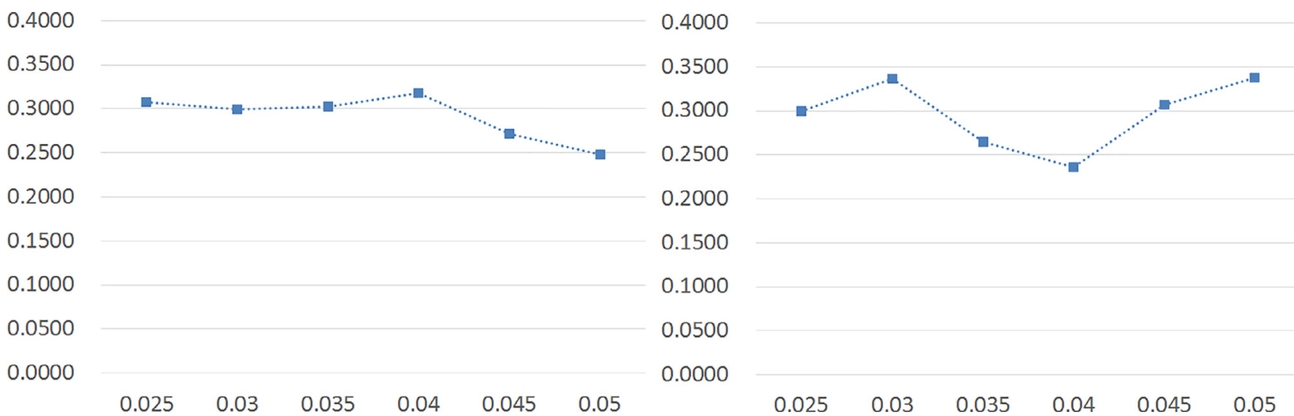

Fig. 3. Hit rate with 2 hidden layers (left) and 3 hidden layers (right) using location dataset.

Table 8
Hit rate in neural network with 2 hidden layers and 0.015 of learning rate.

Participants	Park	bigjam	bjh	bobtong	bongtangE
Hit rate	0.689	0.500	0.500	0.592	0.280
Participants	byeolmyeong	cdy	cuda	jsh	judong
Hit rate	0.500	0.500	0.500	0.543	0.916
Participants	kdy	ksy	leb	ljs	samhwa
Hit rate	0.500	0.539	0.705	0.500	0.066
Participants	shy	ydb			
Hit rate	0.500	0.500			

Table 9
Observed vs. predicted values with optimal neural network for participant 'leb'

	H to H	H to W	H to E	W to H	W to W
Observed value	0.166422	0.001468	0.039648	0.000979	0.028879
Predictive value	0.000159	0.000728	0.016401	0.001143	0.000006
	W to E	E to H	E to W	E to E	
Observed value	0.008321	0.040627	0.007832	0.705825	
Predictive value	0.000748	0.004727	0.004465	0.971621	

Table 10
A observed value and predictive value with optimal neural network by participant 'cuda'

	Home	Work	Restaurant	etc.
Observed value	0.140741	0.185185	0.148148	0.525926
Predictive value	0.008533	0.243748	0.000057	0.747663

the hit rate for two hidden layers and a learning rate of 0.015. A hit rate of 0.5 can be observed for several participants.

On the contrary, learning rates higher than 0.05 lead to lower hit rates. In summary, for both datasets, optimal neural networks have two hidden layers and the learning late is 0.04.

With the optimal network, the hit rate for the positioning dataset is 41%, and the hit rate for the location dataset is 31%. Considering the pseudorandomness of human mobility, it appears as if we found adequate hit rates, at 31% or 41%.

6.2. Location prediction with positioning dataset

The results from the optimal neural network using the positioning dataset were verified. Table 9 shows the observed values versus the predicted values for the personality data provided by a participant with the nickname 'leb'.

