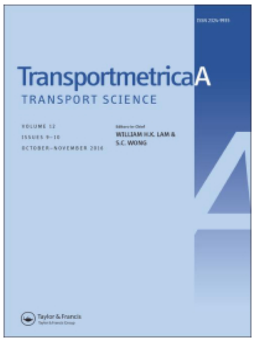

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Combining individual travel behaviour and collective preferences for next location prediction

Qiuping Li, Dan Zou & Yang Xu

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2021.1968066

Combining individual travel behaviour and collective preferences for next location prediction

Qiuping Li[a,b], Dan Zou[a,b] and Yang Xu[c,d]

[a]School of Geography and Planning, Sun Yat-Sen University, Guangzhou, People's Republic of China; [b]Guangdong Provincial Engineering Research Center for Public Security and Disaster, Guangzhou, People's Republic of China; [c]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong; [d]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, People's Republic of China

ABSTRACT

Many mobility prediction models have emerged over the past decade to predict a user's next location through the utilisation of user trajectories. However, the performance is constrained by the quantity of user trajectory data. This research introduces a new approach by combining knowledge of individual travel behaviour and collective preferences of users sharing similar daily activity patterns. First, users are clustered into different groups by their daily activity profiles. Second, each group's collective preferences (i.e. activity and travel distance preferences) are extracted. Then, individual travel behaviour and collective preferences are integrated for the next location prediction. A mobile phone dataset from Shanghai, China, is used for model validation. The results show that the proposed model achieves a prediction accuracy of over 80% during most of the day. Moreover, there is a maximum increase of 16% in prediction accuracy compared with baseline models when users are highly mobile.

ARTICLE HISTORY
Received 19 November 2020
Accepted 7 August 2021

KEYWORDS
Human mobility prediction; distinct daily activity patterns; mobile phone positioning data; user clustering; collective preferences

1. Introduction

The rapid development of sensor networks and mobile internet technologies has provided massive trajectory data for human mobility research and applications. People's movements in urban areas reflect the complicated relationship between human beings and urban space (Shaw and Sui 2018; Yuan 2018). Understanding and predicting human mobility is crucial for transportation demand estimation (Huang et al. 2018), urban planning and management (Yuan, Zheng, and Xie 2012), individual location recommendations (Calabrese, Lorenzo, and Ratti 2010), and disease prevention and control (Wen, Hsu, and Hu 2018). It can also help us deepen the understanding of the dynamic interactions between humans and urban geographical space.

Research on human mobility modelling has progressed very rapidly over the past decade. A large number of human mobility models have emerged. These studies are either at the collective level or the individual level. Models at the collective level, such

CONTACT Qiuping Li liqp3@mail.sysu.edu.cn School of Geography and Planning, Center of Integrated Geographic Information Analysis, Sun Yat-sen University, Guangzhou 510275, People's Republic of China

Table 1. Summary of selected research in individual next location prediction.

Related work	Basic Model	Main data source	Model features
Huang (2017)	Markov model	Social media data	Considers the user's historical trajectories and his/her online behaviours
Chen et al. (2019)	Bayesian model	Social media data	Considers the user's historical trajectories and his/her online behaviours
Li, Lu, et al. (2020)	LSTM	Cellular data	Considers the user's historical trajectories and the closeness and periodic movement patterns in the trajectories
Yu et al. (2015)	Markov model	GPS trajectory data	Combines the user's activity pattern and the common activity pattern of all available users
Hawelka et al. (2017)	EW forecaster	Cellular data	Leverages the user's historical trajectories and the mobility traces of others
Wang et al. (2020)	Bayesian mixture model	Cellular data	Leverages the user's historical trajectories and the mobility traces of others

as the gravity model and radiation model (Zipf 1946; Simini et al. 2012), aim to understand the mobility flow between two different locations. However, they are mostly static, and the detailed dynamic features of human movements at the individual level are often lost (Yan et al. 2017). A large portion of human mobility prediction models at the individual level aim to predict an individual's next location based on their historical trajectories (Calabrese, Lorenzo, and Ratti 2010; Huang 2017; Lv et al. 2017; Zhao, Koutsopoulos, and Zhao 2018; Chen et al. 2019; Li, Gui, et al. 2020; Wang et al. 2020). The main related works, including basic models, main data sources, and features of models, are listed in Table 1. As seen in Table 1, the individual location prediction models can be classified into two groups. In the first group, the prediction models use only the user's past locations and other information accompanying the user's trajectory to estimate the next location. For instance, Huang (2017) proposed a Markov model for predicting the next location of a social media user using their long-term cumulative sparse trajectory data and online behaviours. Recently, recurrent neural network and long short-term memory (LSTM) models have become increasingly popular in next location prediction because they are stable and robust for modelling long-range dependencies (Li et al. 2020; Choi, Yeo, and Kim 2018; Li, Lu, et al. 2020). Li, Lu, et al. (2020) proposed a fuzzy LSTM model that considers the user's trajectory and the closeness and periodic movement patterns revealed in the trajectory. The prediction accuracy improved compared to the traditional Markov model. The main disadvantage of these models, however, is that their prediction performance is constrained by the quantity of user trajectory data. If the historical trajectory data are short or very sparse, the parameters in the abovementioned models (e.g. Markov model and LSTM model) are difficult to estimate, leading to low prediction accuracy (Baumann et al. 2018; Wang et al. 2020). Furthermore, the predicted locations can only be selected from user historical locations. Therefore, it is difficult for these models to accurately predict when people are continually exploring new locations during movement (Baumann, Kleiminger, and Santini 2013; Cuttone, Lehmann, and González 2018).

The second group of models leverages not only the user's trajectory but also the trajectories of others to enhance the prediction performance (Calabrese, Lorenzo, and Ratti 2010; De Domenico, Lima, and Musolesi 2013; Hawelka et al. 2017; Yu et al. 2015; Wang et al. 2020). For instance, Calabrese, Lorenzo, and Ratti (2010) proposed a weighted combination of the individual's past trajectory and collective geographic preferences for user next location prediction. Yu et al. (2015) combined the individual's activity pattern and the common activity pattern in their model, and the weights of these two components were dynamically adjusted depending on the volume of the user's historical trajectories. For users' location prediction with a short and nonrepetitive data history, such as tourists in a foreign country, Hawelka et al. (2017) proposed a sequential learning algorithm that leveraged the mobility traces of other users. A recent study by Wang et al. (2020) proposed a multitask learning-based algorithm to predict users' mobility by learning the mobility behaviours of others to overcome the sparseness issue and improve prediction performance. The results in these studies show the advantage of leveraging the trajectories of other users in the prediction model. However, the population is viewed as a single and homogeneous traveller group in these models. The collective mobility patterns or preferences are extracted from the entire population, and the heterogeneity of users' daily activity patterns is ignored. In fact, people with different daily activity patterns, such as regular workers and stay-at-home workers, have distinct travel behaviours and location preferences (Jiang, Joseph Ferreira, and Gonzalez 2012; Lv et al. 2017; Chen et al. 2018; Yang, Yan, and Ukkusuri 2018). Further consideration of the heterogeneous daily travel patterns of users has the potential to improve the accuracy of the abovementioned location prediction models.

In this paper, we introduce a new approach for next location prediction by combining the knowledge of individual travel behaviour with the location preferences of users that share similar daily activity patterns. In contrast to most existing models, for users with limited trajectory data, the proposed method can leverage the location information of people who have similar daily activity patterns. First, users are clustered into different groups by their daily activity patterns. Second, each group's collective location preferences (i.e. activity and travel distance preferences) are extracted. Then, for each user, their travel behaviours and the group's collective location preferences to which they belong are integrated to predict their next location. Finally, mobile phone positioning data of users in Shanghai, China, are used to evaluate the proposed model's performance.

The rest of this paper is organised as follows. Section 2 describes the methodology, including user clustering, the individual mobility prediction model, and the collective location preference calculation. Section 3 introduces the study area and used data. The experimental results and analysis are presented in Section 4. Section 5 discusses the impacts of the estimation of missing locations and activities and the impacts of collective location preferences on the proposed model. The final section provides concluding remarks and future research directions.

2. Methodology

In this study, we aim to predict the next location of a user given their historical trajectories and the group's collective location preferences to which they belong. For simplicity, the urban space is divided into N grids, and one grid represents a location. Before establishing the mathematical model, we first introduce the definitions in this research.

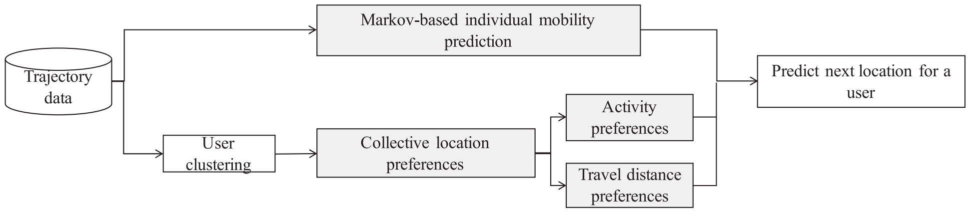

Figure 1. The overall workflow for building the prediction model.

