Novel model for predicting individuals’ movements in dynamic regions of interest

Xiaoqi Shen a, Wenzhong Shi b, Pengfei Chen b,c, Zhewei Liu a and Lukang Wang a

a School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu, China; b Department of Land Surveying and Geo-Informatics, Otto Poon Charitable Foundation Smart Cities Research Institute, the Hong Kong Polytechnic University, Hong Kong, China; c School of Geospatial Engineering and Science, Sun Yat-Sen University, Guangzhou, Guangdong, China

ABSTRACT
The increasing amount of geotagged social media data provides a possible resource for location prediction. However, existing location prediction methods rarely incorporate temporal changes in mobility patterns, which could lead to unreliable predictions. In particular, human mobility patterns have changed greatly in the COVID-19 era. We propose a novel model to predict individuals’ movements in dynamic regions of interest (ROIs), taking into account changes in activity areas and movement regularity. To address changes in the activity areas, we design a new updating strategy that can ensure the realistic extraction of an individual’s ROIs. Then, we develop an integration model for changes in the movement regularity based on two newly proposed prediction methods that consider both rapid and slow changes. The proposed integration model is evaluated based on five real-world social media datasets; three Weibo datasets related to COVID-19 collected in three Chinese cities, one Twitter dataset collected in New York and one dense GPS dataset. The results demonstrate that the proposed model can achieve better performances than state-of-the-art models, especially when mobility patterns change greatly. Combined with related pandemic data, this study will benefit pandemic prevention and control.

1. Introduction
Understanding human mobility patterns is useful for the efficient development of smart cities and studies on social behavior; moreover, it is widely applicable in various fields. Accordingly, numerous studies have been conducted on accessibility analyses (B. Y. Chen et al. 2018), crime dynamics (Caminha et al. 2017), and socioeconomic status (Xu et al. 2018). Recently, human mobility data have been used to study the influence of Coronavirus disease of 2019 (COVID-19) on economy and society (Bonaccorsi et al. 2020). Another study presented the correlation between Twitter and COVID-19 confirmed cases and proved that Twitter is helpful in analyzing epidemic outbreaks (Comito 2021). One of the main related issues is location prediction (P. Chen et al. 2019; Lv et al. 2017), that is, inferring the subsequent location most likely to be visited by an individual based on historical mobility data.

Currently, many studies related to location prediction populate the literature. First, locations are often recorded by longitude and latitude and should be processed for prediction. Some studies map the coordinates to grids (Nizetic, Fertalj, and Kalpic 2009; Do and Gatica-Perez 2012; Yao et al. 2017), but it is difficult to describe the actual activity space using the scale and shape of grids. Points of interest (POIs) are also popular for representing locations (Ahas et al. 2010; Q. Liu et al. 2016). However, for places where POIs are very dense, such as a large shopping mall, the matching method may not work. Regions of interest (ROIs) extracted by clustering coordinates are now popular because of their arbitrary shape and no additional data (Q. Huang 2017; P. Chen et al. 2019). The individuals’ mobility patterns are then modeled for location prediction. The prediction models include Markov-based, classification, sequence-based, and neural networks. Many Markov- and sequence-based models cannot extract temporal patterns well (Lv et al. 2017; Q. Huang 2017; Kautsar and Akbar 2017), and neural networks rely on very dense data for model training (Karatzoglou et al. 2017), which is a difficult condition for an individual. Classification models can utilize spatial–temporal features and are...
adapted to both sparse and dense data (Comito 2018; P. Chen et al. 2019). However, they still face some difficulties that need to be addressed.

A significant problem in location prediction is dynamic activities, including changes in ROIs and movement regularity. Especially under COVID-19, human mobility patterns have changed greatly, including tourism (M. Yang et al. 2021), traffic (Rojas-Rueda and Morales-Zamora 2021; Harkins et al. 2021), and workplace (Fukumura et al. 2021). As a result, the ROIs of the individuals constantly changed. Many existing prediction methods do not consider the ROI changes, caused by reasons such as working from home under COVID-19 (Q. Huang 2017; P. Chen et al. 2019), which leads to unreliable prediction results. Some studies take activity changes into account but only in training data, and changes in new data cannot be identified (W. Huang et al. 2015). Additionally, the regularity of an individual’s movement varies with time owing to the physical conditions or road reconstruction (Sevtsuk and Ratti 2010). Traditional methods encounter inconsistency between old and new data (Cho 2016), and one solution is a data window that can collect new data and further delete old data (Chen et al. 2019). However, it is difficult to determine the length of the window and methods such as a naïve Bayes model with adaptive windowing (NB-ADWIN) are necessary (Bifet and Gavalda 2007). In addition, using the decay factor can help reduce the effect of old data (Chen and Tu 2007). However, both NB-ADWIN and decay factor still suffer from very complex changes, and adapting to location prediction is still a problem.