Each value represents the probability of transition between locations. The observed values show that the probability that the participant will stay at E is 70%, H is 16%, and W is 2%. The observed transition probability from E to H is 4%. The predicted values show that the stay probability at E is 97%, that at W is 0.0006%, and that at H is 0.01%. The predicted transition probability from E to H is 0.4%. The other values are small enough to ignore. Fig. 4 shows the Markov chain representation of the predicted transition values.

6.3. Location prediction with location dataset

A similar verification is done for the location dataset. Table 10 shows the predicted values and observed values for a participant 'cuda'.

Each value stands for the probability of a location visit. The observed probabilities are: 52% for E and 18% for W. The predicted probabilities are: 72% for E and 24% for W.

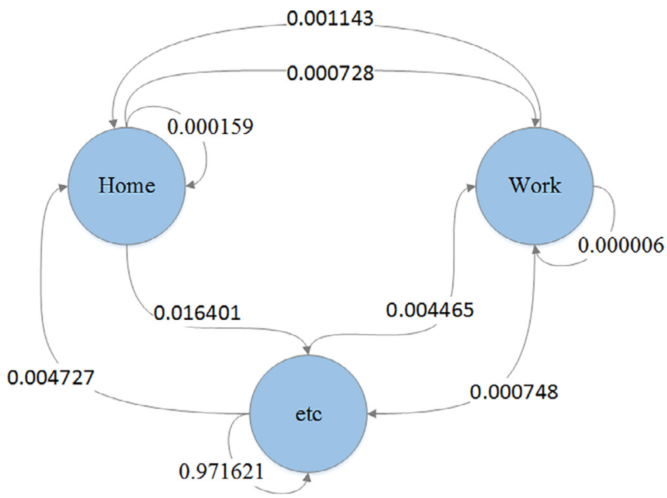

Fig. 4. A Markov model representation of mobility prediction by participant 'leb'.

Table 11
A synthetic BFF dataset for experiments.

BFF	O	C	E	A	N
Normal	2.5	2.5	2.5	2.5	2.5
High-O	4.0	1.0	1.0	1.0	1.0
High-C	1.0	4.0	1.0	1.0	1.0
High-E	1.0	1.0	4.0	1.0	1.0
High-A	1.0	1.0	1.0	4.0	1.0
High-N	1.0	1.0	1.0	1.0	4.0

Table 12
A correlation analysis with neural network trained by positioning dataset.

BFF	H to H	H to W	H to E	W to H	W to W
Normal	0.446341	0.021166	0.058967	0.019322	0.100046
High-O	0.924673	0.015697	0.003324	0.015336	0.005256
High-C	0.950084	0.001696	0.001775	0.001523	0.000155
High-E	0.655989	0.052477	0.073419	0.042243	0.000133
High-A	0.163857	0.000122	0.000070	0.000135	0.835387
High-N	0.985574	0.005699	0.000010	0.004049	0.000016
BFF	W to E	E to H	E to W	E to E	
Normal	0.013159	0.054829	0.013179	0.272989	
High-O	0.011057	0.002773	0.011147	0.010736	
High-C	0.001636	0.000234	0.001595	0.041303	
High-E	0.033676	0.103759	0.027050	0.011254	
High-A	0.000062	0.000127	0.000082	0.000158	
High-N	0.002208	0.000006	0.001493	0.000945	

6.4. Analysis of personality–location relationship

With an optimal network, it is possible to identify the effects of a human's personality on a human's favorite locations. For this purpose, we prepared a synthetic personality model, which has artificial values, to emphasize a specific factor of BFF as shown in Table 11. A BFF set for a normal personality, as well as BFF sets for emphasizing each of O, C, E, A, and N are prepared.

In terms of BFF, 2.5 is the average value. In order to emphasize, for example, the openness, a value of 4.0 can be set for O while 1.0 is set to minimize the effects of other factors, C, E, A, and N.

Our optimal neural network has inputs from Table 11, and Table 12 shows the results of the experiment with the positioning dataset. Table 12 shows the following results: openness, conscientiousness, and neuroticism affects the probability of staying at home highly. In the case of high extraversion, transitions from W to E and from E to H show high probabilities. High agreeableness causes a high probability to stay at W.