Definition 1. *A location set DL_set (u) is the set of historical trajectories of user u in m days. It is defined as follows:*

$$DL_{set(u)} = \left\{Day_1(u) = \left\langle x_u^1, \ldots, x_u^t, \ldots, x_u^n \right\rangle, \ldots, Day_m(u) = \left\langle x_u^1, \ldots x_u^{t-1} \right\rangle\right\},$$

where $Day_m(u)$ is a sequence of locations of user u on the mth *day, and x_u^t represents the location of user u at time t-1.*

Definition 2. *An activity set DA (u) is the set of activities of user u in m days. It is defined as follows:*

$$DA\_set(u) = \left\{Day_1(u) = \left\langle a_u^1, \ldots, a_u^t, \ldots, a_u^n \right\rangle, \ldots, Day_m(u) = \left\langle a_u^1, \ldots, a_u^{t-1} \right\rangle\right\},$$

where $Day_m(u) = \left\langle a_u^1, \ldots, a_u^{t-1} \right\rangle$ is a sequence of activities of user u on the mth *day, and a_u^t is the activity that user u participates in at time t-1. In this definition, the activity type is 'home', 'work', or 'other'.*

The following section presents the methodology for predicting a user's next location by combining their travel behaviour and collective location preferences (i.e. activity and travel distance preferences). The workflow of the proposed model is shown in Figure 1. The three main components of the model are user clustering, Markov-based location prediction based on an individual's historical trajectories, and collective location preference extraction for each cluster of users.

2.1. User clustering

Users are clustered based on their daily activity profiles. Before clustering, the location set of user *u* should be converted to an activity set. Because activity type information is unavailable in mobile phone trajectories, activities such as 'home' and 'work' are usually identified by space and time constraints (Liu et al. 2014; Alexander et al. 2015; Tu et al. 2017). For the sequence of locations of user *u* on the *i*th day, if the distance of several consecutive location records is less than 500 m and the duration from the first to the last record of these locations is more than 4 h at night, their activity type is identified as 'home' (Cao et al. 2017). Similarly, if the distance of several consecutive location records is less than 500 m and the duration from the first to the last record of these locations is more than 3 h during the daytime, their activity type is identified as 'work' (Cao et al. 2017). Then, the activity type of the remaining location records in *DL_set*(u) is identified as 'other'.

After each location in $DL\_set(u)$ is labelled correctly as 'home', 'work', or 'other', the frequency of each type of activity at each time t for user u in a total of m sample days is calculated. For example, the frequency of 'home' activity fh_u^t is calculated as follows:

$$fh_u^t = \frac{\sum_{i=1}^{i=m} A_h(t,i)}{n} \tag{1}$$

$$A_h(t,i) = \begin{cases} 1, & \text{if activity type at time } t \text{ of the ith day is 'home'} \\ 0, & \text{else} \end{cases} \tag{2}$$

Similarly, the frequencies of 'work' and 'other' activities are also calculated. Then, the daily activity frequency sequence of user u can be represented by the following vector:

$$\begin{aligned} AF(u) = \{&Home(u) = \langle fh_u^1, \ldots, fh_u^t, \ldots, fh_u^n \rangle, \\ &Work(u) = \langle fw_u^1, \ldots, fw_u^t, \ldots, fw_u^n \rangle, \\ &Other(u) = \langle fo_u^1, \ldots, fo_u^t, \ldots, fo_u^n \rangle \} \end{aligned}$$

where fw_u^t is the frequency of the 'work' activity and fo_u^t is the frequency of the 'other' activity.

Finally, the k-means clustering method is used to cluster the daily activity frequency sequences of all users $AF = \{AF(1), \ldots, AF(s)\}$, where s is the number of users. Then, the users can be classified into several groups with heterogeneous daily activity patterns.

2.2. Individual mobility prediction based on users' historical trajectories

Markov models have been widely applied to explain human mobility patterns and predict the locations of individuals because of their simplicity and effectiveness (Huang 2017; Qiao et al. 2018). In this study, we chose the Markov model as the first part of our proposed model to predict the user's next location based on their historical trajectories. The location transition probability matrix (Figure 2) is established and updated as the user's historical location sequence increases. For user u with daily location set DL_set (u), their location choice for time t is modelled as the following formula:

$$P_M(x_u^t = j | x_u^{t-1} = i) = p_{ij}^{t-1}, \quad i, j \in [1, \ldots, N] \tag{3}$$

where p_{ij}^{t-1} is the transition probability from locations i to j updated at time t-1. If the historical location sequence of user u is short, the location transition probability matrix is easily biased. Therefore, it is necessary to leverage the collective location preferences of other users with similar daily travel patterns to obtain better prediction performance.

2.3. Collective location preferences

The other part of the proposed model is the collective location preferences extracted from the trajectories of users who share similar daily activity patterns. The collective location preferences here are users' preferences for different types of activities and different travel distances.

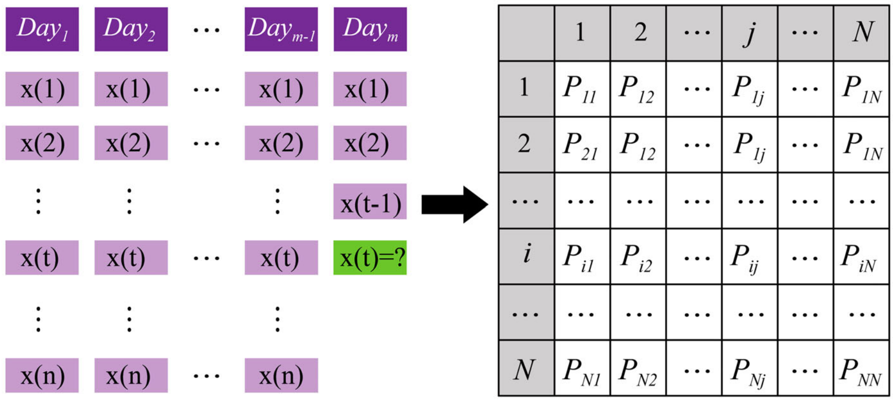

Figure 2. Transition probability distribution based on historical user trajectories.

2.3.1. Activity type preferences

A user's next location is constrained by their activity type (or trip purpose) (Liu et al. 2015; Li et al. 2015; Yu et al. 2015); therefore, the activity information of users with similar daily activity patterns is used as one component of collective location preferences to enhance the performance of the Markov-based individual mobility prediction model in this study.

Users' preferences for 'home', 'work', and 'other' activities are represented by the probability that they choose these activities. The activity transition probability matrix is established from the daily activity sequence of all users in the same group. Given the activity transition matrix of group S, the probability that user u will choose activity a_q for time t is calculated as:

$$P(a_u^t = a_q | a_u^{t-1} = a_p) = p_{a_p a_q}^{t-1} \tag{4}$$

where $p_{a_p a_q}^{t-1}$ is the activity transition probability from activity a_p to activity a_q at time t-1.

Subsequently, the user's next location can be predicted based on their next activity choice. If the current location of individual u is i, the probability they will choose location j as the next location is:

$$P_{A_S}(x_u^t = j | x_u^{t-1} = i) = P(a_u^{t-1} = a_p | x_u^{t-1} = i) \cdot P(a_u^t = a_q | a_u^{t-1} = a_p) \cdot P(x_u^t = j | a_u^t = a_q) \tag{5}$$

where $P(a_u^{t-1} = a_p | x_u^{t-1} = i)$ is the probability that location i is labelled activity a_p. $P(x_u^t = j | a_u^t = a_q)$ is the probability that location j will be selected at time t if the current activity of user u is a_q. It is calculated as follows:

$$P(x_u^t = j | a_u^t = a_q) = \begin{cases} 1 & \text{if } j \in DL\_Set(S) \\ 0 & else \end{cases} \tag{6}$$

where $DL\_Set(S)$ is the location candidate set of all visited locations of users in group S.

2.3.2. Travel distance preferences

In many circumstances, people are likely to travel short distances, which shows a decay pattern in the travel distance of people's trips (Mckercher and Lew 2003; Zheng and Zhou

2017). There is a decaying relationship between travel distance and its frequency, and it differs among users with different daily activity patterns (Yan 2011). Therefore, we build one travel distance decay function for each group of users. The travel distance preferences of each group are calculated by their travel distance decay functions.

For user u in group S, if they go from location i at time t-1, then the probability they will appear at location j at time t is as follows:

$$P_{D_S}(x_u^t = j | x_u^{t-1} = i) = f_d^S(d_{ij}, t-1) \tag{7}$$

where $f_d^S(d_{ij}, t-1)$ is the travel distance decay function of users in group S and d_{ij} is the distance from location i to location j. The travel distance decay function $f_d^S(d_{ij}, t-1)$ is defined as follows:

$$f_d^S(d_{ij}, t-1) = \frac{1}{w * m} \sum_{u=1}^{w} \sum_{k=1}^{m} P(d_{u_{ij}}^{t-1}(k) = d) \tag{8}$$

where $d_{u_{ij}}^{t-1}(k)$ is the distance from location i at time t-1 to location j at time t in the trajectory of user u of the k^{th} day and w is the number of users in group S.

The next location of individual u is then predicted by the linear weighting function of P_M, P_{A_S}, and P_{D_S} as follows:

$$P(x_u^t = j | x_u^{t-1} = i) = \alpha(t) P_M + \beta(t) P_{A_S} + \gamma(t) P_{D_S}, \quad u \in S \tag{9}$$

where $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ are the weights of P_M, P_{A_S} and P_{D_S}, respectively. They are time-variant parameters, and their sum equals 1. We denote our proposed mobility prediction model based on users' travel behaviour and collective location preferences of users with similar daily activity patterns as MUCS.