To resolve problems caused by dynamic activities, we propose a model to predict individuals’ movements by considering dynamic ROIs (PMDR). The main contributions of this study are as follows.

(1) An updating strategy was designed to ensure the realistic extraction of an individual’s ROIs, which fills the gap in the change detection of updated activity regions. This strategy uses a new index based on the spatial similarity between the old and new ROIs to detect whether the ROIs change. Previous studies either used unchanged ROIs or only considered changes in training data. In contrast, our model can check the changes in ROIs with new data and update the ROIs.

(2) Two new prediction models were proposed for different degrees of change in the mobility pattern. One prediction model extends the NB-ADWIN with adaptive weights and updated ROIs (PMAU) for rapid changes. Then, a prediction model with decay factors (PMDF) was proposed for predictions when changes in the movement regularity are slow. Most previous studies used a fixed window length; however, the limitation of such approaches is that a fixed length can be suitable for complex changes in ROIs and movement regularities. The two proposed models use adaptive values of the window length or decay factors to ensure that the models are more suitable for prediction in dynamic ROIs.

(3) An integration model, PMDR, based on PMAU and PMDF was developed to choose the most credible result from the two models. Consequently, more stable and reliable prediction results were obtained compared to the results of the separate models. Most existing location prediction models either use a signal model or train different models’ weights in the training data. The proposed model further updates the weights for each new data to dynamically adjust the integration model with the individuals’ mobility patterns.

The proposed model was evaluated on five real-life datasets and showed a better performance than those of state-of-the-art models. The experiment datasets included three Weibo (a social media platform in China) datasets, one Twitter dataset, and one GPS dataset. The data comprising Weibo datasets obtained when the COVID-19 pandemic first spread through China were collected in three different cities, including Beijing, Shenzhen, and Shanghai. The Twitter dataset was collected from July 2018 to January 2019 in New York. The GPS dataset called Geolife (Zheng et al. 2009) is a public dataset with a period from April 2007 to August 2012. The proposed model performed better than the current models for all five datasets, especially when the mobility patterns changed significantly.

The remainder of this paper is organized as follows. Section 2 includes a review of the related studies and Section 3 provides an introduction of the experimental datasets. The proposed model is discussed in Section 4, and the experiments and results are presented in Section 5. We conclude the paper in Section 6.
2. Related studies

2.1. Human mobility analysis using social media data

The data obtained from social media have recently become a popular data source in human mobility studies owing to their extensiveness (Kemp 2018) and various methods by which they can be accessed, such as application programming interfaces (APIs) (Martí, Serrano-Estrada, and Nolasco-Cirugeda 2019). Although there are different types of data that can reflect human mobility, such as traffic trajectory (Liu et al. 2017), mobile phone (Jiang, Ferreira, and Gonzalez 2016), and smart card (Zhong et al. 2016) data, access to these data is limited owing to their high cost and privacy issues. By contrast, social media data are available at a price and can be accessed through crawling and APIs (Martí, Serrano-Estrada, and Nolasco-Cirugeda 2019). In particular, the long-term series of user trajectory data from social media data can be used for location prediction (Chen et al. 2019). In comparison to other types of data, social media data cover the wide spatiotemporal range of human mobility; moreover, they can be easily accessed by researchers.