For the location dataset, Table 13 shows the results.

Table 13
A correlation analysis with neural network trained by location dataset.

BFF	Home	Work	Restaurant	etc.
Normal	0.000119	0.837706	0.106797	0.055378
High-O	0.015109	0.000435	0.004197	0.980259
High-C	0.243247	0.040468	0.663421	0.052865
High-E	0.000769	0.000001	0.000134	0.999096
High-A	0.000011	0.999902	0.000087	0.000001
High-N	0.000049	0.579682	0.000549	0.419764

Table 14
Collected mobility information for new participants.

Participants	Number of positions	Number of check-ins
khb	16,326	187
sck	23,149	81

Table 15
BFF values of new participants for verification.

Participants	O	C	E	A	N
khb	3.300	2.600	3.444	4.100	3.780
sck	2.800	3.500	4.100	3.444	3.750

Table 16
A prediction of mobility pattern by positioning dataset.

Participants	H to H	H to W	H to E	W to H	W to W
khb	0.97457	0.00402	0.00410	0.00414	0.00002
sck	0.96570	0.00061	0.00023	0.00156	0.03093
	W to E	E to H	E to W	E to E	hit rate
khb	0.00414	0.00213	0.00384	0.00299	0.46335
sck	0.00027	0.00028	0.00036	0.00003	0.14925

High openness and high extraversion cause high probability to visit etc. locations. High conscientiousness causes high probability to visit restaurants High agreeableness and high neuroticism causes high probability to visit work.

Considering both cases with positioning data and location data in common, high agreeableness causes frequent visits to work. High openness and high conscientiousness causes relatively frequent visits to home. However, in the case of openness and conscientiousness, the two types of datasets show different results. This maybe because of the nature of data collection. In the case of passively collected positioning data, the duration of stay at certain locations is recorded, whereas a visit to a certain location is recorded only once in the case of actively collected location data. For example, continual staying at home may cause several hours of stay while the number of check-ins is only one.

6.5. Verification of trained neural network with additional data

For the verification of our results, extra experiments were conducted. Two new datasets were acquired from two new participants. In addition, the positioning dataset, location dataset, and the BFF of two new participants were prepared. Tables 14 and 15 show a summary of the two new datasets: number of positions, number of check-ins, and BFF values.

Table 16 shows the results for the positioning dataset analysis for two new participants with already trained neural networks. Table 16 shows the highest predicted value for home stay. Previous experiments showed that the related BFFs were openness, conscientiousness, and neuroticism. Both new participants showed high neuroticism, and thus, the new participants tended to stay at home with highly probabilities. This result agrees with the previous

Table 17
A prediction of mobility pattern by location dataset.

Participants	Home	Work	Restaurant	etc.	Hit rate
khb	0.00001	0.90599	0.06095	0.03304	0.39038
sck	0.74588	0.00013	0.00039	0.25358	0.03703

result, and a similar interpretation can be found in [10], that neuroticism increases the probability of staying at home.

Location dataset analysis was also done. A trained neural network was used.

Table 17 shows the results with the location dataset. Participant “khb” is predicted to be at work with high probability, owing to the high agreeableness and neuroticism of BFF. Previous experiments showed that the location work was highly related with agreeableness and neuroticism. Participant “sck” is predicted to be at home with high probability, since sck’s BFF has a high conscientiousness of 3.5. It can be inferred that high conscientiousness leads to a high tendency of staying at home. However, sck has high extraversion and sck is predicted to be at etc. according to the previous results shown in Table 13. This inconsistency is due to the small number of collected location dataset, which leads to low hit rates of 0.03703. In addition, the location dataset tends to be highly skewed, since the participants must check-in actively, which cannot be forced on volunteers.

7. Conclusions

In this paper, the relationship between human personality and human mobility was studied using DBNs and DNNs along with connected BFFs and the mobility information provided by volunteers. The mobility information was divided into two categories: positioning data, which were collected passively and location data, which were collected actively by participants. We also tried to predict an individual’s mobility patterns based on trained neural networks.