3. Data description

In this study, we used Shanghai mobile phone positioning data to validate our model's efficiency. Shanghai Unicom provided the data. Shanghai is one of the largest economic centres in China. It has comprehensive urban functions and well-developed transportation systems, providing a wide range of opportunities for residents' daily travel and activities. The mobile phone positioning data include five working days on December 28, 29, and 30 in 2015 and January 4 and 5 in 2016. The experimental area includes all 15 administrative districts of Shanghai except Chongming, as shown in Figure 3.

The positions of mobile phone users were recorded at hourly intervals. The dataset contains 5,776,605 position records of 631,645 phone users, and each user had 8.5 position records a day on average. For privacy protection, this study did not contain any personal information, and each phone user in the data set was assigned a user ID. Among the 631,645 phone users, we selected users with at least 20 position records per day during five working days. Then, the nearest-neighbour interpolation method (Hoteit et al. 2014; Liu et al. 2018) was utilised to fill in the missing locations of these selected users' trajectories at hourly intervals, which is a commonly used trajectory interpolation method. By using this method, a missing record can be interpolated by the value of its nearest sampling position in time. Finally, the oscillations (i.e. ping-pong effect) of mobile phone positioning data were approached using the point-clustering method proposed by Xu et al. (2020). After

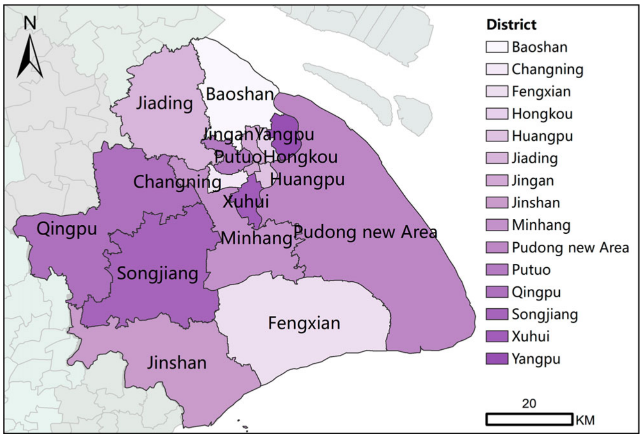

Figure 3. Experiment area.

Table 2. Example of mobile phone positioning data.

User ID	Date	Time	Longitude	Latitude
User 1	2015/12/28	1:00	120.***	30.***
User 1	2015/12/28	2:00	120.***	30.***
…	…	…	…	…
User 1	2015/12/28	24:00	120.***	30.***

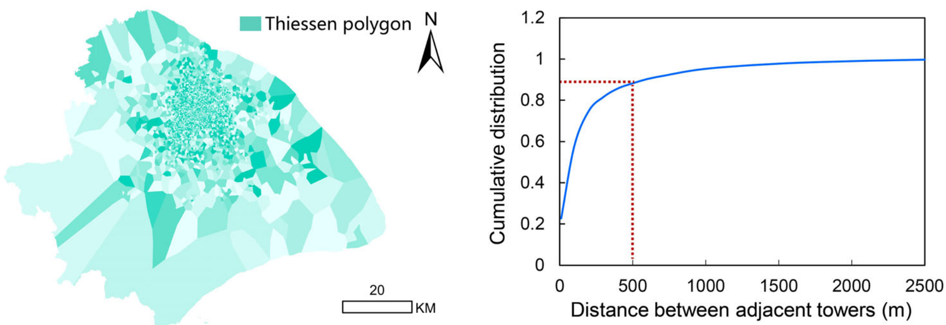

Figure 4. (a) Spatial distribution of mobile phone towers in the experimental area. (b) Cumulative distribution of the distance between two adjacent mobile phone towers.

data pre-processing, we obtained 4,145 mobile phone users with good data quality and used them as the samples for this study. An example of mobile phone position records of an individual in a day is shown in Table 2.

Figure 4(a) shows the spatial distribution of the mobile phone towers in this study. Each tower is represented by a Thiessen polygon to denote its service area. The distance between any two adjacent mobile phone towers was calculated, and the distance distribution is shown in Figure 4(b). It can be found from the curve of Figure 4(b) that approximately 88% of the distance between two adjacent mobile phone towers was less than 500 m. We divided the study area into 500 m by 500 m grids for simplicity and to speed up the calculations. Ultimately, there were 22,098 grid cells (5,603.3 km^2) in the experimental area.

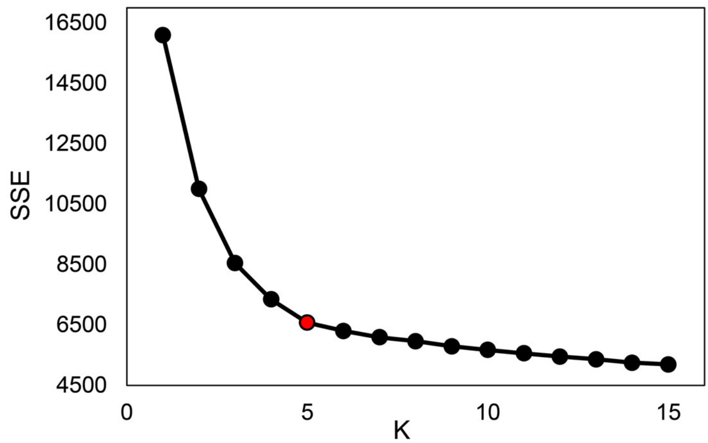

Figure 5. The SSEs with different values of *k*.

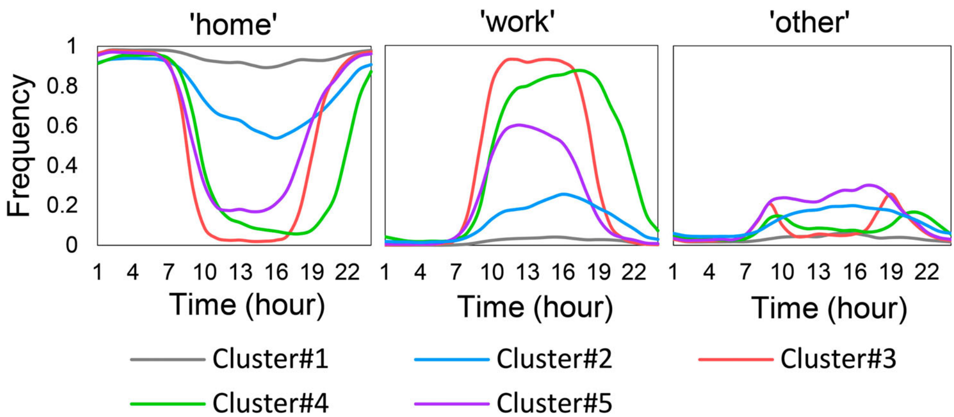

Figure 6. Activity frequency profiles of the five different user clusters.

4. Experiments and results

4.1. User clustering results

For each user u, the activity frequency vector $AF(u)$ was calculated according to the formulas described in Section 2.1. Then, the k-means clustering method was used to classify vector $AF(u)$ of all users. The sum of squared error (SSE) was used to determine the best k. As shown in Figure 5, the SSE gradually decreased as k increased from 2 to 15. According to the elbow rule, when the clustering number was 5, it provided a more stable clustering result and a richer partition of individuals' daily activity patterns.

The curves of the cluster centre of five clusters are shown in Figure 6. The properties of users of each cluster are described as follows:

Family person: Users in Cluster #1 have a very high frequency ($0.89 \sim 0.98$) of 'home' activity all day long, which means they spend most of their time at home. This cluster is large and includes 43.5% of all samples.

Active family person: For users in Cluster #2, the frequency of 'home' activity is also high (approximately 0.6), and the frequency of 'work' and 'other' activities is 0.2 for each. These users often stay at home and sometimes work or perform other activities during the day. The proportion of users in Cluster #2 is 14.3%.

Regular worker: Users in Cluster #3 have a regular schedule. They often leave home at approximately 8:00 and return home at approximately 18:00. During the day, they are always at work. The users in this cluster share 19.5% of all samples.

Afternoon worker: Users in Cluster #4 also have a regular schedule. The difference is that users in Cluster #4 start work approximately two hours later than those in Cluster #4. The major working time for them is in the afternoon. The proportion of users in Cluster #4 is 10.4%.

Part-time worker: Users in Cluster #5 are 'part-time workers' because they do not always work. The frequency of 'work' activity is only approximately 0.6 during the daytime, while the frequency of 'other' activity is 0.2, which is the highest frequency among all five clusters. The users in Cluster #5 account for 12.3% of all samples.

Next, the activity preferences of the abovementioned five clusters of users were calculated and compared. The hourly transition probabilities between three types of activities (i.e. 'home', 'work', and 'other') were calculated for each group. In Figure 7, the transition probability matrix per hour is visualised by a 3×3 square, in which each row represents 'home', 'work' and 'other' activities in turn, as does each column. The transition probabilities between the three types of activities are rendered by using gradient colours.

As expected, the activity preferences of the five groups of users are different. The transition matrix of 'family person' changes little in 24 h, and the probability of 'home' to 'home' is high all the time. The transition matrix of 'active family person' has a similar pattern to that of 'family person', except for higher frequencies of 'work' and 'other' activities. For 'regular worker' and 'afternoon worker', the probabilities of 'home' to 'work', 'work' to 'work', and 'other' to 'work' are high during the daytime (9:00–17:00 for 'regular worker' and 11:00–20:00 for 'afternoon worker'). Moreover, the probabilities of 'work' to 'home' and 'other' to 'home' increase from 18:00 for 'regular worker' and three hours later for 'afternoon worker'. For 'part-time worker', the probabilities of 'home' to 'work', 'work' to 'work', and 'other' to 'work' in the daytime are not as high as those of 'regular worker' and 'afternoon worker'. Apart from 'work', some users in 'part-time worker' also participate in 'other' activities. The activity preferences of different groups of people can help us understand their travel motivations, making the prediction of the user's next location more accurate.