Various studies have performed human mobility analyses based on social media data. Wu et al. (2014) proposed a method to model location and activity transitions using check-in data. They utilized two types of travel demands and three types of trip patterns to build a model based on the two parts of transition probabilities associated with activity and movement. Yang et al. (2019) mined the mobility patterns of a college community and determined how distance influences the active areas. Huang and Wong (2015) introduced a method to build space-time paths represented by cones of different sizes and colors. An approach, called TPM, was presented to extract dense regions and mobility patterns in the regions and was evaluated using a new validation methodology. In addition, the extracted patterns were evaluated by a novel similarity measure method (Cesario, Comito, and Talia 2017). Cui, Xie, and Liu (2018) analyzed human mobility using a novel framework based on Weibo check-in records. The framework can identify activity-specific locations and modifies the bias of population representativeness. A recent review surveyed and classified event detection methods based on Twitter data, and also revealed the subtasks and shortcomings of the detection methods (Hasan, Orgun, and Schwitter 2018).

2.2. Location identification

Some applications, such as location prediction and recommender systems, prioritize the identification of location (i.e. by extracting meaningful locations from raw data) before using the origin location data. For this research, the original locations were recorded using coordinates (longitude and latitude). One location may be represented by different coordinates, making it necessary to extract activity areas from the raw data. This can be done by extracting stay points, which are locations where users stay beyond a time threshold (Zheng et al. 2009). After extracting the stay points of all users, a clustering method is used to divide the stay points into different groups. Thomason, Griffiths, and Sanchez (2016) presented a novel method to extract stay points by defining a buffer of specific points and deciding whether a trajectory point belongs to a visit point according to the buffer gradient.

Several methods of extracting stay points are based on dense cell phone data, but the individual trajectories of social media data are irregular and sparse (Huang 2017). The time interval of continuous trajectory points is relatively long, and the interval length is not fixed. Clustering methods are commonly used for social media data, such as density clustering (Huang and Wong 2015) and graph clustering (Papadopoulos et al. 2012). Huang (2017) used density-based spatial clustering of applications with noise (DBSCAN) to cluster the locations of an individual’s Tweets and formed ROIs according to the boundary of each cluster. The clustering parameters were determined by the empirical values and all individual parameters were the same. Chen et al. (2019) utilized hierarchical DBSCAN (HDBSCAN) with an adaptive parameter based on two criteria (HDAP) to replace DBSCAN. However, individuals may not visit a location that was once frequently visited for several reasons, such as moving to a new home (Huang et al. 2015) or the urge to visit new locations. The dynamic nature of human mobility patterns and gradual changes in activity areas are prominent issues in existing studies. Therefore, the current study utilizes an updating strategy to solve the problem related to changes in the activity area.
2.3. Human mobility prediction

Several studies have been conducted on human mobility prediction, and some studies have obtained positive results. By studying the trajectories of mobile phone users, González, Hidalgo, and László Barabási (2008) found that human mobility has a high temporal and spatial regularity. Song et al. (2010) predicted a potential average human mobility of 93% using mobile phone carrier records by combining the empirically determined user entropy and Fano’s inequality. Based on the predictability of human mobility, various prediction methods have been proposed that can be classified according to two methods: objective and basic.

There are two main prediction models based on objective methods: next-place and spatiotemporal location predictions (Chen et al. 2019; Lv et al. 2017). Next-place predictions infer the next location of an individual based on their history of visited places. A next-place prediction method was proposed by combining several features describing the mobility patterns based on MS model trees (Comito 2018). Chauhan, Tejwani, and Toshniwal (2016) proposed a model using classifiers combined with Markov models that was tested on a Twitter dataset. Spatiotemporal location predictions aim to predict the location of an individual at any instance. Unlike next-place prediction, spatiotemporal location predictions consider more temporal information, such as the time interval between the current and next visits. Ma et al. (2015) proposed a model named W4 with multiple functions, including location prediction for a given instance. Liu et al. (2016) extended a recurrent neural network to spatial–temporal recurrent neural networks by modeling the spatiotemporal information and analyzed the influence of different window widths on the accuracy.