Among the various deep learning models, DBN was used to generate the neural networks and initialize weights. DNN was used to train using data and verify the results. We also tried to find the highest hit rate by identifying the optimal parameters for the neural networks. With the resulting optimal neural network, the relationship between personality and mobility was found. In detail, verification was done with the input dataset and the effect of each factor of personality on the human mobility was expressed.

In the case of training using the positioning dataset, the neural network showed an average hit rate of 41%, a maximum hit rate of 91%, and a minimum hit rate of 3%. In the case of training with the location dataset, the neural network showed an average hit rate of 31%, a maximum hit rate of 81%, and a minimum hit rate of 1%. With these results, it could be assumed that knowledge of human personality, in combination with the human mobility model, will be able to predict the mobility pattern. Among the five personality factors, agreeableness was related to the work place, and openness and conscientiousness were related to home.

Our result can provide a foundation for the differentiated customized services of location-based service (LBS). As introduced in Section 6, passive mobility information can construct a Markov model and enable location prediction from the personality model [34].

In addition, we tried a new application of machine learning. As we have prepared the base for a method to figure out the personality–mobility relationship using deep learning techniques, the results will be more precise with the evolution of deep learning. The deep learning models presented in Section 2 – the deep convolution neural network (DCNN), deep recurrent neural network (DRNN), regions with CNN (R-CNN), and faster R-CNN – are other candidates for our purpose [35–37].

We have several future research options. First, we plan to recruit more participants so that the results will be more precise. Second, the sigmoid function is now a function for neural network training; however, we plan to use the rectified linear unit (ReLU) to solve the vanishing gradient problem and obtain a higher hit rate [38]. Finally, we can use the GPGPU or other performance-oriented computer systems for the high computational requirements of deep learning. With the current DNN and DBN, it takes two minutes for training a neural network with two hidden layers, and five minutes for training neural networks with three hidden layers. Since more participants directly implies more training datasets and a larger number of hidden layers, we need to reduce the processing time, which is actually the training time for the actual experiments.

Acknowledgment

Authors give special thanks to participants who collected and submitted their personality and mobility information voluntarily. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2017R1D1A1B03029788).

References

[1] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of go with deep neural networks and tree search, Nature 529 (7587) (2016) 484–489.
[2] S.Y. Kim, H.J. Koo, H.Y. Song, A study on influence of human personality to location selection, J. Ambient Intell. Hum. Comput. 7 (2) (2016) 267–285.
[3] H.Y. Song, Probabilistic space-time analysis of human mobility patterns, WSEAS Trans. Comput. 15 (2016) 222–238.
[4] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. R. Stat. Soc. Ser. B Methodol. (1977) 1–38.
[5] H.Y. Song, D.Y. Kim, A method for group mobility model construction and model representation from positioning data set using GPGPU, KIPS Trans. Softw. Data Eng. 6 (3) (2017) 141–148.
[6] R.W. Byrne, Memory for urban geography, Q. J. Exp. Psychol. 31 (1) (1979) 147–154.
[7] J. Anable, ‘Complacent car addicts’ or ‘aspiring environmentalists’ identifying travel behaviour segments using attitude theory, Transp. Policy 12 (1) (2005) 65–78.
[8] B. Verplanken, H. Aarts, A. Van Knippenberg, Habit, information acquisition, and the process of making travel mode choices, Eur. J. Soc. Psychol. 27 (5) (1997) 539–560.
[9] P.S. Wolf, W.J. Jacobs, GPS technology and human psychological research: a methodological proposal, J. Methods Measur. Soc. Sci. 1 (1) (2010) 1–7.
[10] H.Y. Song, E.B. Lee, An analysis of the relationship between human personality and favored location, Proceedings of the AFIN 2015, 2015, p. 12.
[11] Y. Bengio, N. Boulanger-Lewandowski, R. Pascanu, Advances in optimizing recurrent networks, in: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 8624–8628.
[12] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting., J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[13] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[14] C. Szegedy, A. Toshev, D. Erhan, Deep neural networks for object detection, in: Proceedings of the Advances in Neural Information Processing Systems, 2013, pp. 2553–2561.
[15] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.
[16] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[17] T.N. Sainath, A.-r. Mohamed, B. Kingsbury, B. Ramabhadran, Deep convolutional neural networks for LVCSR, in: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013, pp. 8614–8618.
[18] T. Mikolov, M. Karafiát, L. Burget, J. Cernockỳ, S. Khudanpur, Recurrent neural network based language model., in: Proceedings of the Interspeech, vol. 2, 2010, p. 3.
[19] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber, A novel connectionist system for unconstrained handwriting recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (5) (2009) 855–868.
[20] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, Neural Comput. 18 (7) (2006) 1527–1554.