4.2. Prediction results evaluation and analysis

4.2.1. Evaluation metric and baseline models

We first introduce the performance metric and baseline models for evaluating the efficiency of the proposed model. The average location prediction accuracy for the next time t is defined by the ratio of the number of users whose locations are correctly predicted to the number of all users. The formula is as follows:

$$accuracy(t) = \frac{N_{correct}(t)}{N_{all}(t)} \tag{10}$$

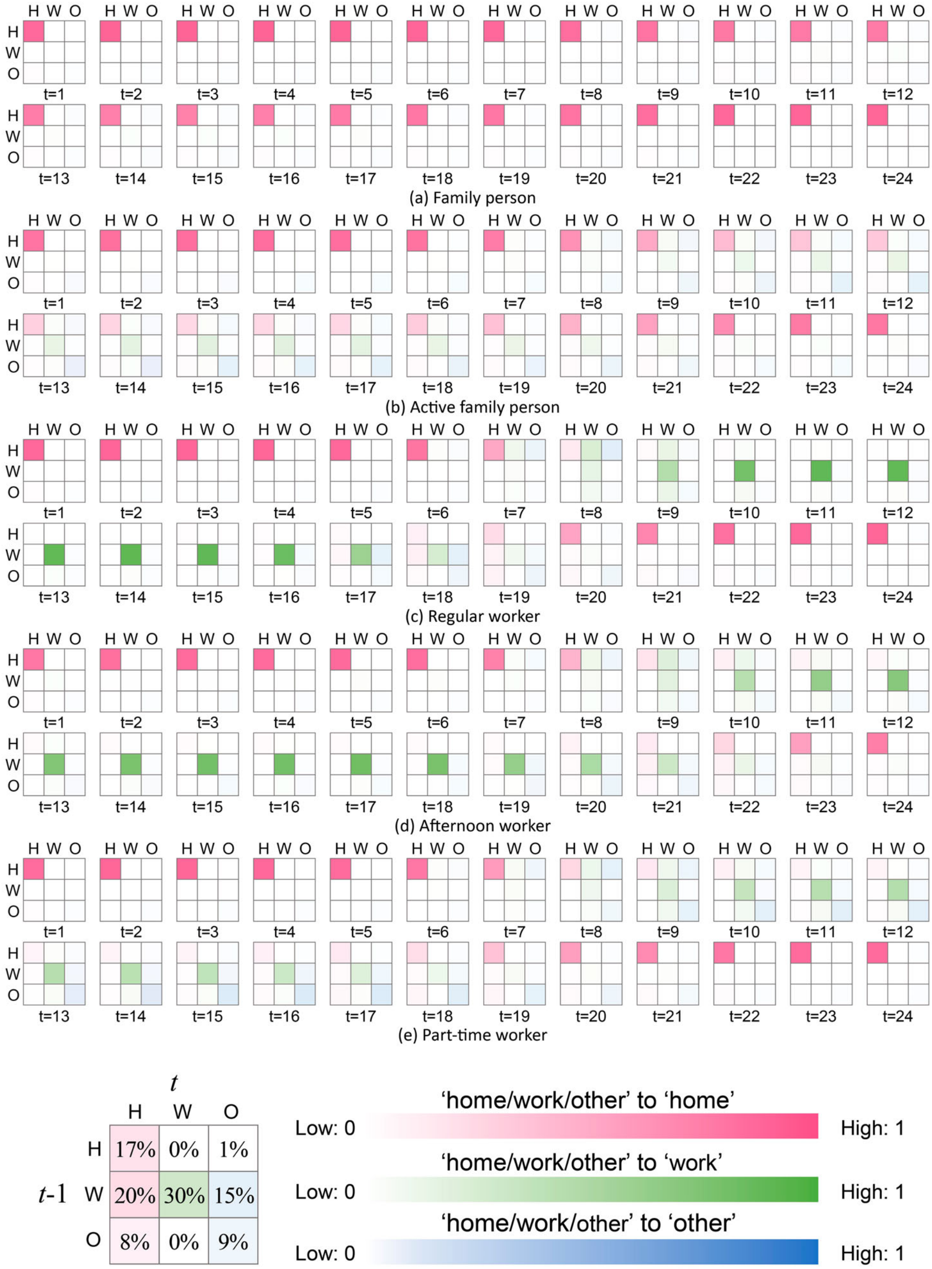

Figure 7. The activity preferences of different groups of people.

where $N_{correct}(t)$ is the number of users whose locations are correctly predicted, and $N_{all}(t)$ is the number of all users.

We used 70% of each group's samples as the training dataset and the rest as the testing set. Since mobility behaviours vary among the five groups of users, the values of α, β, and γ differed in each group's prediction models. An iterative procedure was utilised to determine the optimal values of these three parameters. From the training dataset, the data of the first m-1 days were chosen to build the prediction models with varying α, β, and γ values, and the data of the m^{th} day were utilised to validate the corresponding models. For each group, α, β, and γ were set to 0 in the beginning, and then a triple loop was used to increment their values by 0.1. At each iteration, the average prediction accuracy of users in this group was calculated. Subsequently, the combination of α, β, and γ with the highest accuracy was selected.

Our model was compared with the following three baseline models:

(a) Markov model (MM): This is the first component of the proposed model, in which the trajectory of each user is modelled as a 1-order Markov chain when conducting next-place prediction (Gambs, Killijian, and Del Prado Cortez 2012).
(b) Long short-term memory model (LSTM): This model is a recurrent neural network architecture used in the field of deep learning (Sutskever, Martens, and Hinton 2011). In this model, the trajectory of each user is modelled as a time sequence, and the long-range dependencies are considered.
(c) Mobility prediction based on historical user trajectories and collective location preferences of all users (MUCA): This model is similar to the proposed MUCS model, except for the collective location preferences extracted from all users.

4.2.2. Comparison with baseline models at different times of the day

Figure 8 shows the prediction accuracy of the four models at different times of the day. The evening hours (23:00–6:00) were not included because the four models' prediction accuracies were similar and are all over 90% during this period. As shown in Figure 8, the MUCS achieved a prediction accuracy of over 80% during most of the day. The prediction accuracy of the four models varied at different times, but a general trend can be observed: the accuracy decreased gradually from 7:00 to 9:00, then rose from 9:00 to 15:00, then decreased again from 15:00 to 18:00, and finally rose again after 18:00. The two lowest points on the three models' curves were at 9:00 and 18:00, exactly the most active periods of daily human movements. Both MUCA and MUCS performed better than MM and LSTM, which indicates that it is effective to consider collective preferences. LSTM was slightly better than MM because LSTM considers the long-range dependencies of individuals' location sequences. However, as an advanced machine-learning method, LSTM was not very effective due to insufficient training data. Moreover, the MUCS achieved a 1% to 3% improvement compared with the MUCA when people were highly active, e.g. 9:00–10:00 and 16:00–20:00. The maximum improvement of the MUCS over the MUCA occurred at 17:00 when the prediction accuracy of the MUCS model was 84.3% and that of the MUCA model was 81.3%. We believe this is the result of considering the distinct daily activity patterns of users.

Next, we further compared the performance of the four models on five groups of users with distinct daily activity patterns. As shown in Figure 9(a)–(e), MUCS performed the best in five groups of users, followed by MUCA, while the worst was MM and LSTM. For the 'family person' group, the curves of prediction accuracy of the four models were stable, and the accuracy was high throughout the day, in the range of 90% to 100% (Figure 9 (a)). Figure 9(b) shows the prediction results of the 'active family person' group. The prediction accuracy was as high as that of the 'family person' group at night but decreased to approximately 75% during the daytime. Due to the highly stationary nature of users' trajectories in groups

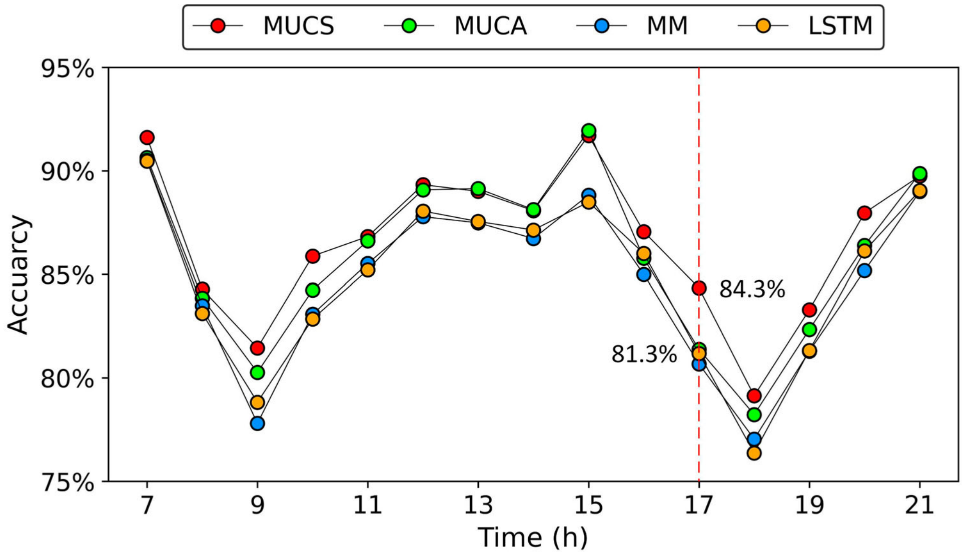

Figure 8. Prediction accuracy in the testing set from 7:00 to 21:00.

of 'family person' and 'active family person', the prediction accuracies of MM and LSTM were significantly high, and the improvements of MUCS and MUCA to MM and LSTM were relatively small. Figure 9(c)–(e) shows the prediction results of the 'regular worker', 'afternoon worker', and 'part-time worker' groups, of which the curves show similar 'W' patterns. The prediction accuracy was high when people stayed at home or work and decreased when they were travelling or participating in other activities. For these three groups, the improvement of MUCS was evident, and the maximum accuracy increases in MUCA, MM, and LSTM were up to 6%, 16%, and 10%, respectively.