The main types of human mobility prediction models under the basic methods include Markov-based (Lv et al. 2017), classification (Comito 2018), sequence-based (Kautsar and Akbar 2017), and neural network-based models (Karatzoglou et al. 2017). Among these, Markov-based models are popular owing to their simplicity. Huang (2017) predicted an individual’s locations by combining a Markov chain model with temporal information and placing importance on the ROI. Chen et al. (2019) combined the transition probabilities with the temporal habit and time interval based on a naïve Bayes algorithm. Sequence-based models commonly use sequential pattern-mining technologies to predict future locations by measuring the similarity between the current trajectory and patterns. Ying, Lee, and Tseng (2013) mined the patterns in space, time, and semantics, and built a prefix tree to index them. Comito (2020) proposed a next-place prediction framework, based on frequent pattern mining and classification. This model combined individuals’ mobility patterns and all Twitter users’ mobilities to obtain better performance. Recently, neural networks have been studied for predicting individual locations owing to their strong learning abilities. Yao et al. (2017) presented a prediction framework with the multiple factors considered in recurrent neural networks (RNNs) for semantic trajectories. The framework was tested on two datasets, Foursquare and Twitter, which yielded a better performance than those of other methods. Despite the existence of several methods for predicting individual locations, changes in human mobility patterns have been overlooked. Thus, this study proposes an integration model to overcome the problem of changes in movement regularity.

3. Data description

Five datasets were used in this study, including three Weibo (a social media platform in China) datasets, one Twitter dataset and one GPS data, called Geolife. The Weibo datasets were collected in three different cities, including Beijing, Shenzhen, and Shanghai. The three cities are all important and famous cities in China in terms of economy and culture, and the time range of the three datasets covered the outbreak of COVID-19. Weibo data from Wuhan, a large city affected by the COVID-19 pandemic, was not collected because many individuals stayed at home, and the mobility data were too small to test. The Twitter dataset was collected in New York and the main data of Geolife were recorded in Beijing by GPS receivers and GPS phones.

To ensure the validity of the experiment, preprocessing was implemented on the five datasets. First, dirty data were deleted, including repeated data and data with missing and abnormal values. Then, individuals with a time span of fewer than two months
were filtered to ensure that prediction models could learn mobility patterns from the data. Finally, 19,954 stay points were extracted from the Geolife dataset using the method in the work that published the dataset (Zheng et al. 2009). The main statistical data after preprocessing are listed in Table 1. The time span of the Geolife dataset is very large (5 years), and that of the Shanghai dataset also reaches 21 months with 8,723 individuals. Figure 1 depicts the heatmaps of the five datasets, showing the different spatial distributions of hot spots. For example, Figure 1(b) shows many small hot spots distributed in different places, and in Figure 1(d) there are two main hot spots and a small one. Figure 2 shows the data distribution of the datasets, and a few individuals’ data sizes are much larger than others, and they account for most of the total posts. According to Figure 3, the time-frequency of the New York data- set mainly ranges from 0.5 to 3 h and that of the Shanghai dataset ranges from 0 to 6 h. In addition, the sample rate of the Geolife dataset is approximately every 1–5 s at a data point. In summary, the datasets including both the sparse (social media data) and dense (GPS data) one with different hot spot spatial distributions; three of these were influenced by the COVID-19 pandemic and the Spring Festival travel rush. Therefore, the five datasets can efficiently test the reliability and generality of the proposed model.

4. Methodology

As illustrated in Figure 4, the workflow of the proposed model includes the following three parts.

1. Change detection of ROIs: Individuals’ historical social media data are first clustered to extract the ROIs. When new data are obtained, the ROIs are updated. Then, a method is proposed to detect whether the ROIs change by measuring the similarity between the old and new ROI sets. Once a change occurs, the prediction model described in Section 4.2, is rebuilt; otherwise, the new data are used to update the model.

2. PMDR model construction: The spatiotem- poral features characterizing mobility are first extracted. The PMAU model with multiple weighted features is proposed and trained for prediction when rapid changes occur in the mobility pattern. Then, the newly proposed PMDF model that weakens the influence of old data is trained for prediction when the changes in human mobility patterns are slow. Finally, the two models were integrated to build the proposed integra- tion model, PMDR. Consequently, the PMDR model can obtain stable and reliable prediction results when the mobility pattern changes.

3. Prediction and model updating: A prediction request requires the PMDR model to infer the subsequent ROIs at a specific time. Once a new record is published, the prediction results based on the time of the new record and true ROI are combined to update the prediction model.