[21] Sportstracker, (http://www.sports-tracker.com).
[22] Garmin gpsmap810, (https://buy.garmin.com/en-US/US/into-sports/cycling/edge-810/prod112912.html).
[23] Swarm app, (https://www.swarmapp.com/).
[24] D.P. Schmitt, J. Allik, R.R. McCrae, V. Benet-Martínez, The geographic distribution of big five personality traits patterns and profiles of human self-description across 56 nations, J. Cross Cultural Psychol. 38 (2) (2007) 173–212.
[25] J. Block, A contrarian view of the five-factor approach to personality description., Psychol. Bull. 117 (2) (1995) 187.
[26] G. Chittaranjan, J. Blom, D. Gatica-Perez, Who's who with big-five: analyzing and classifying personality traits with smartphones, in: Proceedings of the 2011 Fifteenth Annual International Symposium on Wearable Computers, IEEE, 2011, pp. 29–36.
[27] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[28] G.E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800.
[29] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, et al., Greedy layer-wise training of deep networks, Adv. Neural Inf. Process. Syst. 19 (2007) 153.
[30] P.J. Werbos, Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, PhD thesis, Harvard University, 1974.
[31] G.E. Nasr, E. Badr, C. Joun, Cross entropy error function in neural networks: forecasting gasoline demand., in: Proceedings of the FLAIRS Conference, 2002, pp. 381–384.
[32] S. Xu, L. Chen, A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining, in: Proceedings of the 5th International Conference on Information Technology and Applications (ICITA '08), 2008, pp. 683–686.
[33] G. Kenyon, H. Paugam-Moisy, Multilayer neural networks: one or two hidden layers, Adv. Neural Inf. Process. Syst. 9 (1997) 148–154.
[34] S. Bellahsene, L. Kloul, A new Markov-based mobility prediction algorithm for mobile networks, in: Proceedings of the European Performance Engineering Workshop, Springer, 2010, pp. 37–50.
[35] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[36] M. Hermans, B. Schrauwen, Training and analysing deep recurrent neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2013, pp. 190–198.
[37] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2015, pp. 91–99.
[38] V. Nair, G.E. Hinton, Rectified linear units improve Restricted Boltzmann Machines, in: Proc Icml, 2010, pp. 807–814.

Dongyup Kim received his B.S. degree in Computer Engineering in 2015 and received his M.S. degree in Computer Engineering in 2017, respectively, both from Hongik University, Seoul, Korea. His research interests are in the areas of parallel computing, artificial network, internet of things and human mobility modeling.

Hayoon Song received his B.S. degree in Computer Science and Statistics in 1991 and received his M.S. degree in Computer Science in 1993, both from Seoul National University, Seoul, Korea. He received Ph.D. degree in Computer Science from University of California at Los Angeles, USA in 2001. From 2001, he has worked at Department of Computer Engineering, Hongik University, Seoul, Korea and is now a professor. In his sabbatical year 2009, he worked at Institute of Computer Technology, Vienna University of Technology, Austria as a visiting scholar. His research interests are in the areas of mobile computing, application of machine learning, internet of things and human mobility modeling.

Please cite this article as: D.Y. Kim, H.Y. Song, Method of predicting human mobility patterns using deep learning, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.07.069