We also report the performance of the MUCS and three baseline models at 9:00, 13:00, 17:00, and 21:00 in Table 3. It is clearly shown in Table 3 that MUCS outperformed the three baseline models in most cases. The largest improvement occurred at 9:00 for the 'regular worker' group. At 9:00, the prediction accuracies of MUCS, MUCA, MM, and LSTM were 72.76%, 66.77%, 56.79%, and 65.02%, respectively. Compared with MM, MUCS increased the prediction accuracy by 16%, which was the maximum accuracy increase in the MUCS of the three baseline models.

4.2.3. Comparative performance on different historical sequence lengths

The performance of the prediction model may vary with the length of the historical trajectory of users. Therefore, three models' prediction performances (MUCS, MUCA, and MM) with different lengths of user historical sequences are discussed in this section. We randomly sampled 50% of users in each group and tested the prediction accuracy of these users by assuming that each user has different lengths of trajectories ranging from 6 to 96. The LSTM model was not included for comparison because it is very demanding on the length of training trajectories. In most cases, the lengths of the trajectories (from 6 to 96) were not sufficient for the LSTM model. Therefore, comparing with the LSTM model here does not make sense.

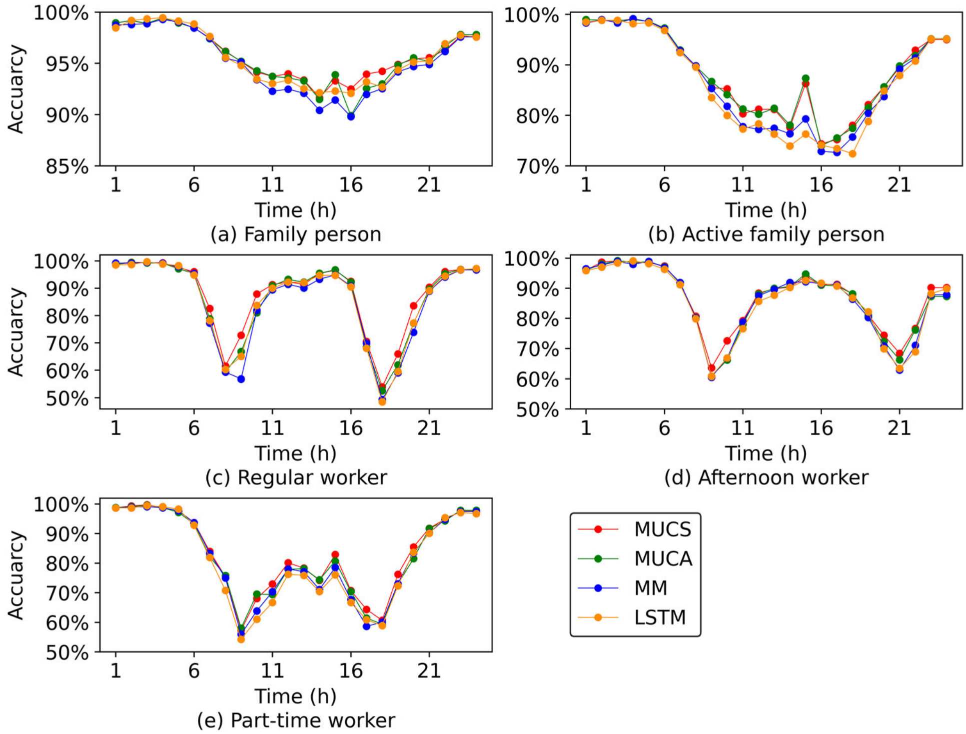

Figure 9. Prediction accuracy of five groups of users at different times of the day.

Table 3. Prediction accuracy of the four models on five groups of users.

		Accuracy (%)			
Groups	Models	9:00	13:00	17:00	21:00
Family person	MUCS	95.17	**93.37**	**93.94**	**95.55**
	MUCA	**95.19**	93.28	92.54	95.17
	MM	95.11	92.08	91.97	94.88
	LSTM	94.79	92.52	93.18	95.23
Active family person	MUCS	85.49	81.17	75.18	89.50
	MUCA	**86.69**	**81.38**	**75.50**	**89.78**
	MM	85.31	77.45	72.63	89.17
	LSTM	83.50	76.26	73.40	87.88
Regular worker	MUCS	**72.76**	91.39	**70.47**	**90.31**
	MUCA	66.77	**92.18**	68.48	89.61
	MM	56.79	90.06	69.84	88.95
	LSTM	65.02	92.09	67.99	89.12
Afternoon worker	MUCS	**63.52**	89.78	**91.29**	**68.33**
	MUCA	60.78	**89.92**	90.77	66.19
	MM	60.47	89.59	91.03	62.79
	LSTM	60.79	87.70	90.72	63.34
Part-time worker	MUCS	57.31	**78.29**	**64.30**	**91.76**
	MUCA	**58.00**	78.10	61.38	91.65
	MM	55.94	77.09	58.63	90.17
	LSTM	54.19	75.83	60.82	90.06

As shown in Figure 10, the two models (i.e. MUCS and MUCA) based on both the user's historical trajectories and the collective location preferences performed more robustly than MM. Especially in the early stage when the length of the user's historical sequence was very short, the improvement by MUCS and MUCA was evident. It is noted that MUCS performed

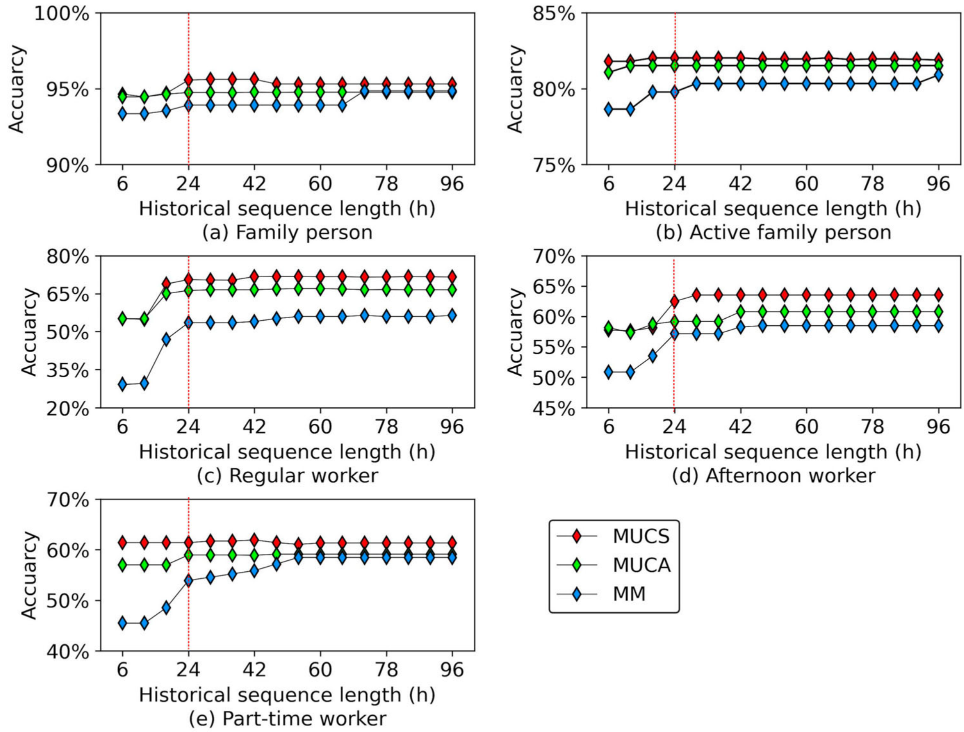

Figure 10. Prediction accuracy of five groups of users with different lengths of historical trajectory.

better than MUCA for all five groups due to exploiting the collective location preferences of users with similar daily activity patterns. Based only on the user's historical trajectories, MM encounters the cold start problem in the early stage (sequence length less than 24 h). The prediction accuracy rose greatly with the increasing length of the user's historical trajectory.

For the 'family person' group (Figure 10(a)), the prediction accuracy of all three models was high even in the early stage, and it reached more than 93%. The limited historical trajectory had little impact on the next location prediction in this group. Similarly, the prediction accuracy of the MUCS and MUCA was also high (more than 80%) for the users of the 'active family person' group in the early stage. This means that even with the limited historical trajectory length, the 'active family person' group's prediction accuracy can be guaranteed by considering collective location preferences. Compared with 'family person' and 'active family person', the users in the 'regular worker', 'afternoon worker', and 'part-time worker' groups visited many more locations. Hence, their prediction accuracy was lower in the early stage, and the prediction accuracy improvements of MUCS, MUCA, and MM were more evident.

4.2.4. Analysis of parameters of MUCS

The optimal values of parameters α, β, and γ of the MUCS may vary at different times of the day. Therefore, the optimal values of these three parameters from 1:00 to 24:00 are calculated to find the best parameter setting of MUCS for each prediction time. In the experiment, we found that the model obtained the highest prediction accuracy in more than one

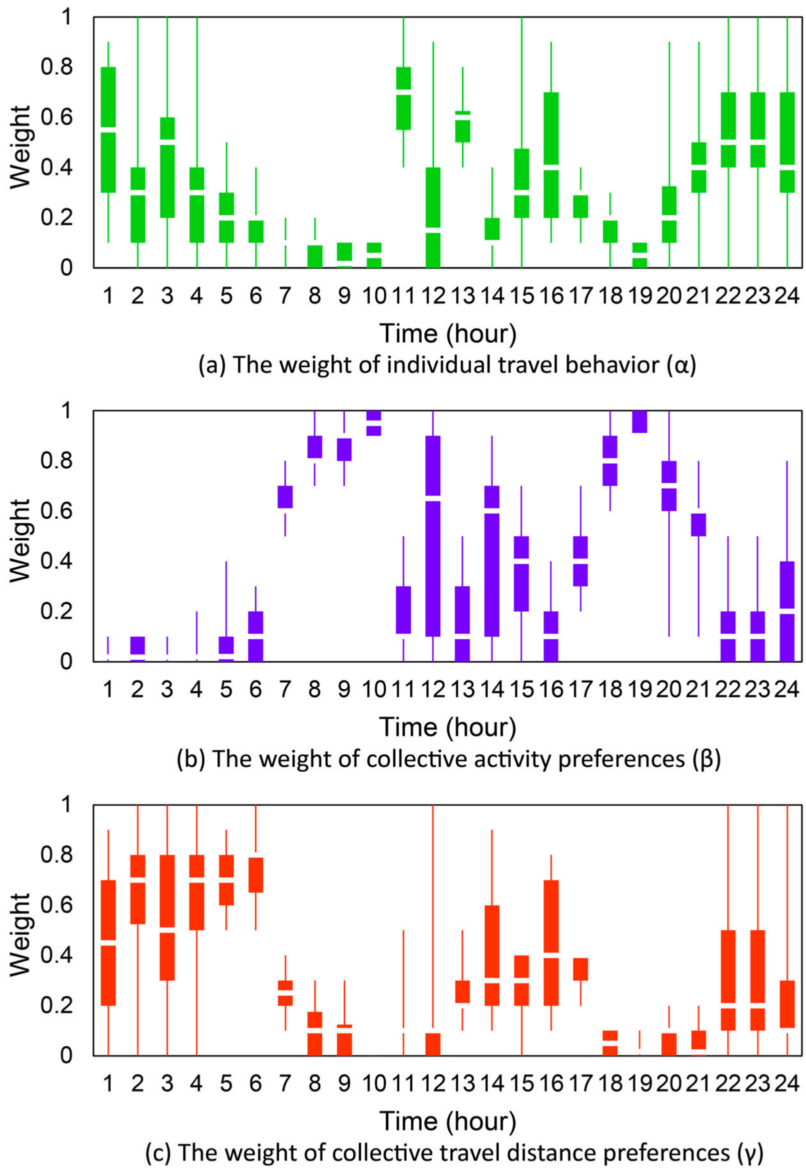

Figure 11. Variations of α, β and γ from 1:00 to 24:00.

combination of parameters. Therefore, we used a boxplot to show the distributions of α, β, and γ. Figure 11 shows the distributions of α, β, and γ from 1:00 to 24:00 of the 'regular worker' group (see Appendix 1 for distributions for 'family person', 'active family person', 'afternoon worker', and 'part-time worker').

As shown in Figure 11, the median values of α, β, and γ show distinct variation patterns. These parameters vary among users in different clusters and vary with time, which helps us understand how these behavioural factors impact an individual's next location prediction at different times. By combining the analysis in Figure 6, it was found that the median values of α and γ showed negative correlations with the 'other' activity profile of 'regular worker'. The median values of α and γ were relatively high when users stayed at home at night (22:00–6:00) or workplaces (12:00–17:00), which indicates that individual travel behaviour and collective travel distance preference are important for predicting the locations of the 'regular worker' group at these times. During these two time periods, the probability of users exploring new locations was low; thus, higher α values can help predict their next locations by giving higher probabilities to grid cells they visited before. Similarly, users were also less likely to travel very long distances at these times, and higher γ values achieved higher probabilities on grid cells close to their current locations. The median value of β had a bimodal variation pattern and was positively correlated with the 'other' activity profile of 'regular worker'. The value of β was low when people stayed at home at night (22:00–6:00) but high when they commuted or participated in 'other' activities (7:00–10:00 and 18:00–21:00), which suggests that the activity preference is important for location prediction in the 'regular worker' group in two periods (7:00–10:00 and 18:00–21:00). Higher values of β indicated that the prediction model achieved higher probabilities to grid cells where users' preferred activities took place previously.

In Appendix 1, we can also observe that the value of β usually increased when users' locations changed and decreased when users stayed at home or workplaces. In contrast, the variation patterns of α and γ were opposite to that of β. Among the five groups of users, 'active family person' and 'part-time worker' had the most similar variation patterns of α, β, and γ. These two groups of users both had a relatively high proportion of 'other' activities during the daytime; therefore, the values of γ remained at a low level.

5. Discussion

In this study, the missing locations of 4,145 selected mobile phone users were estimated using the nearest-neighbour interpolation method. To evaluate the impacts of the missing location estimation accuracy on the performance of MUCS, two more experiments were implemented.

First, we compared the missing location estimation accuracy of the nearest-neighbour interpolation method with the other two interpolation methods, i.e. linear interpolation and gradient boosting decision tree (GBDT) (Li et al. 2019). Linear interpolation uses spatial–temporal correlations among data to interpolate the missing records, whereas GBDT is a machine-learning approach that uses some relevant features to predict missing records. In this study, the features of GBDT were the same as those in Li et al. (2019). The mean absolute error (MAE) was used to evaluate the performance of these interpolation methods. After interpolation, the MAE values of the nearest-neighbour interpolation, linear interpolation and GBDT methods were 450, 440, and 700 m, respectively. The nearest-neighbour interpolation and linear interpolation methods performed better than GBDT. This was because the missing ratio of our mobile phone positioning data of 4,145 users was low (only 11%), and few consecutive records were missing. Thus, simple interpolation methods, such as nearest-neighbour interpolation and linear interpolation, actually obtained better results than machine-learning-based methods.

Then, the performance of MUCS on trajectory data sets interpolated by these three methods was analysed. The average prediction accuracy of all mobile phone users and the maximum differences among the three datasets are reported in Table 4. As shown in

Table 4. Performance of MUCS on three interpolated trajectory datasets

Dataset	Accuracy (%) 5:00	9:00	13:00	17:00	21:00
Nearest_data	**98.40**	**81.45**	**89.00**	**84.34**	**90.37**
Linear_data	98.29	80.53	88.64	**81.60**	90.04
GBDT_data	**98.23**	**79.98**	**87.21**	82.05	**89.82**
Maximum differences	0.17	1.47	1.79	2.74	0.55

Table 5. Parameter settings in three scenarios.

Scenario	User's travel behaviour	Collective preferences Activity type	Travel distance
MUCS ($\beta = 0$)	Y	N	Y
MUCS ($\gamma = 0$)	Y	Y	N
MUCS	Y	Y	Y

Table 4, the maximum differences among the three datasets at five different times (5:00, 9:00, 13:00, 17:00, and 21:00) were less than 3%. MUCS performed better on Nearest_data (the trajectory data reconstructed by the nearest-neighbour interpolation method) than Linear_data, although the MAE value of the former is a slightly larger than that of the latter. This is because the linear interpolation method generates new locations that are difficult to predict accurately. In contrast, nearest-neighbour interpolation uses the nearest sampling position in time and does not generate new locations. The proposed approach performed the worst on GBDT_data because its MAE was the largest.

We also discuss is the impacts of the activity estimation accuracy on the MUCS performance. Since activity type information is unavailable in mobile phone trajectories, we identified activities using spatial and temporal constraints. Two types of highly regular activities (i.e. 'home' and 'work') were identified in this study. In general, residents usually stayed home at night and worked during the daytime, so the identification of 'home' and 'work' was much easier and more accurate than other types of activities, such as 'shopping' and 'eating out'. In Yin et al. (2016), the identification accuracies of 'home' and 'work' by space and time constraints were both up to 99%. In Liu et al. (2013), the identification accuracies of 'home' and 'work' were also more than 80%. We did not validate our activity identification results because of the difficulty in acquiring ground truth human activity data. According to the high accuracy obtained by the abovementioned studies, we think the estimation of highly regular activities, such as 'home' and 'work', has a limited impact on subsequent location prediction.

Finally, we assessed the impacts of collective location preferences on model performance. The collective location preferences in this study consisted of two parts: activity type preference and travel distance preference. We designed three scenarios for a performance comparison (Table 5).

It is clearly shown in Figure 12 that the decrease in prediction accuracy is apparent for 'family person', 'regular worker', and 'afternoon worker' when β is 0. Take 'regular worker' as an example. The decline was concentrated when users' activities and locations changed, such as 9:00 and 10:00 in the morning and 19:00 and 20:00. At 9:00 and 10:00, the primary activity transitions are 'work' to 'work', 'home' to 'work', and 'other' to 'work' (see Figure 7), so 'work' is a dominant activity. The activity preference for 'work' plays a vital

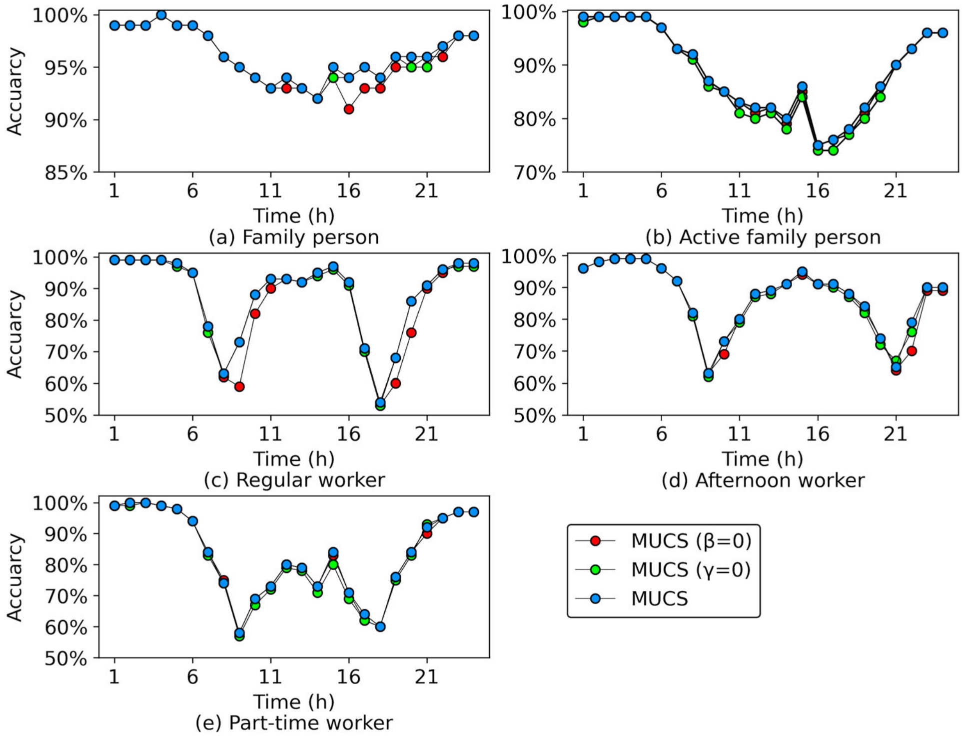

Figure 12. Prediction accuracy of five groups of users in three scenarios.

role in next location prediction because it gives higher probabilities to grid cells where 'work' took place before. Similarly, at 19:00 and 20:00, users' activities shifted from 'work' and 'other' to 'home', and 'home' was a dominant activity. For 'active family person' and 'part-time worker', the decrease in prediction accuracy is apparent when γ is 0 because the frequencies of 'other' activities are relatively high in the daytime, and the activity transitions are diverse. Thus, the improvement in prediction accuracy mainly depends on travel distance preferences. With the help of activity and travel distance preferences extracted from users with similar daily activity patterns, the prediction model can capture hidden location preferences, and the prediction accuracy can be improved.

6. Conclusions

In this paper, we proposed a human mobility prediction model by combining individual travel behaviour with the location preferences of users who share similar daily activity patterns. First, users were clustered into groups with distinct daily activity patterns. Second, each group's collective location preferences (i.e. activity and travel distance preferences) were extracted. Then, the user's travel behaviour and collective location preferences were integrated to predict their next location. A mobile phone positioning dataset of users in Shanghai, China, was used to validate the proposed model. The results showed the following conclusions. (1) The prediction accuracy concerning a user's next locations is closely related to their daily activity pattern. For example, the location prediction accuracy of a 'family person' is much higher than that of users in other groups. (2) The proposed model achieves a prediction accuracy of over 80% during most of the day. There is a maximum increase of 16% in the prediction accuracy compared with the three baseline models when users' mobility is highly active, demonstrating the effectiveness of the proposed model. (3) The importance of individual travel behaviour, collective activity preferences, and collective travel distance preferences varies at different times of the day. Our model can contribute to a more in-depth understanding of the relationship between a user's travel behaviour and collective location preferences in next location prediction.

The proposed model experiment was based on a 5-day mobile phone positioning dataset. Although users continued to explore new locations on other days, our study showed that several major daily activity patterns can be captured even with a 5-day dataset, especially given that we only focused on the patterns of 'home', 'work', and 'other' activities. In addition, our model showed good performance based only on a 5-day mobile phone positioning dataset. We believe a longer dataset may be even more helpful.

We plan to improve our model in several directions. First, in our current work, only 'home', 'work', and 'other' activities were considered in user clustering and collective activity preference extraction owing to the inadequate spatio-temporal resolution of mobile phone positioning data. If higher spatio-temporal resolution data are available in the future, we can use more contextual information such as POI and land use, to differentiate various types of 'other' activity, such as 'shopping' and 'leisure', and extend our prediction model. Second, the proposed model is primarily for mobile phone users with relatively high data quality (for example, at least 20 position records per day). We believe proposing a model that can predict locations for a large proportion of mobile phone users with heterogeneous data quality is worth researching in the future. Third, we will validate the performance of the proposed model using datasets from different cities and different sources.

Disclosure statement

No potential conflict of interest was reported by the authors .

Funding

This work was supported by National Natural Science Foundation of China [grant number 41971345, 71961137003]; Guangdong Basic and Applied Basic Research Foundation [grant number 2020A1515010695].

ORCID

Yang Xu http://orcid.org/0000-0003-3898-022X

References

Alexander, L., S. Jiang, M. Murga, and M. C. González. 2015. "Origin-destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data." *Transportation Research Part C: Emerging Technologies* 58: 240–250.

Baumann, P., W. Kleiminger, and S. Santini. 2013. "The Influence of Temporal and Spatial Features on the Performance of Next-Place Prediction Algorithms." Proceedings of the 2013 ACM International Joint Conference on Pervasive and ubiquitous computing, Zurich, Switzerland, 449–458.

Baumann, P., C. Koehler, A. K. Dey, and S. Santini. 2018. "Selecting Individual and Population Models for Predicting Human Mobility." *IEEE Transactions on Mobile Computing* 17 (10): 2408–2422.

Calabrese, F., G. D. Lorenzo, and C. Ratti. 2010. "Human Mobility Prediction Based on Individual and Collective Geographical Preferences." 13th International IEEE Conference on Intelligent Transportation Systems, 312–317.

Cao, J., W. Tu, Q. Li, and R. Cao. 2017. "Spatiotemporal Analysis of Aggregated Human Activities Based on Massive Mobile Phone Tracking Data." *Journal of Geo-Information Science* 19 (4): 467–474.

Chen, H., Y. Fan, J. Jiang, and X. Chen. 2018. "Mobility Prediction Based on POI-Clustered Data." In *International Conference on Machine Learning and Intelligent Communications*, edited by Limin Meng and Yan Zhang, 60–72. Cham: Springer .

Chen, P., W. Shi, X. Zhou, Z. Liu, and X. Fu. 2019. "STLP-GSM: A Method to Predict Future Locations of Individuals Based on Geotagged Social Media Data." *International Journal of Geographical Information Science* 33 (12): 2337–2362.

Choi, S., H. Yeo, and J. Kim. 2018. "Network-Wide Vehicle Trajectory Prediction in Urban Traffic Networks Using Deep Learning." *Transportation Research Record: Journal of the Transportation Research Board* 2672 (45): 173–184.

Cuttone, A., S. Lehmann, and M. C. González. 2018. "Understanding Predictability and Exploration in Human Mobility." *EPJ Data Science* 7 (1): 1–17.

De Domenico, M., A. Lima, and M. Musolesi. 2013. "Interdependence and Predictability of Human Mobility and Social Interactions." *Pervasive and Mobile Computing* 9 (6): 798–807.

Gambs, S., M. O. Killijian, and M. N. Del Prado Cortez. 2012. "Next Place Prediction Using Mobility Markov Chains." Proceedings of the First Workshop on Measurement, Privacy, and Mobility, Bern, Switzerland, 1–6.

Hawelka, B., I. Sitko, P. Kazakopoulos, and E. Beinat. 2017. "Collective Prediction of Individual Mobility Traces for Users with Short Data History." *PLoS One* 12 (1): e0170907.

Hoteit, S., S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle. 2014. "Estimating Human Trajectories and Hotspots Through Mobile Phone Data." *Computer Networks* 64: 296–307.

Huang, Q. 2017. "Mining Online Footprints to Predict User's Next Location." *International Journal of Geographical Information Science* 31 (3): 523–541.

Huang, Z., X. Ling, P. Wang, F. Zhang, Y. Mao, T. Lin, and F. Y. Wang. 2018. "Modeling Real-Time Human Mobility Based on Mobile Phone and Transportation Data Fusion." *Transportation Research Part C: Emerging Technologies* 96: 251–269.

Jiang, S., J. Joseph Ferreira, and M. C. Gonzalez. 2012. "Discovering Urban Spatial-Temporal Structure from Human Activity Patterns." Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 95–102.

Li, M., S. Gao, F. Lu, and H. Zhang. 2019. "Reconstruction of Human Movement Trajectories from Large-Scale low-Frequency Mobile Phone Data." *Computers, Environment and Urban Systems* 77: 101346.

Li, F., Z. Gui, Z. Zhang, D. Peng, S. Tian, K. Yuan, and Y. Lei. 2020. "A Hierarchical Temporal Attention-Based LSTM Encoder-Decoder Model for Individual Mobility Prediction." *Neurocomputing* 403: 153–166.

Li, X., D. Lian, X. Xie, and G. Sun. 2015. "Lifting the Predictability of Human Mobility on Activity Trajectories." 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 1063–1069.

Li, M., F. Lu, H. Zhang, and J. Chen. 2020. "Predicting Future Locations of Moving Objects with Deep Fuzzy-LSTM Networks." *Transportmetrica A: Transport Science* 16 (1): 119–136.

Liu, F., D. Janssens, J. Cui, Y. Wang, G. Wets, and M. Cools. 2014. "Building a Validation Measure for Activity-Based Transportation Models Based on Mobile Phone Data." *Expert Systems with Applications* 41 (14): 6174–6189.

Liu, F., D. Janssens, J. Cui, G. Wets, and M. Cools. 2015. "Characterizing Activity Sequences Using Profile Hidden Markov Models." *Expert Systems with Applications* 42 (13): 5705–5722.

Liu, F., D. Janssens, G. Wets, and M. Cools. 2013. "Annotating Mobile Phone Location Data with Activity Purposes Using Machine Learning Algorithms." *Expert Systems with Applications* 40 (8): 3299–3311.

Liu, Z., T. Ma, Y. Du, T. Pei, J. Yi, and H. Peng. 2018. "Mapping Hourly Dynamics of Urban Population Using Trajectories Reconstructed from Mobile Phone Records." *Transactions in GIS* 22 (2): 494–513.

Lv, Q., Y. Qiao, N. Ansari, J. Liu, and J. Yang. 2017. "Big Data Driven Hidden Markov Model Based Individual Mobility Prediction at Points of Interest." *IEEE Transactions on Vehicular Technology* 66 (6): 5204–5216.

Mckercher, B., and A. A. Lew. 2003. "Distance Decay and the Impact of Effective Tourism Exclusion Zones on International Travel Flows." *Journal of Travel Research* 42 (2): 159–165.
Qiao, Y., Z. Si, Y. Zhang, F. B. Abdesslem, X. Zhang, and J. Yang. 2018. "A Hybrid Markov-Based Model for Human Mobility Prediction." *Neurocomputing* 278: 99–109.
Shaw, S.-L., and D. Sui. 2018. "Editorial: GIScience for Human Dynamics Research in a Changing World." *Transactions in GIS* 22 (4): 891–899.
Simini, F., M. C. González, A. Maritan, and A. L. Barabási. 2012. "A Universal Model for Mobility and Migration Patterns." *Nature* 484 (7392): 96–100.
Sutskever, I., J. Martens, and G. E. Hinton. 2011. "Generating Text with Recurrent Neural Networks." International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28–July, 1017–1024.
Tu, W., J. Cao, Y. Yue, S. L. Shaw, M. Zhou, Z. Wang, X. Chang, Y. Xu, and Q. Li. 2017. "Coupling Mobile Phone and Social Media Data: a new Approach to Understanding Urban Functions and Diurnal Patterns." *International Journal of Geographical Information Science* 31 (12): 2331–2358.
Wang, H., S. Zeng, Y. Li, P. Zhang, and D. Jin. 2020. "Human Mobility Prediction Using Sparse Trajectory Data." *IEEE Transactions on Vehicular Technology* 69 (9): 10155–10166.
Wen, T.-H., C.-S. Hsu, and M.-C. Hu. 2018. "Measurement of the Potential Geographic Accessibility from Call to Definitive Care for Patient with Acute Stroke." *International Journal of Health Geographics* 17 (1): 1–15.
Xu, Y., X. Li, S. L. Shaw, F. Lu, L. Yin, and B. Y. Chen. 2020. "Effects of Data Preprocessing Methods on Addressing Location Uncertainty in Mobile Signaling Data." *Annals of the American Association of Geographers* 111 (2): 515–539.
Yan, X. Y. 2011. "Empirical Statistics on Individual Human Travel Behaviour." *Journal of University of Electronic Science and Technology of China* 40 (2): 168–173.
Yan, X. Y., W. X. Wang, Z. Y. Gao, and Y. C. Lai. 2017. "In Situ Click Chemistry Generation of Cyclooxygenase-2 Inhibitors." *Nature Communications* 8 (1): 1–9.
Yang, C., F. Yan, and S. V. Ukkusuri. 2018. "Unraveling Traveler Mobility Patterns and Predicting User Behavior in the Shenzhen Metro System." *Transportmetrica A: Transport Science* 14 (7): 576–597.
Yin, M., M. Sheehan, S. Feygin, J. F. Paiement, and A. Pozdnoukhov. 2016. "A Generative Model of Urban Activities from Cellular Data." *IEEE Transactions on Intelligent Transportation Systems* 19 (6): 1682–1696.
Yu, C., Y. Liu, D. Yao, L. T. Yang, H. Jin, H. Chen, and Q. Ding. 2015. "Modeling User Activity Patterns for Next-Place Prediction." *IEEE Systems Journal* 11 (2): 1060–1071.
Yuan, M. 2018. "Human Dynamics in Space and Time: A Brief History and a View Forward." *Transactions in GIS* 22 (4): 900–912.
Yuan, J., Y. Zheng, and X. Xie. 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*, edited by Qiang Yang, Deepak Agarwal, and Jian Pei, 186–194. Beijing, China: ACM .
Zhao, Z., H. N. Koutsopoulos, and J. Zhao. 2018. "Individual Mobility Prediction Using Transit Smart Card Data." *Transportation Research Part C: Emerging Technologies* 89: 19–34.
Zheng, Z., and S. Zhou. 2017. "Scaling Laws of Spatial Visitation Frequency: Applications for Trip Frequency Prediction." *Computers, Environment and Urban Systems* 64: 332–343.
Zipf, G. K. 1946. "The P_1 P_2/D Hypothesis: On the Intercity Movement of Persons." *American Sociological Review* 11 (6): 677–686.

Appendix 1

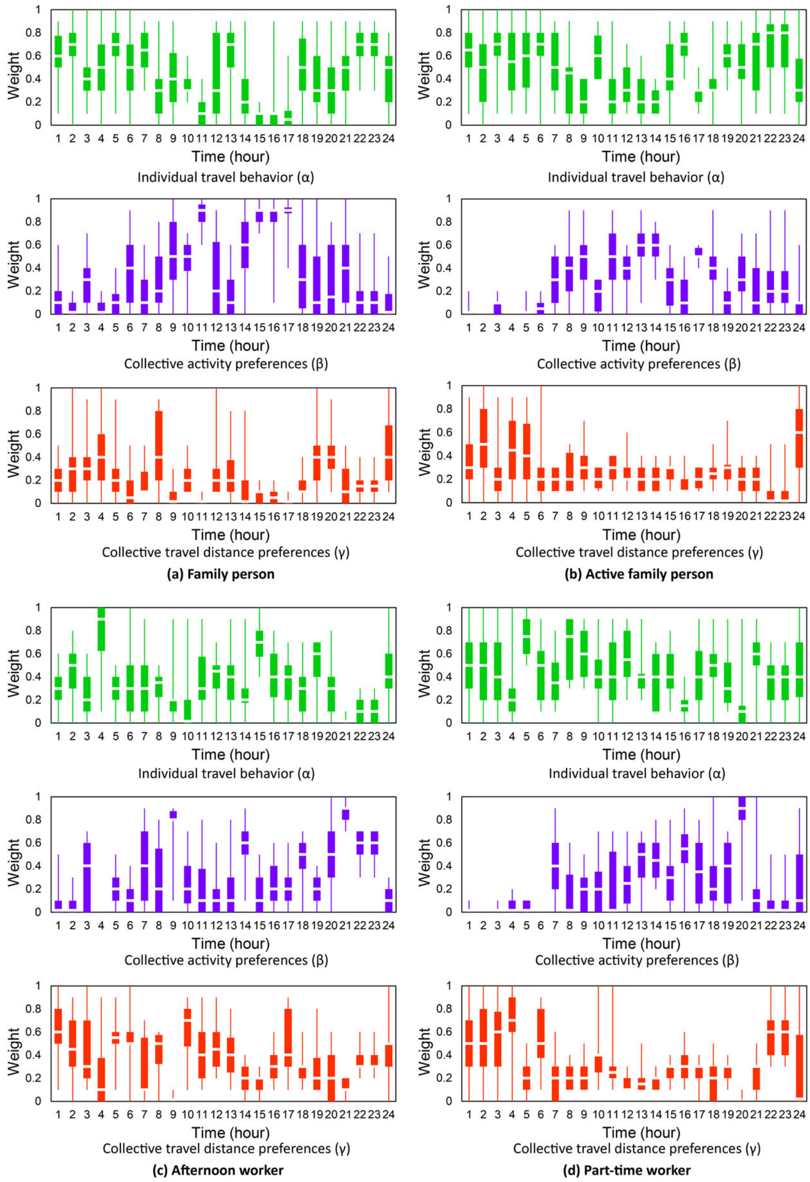

Figure 13. Variations in α, β and γ from 1:00 to 24:00 for four groups of users.