4.1. Change detection of ROIs

In this study, we utilized hierarchical DBSCAN with an adaptive parameter (HDAP) to extract ROIs from individuals’ online footprints, and propose a method to detect the changes in ROIs. The adaptive parameter is the minimum number of points in a cluster. As change detection of ROIs is independently processed on each individual, the concepts introduced next are also personal. After clustering the existing data, new data posted by an individual may change the ROIs, and

| Table 1. Detailed information of the five datasets, showing the differences in statistical characteristics. The time spans of the datasets are reported, as well as the data scales of the entire datasets and the average individual scales. |
| --- |
| **Beijing dataset** | **Shenzhen dataset** | **Shanghai dataset** | **New York dataset** | **Geolife dataset** |
| **Time span** | January 2020 – April 2020 | January 2020 – April 2020 | January 2019 – September 2020 | July 2018 – January 2019 | April 2007 – July 2012 |
| **Number of individuals** | 170 | 118 | 8,723 | 129 | 61 |
| **Total records** | 21,876 | 16,512 | 1,000,394 | 84,104 | 21,273,620 |
| **Average records per individual** | 129 | 140 | 125 | 652 | 348,748 |
the data are re-clustered to ensure realistic extraction of individuals’ activity areas. Because human mobility has a certain regularity, the clustering process is performed again after obtaining a certain amount of data. Here, the amount of data is set as the minimum number of points to form a cluster, which is automatically determined by the HDAP algorithm.

To demonstrate the necessity of updating the ROIs, an individual from the datasets was selected to show the changes in the ROIs. Figure 5 depicts the ROI identification of the individual at different time points. There are three common clusters in both the left and right pictures, including Clusters 1, 3, and 4. Clusters 3 and 4 remain the same at different time periods, and Cluster 1 obtains new points in the right picture. In addition, an evident difference was observed between the two pictures. According to Figure 3(a), there are a small number of points and only one cluster, Cluster 2, in the lower left. However, in Picture 3(b), the surrounding visits increased, and Cluster 2 evolved into two new clusters, Cluster 5 and Cluster 6. If there is no ROI updating strategy, some possible new ROIs, such as Cluster 5, cannot be extracted because the locations in the area are all considered as noise. Therefore, it is necessary to update and detect changes in ROIs to ensure a realistic extraction of individuals’ activity areas.

![Heatmaps of the five datasets with a different blue brightness representing data density differences. The brighter the color, the denser the data. It shows the spatial distribution of individual activity hotspots across the five cities.](image_url)
Figure 2. Distribution of individual data volume across the five datasets, where individuals are sorted by the amount of data from large to small.

Figure 3. Time-frequency of the five datasets. Time intervals for continuous records of individuals are calculated and the maximum, minimum, median and upper and lower quartiles are presented in the box diagram.
Two situations are observed after clustering is performed again: (1) the ROIs change and the prediction model should be rebuilt, and (2) the ROIs do not change, and the previous prediction model is continually used. First, to demonstrate the change detection method, some related concepts are presented. In this research, location \( l \) is represented by longitude and latitude, \( l = (\text{lon}, \text{lat}) \). An ROI \( r \) is a set of locations clustered by HDAP, \( r = \{l_0, l_1, l_2, \ldots, l_k\} \), where \( k \) is the number of locations in the ROI; thus an ROI set with \( n \) ROIs is defined as \( R = \{r_0, r_1, r_2, \ldots, r_n\} \). Then, the proposed method detects whether the ROIs change by calculating the similarity between the old and new ROI sets \( R_{\text{old}} \) and \( R_{\text{new}} \). The ROIs in the sets are called \( r_{\text{old}}(i) \) and \( r_{\text{new}}(j) \), where \( 0 \leq i \leq n_{\text{old}} \), \( 0 \leq j \leq n_{\text{new}} \); \( n_{\text{old}} \) and \( n_{\text{new}} \) are the numbers of ROIs in the old and new ROI sets, respectively. After clustering again, the labels of the two same ROIs are different; therefore, the two same ROIs should first be found and assigned to the same label. The similarity between two ROIs \( r_{\text{old}}(i) \) and \( r_{\text{new}}(j) \) is defined as follows:

\[
S(r_{\text{old}}(i), r_{\text{new}}(j)) = \frac{N(r_{\text{old}}(i) \cap r_{\text{new}}(j))}{N(r_{\text{old}}(i) \cup r_{\text{new}}(j))}
\]  

(1)

Figure 4. Proposed methodology framework with three main parts. The trapezoids refer to historical and new data, the diamond is used to show the different steps according to the detection result, and the squares mean various processing steps.

Figure 5. Real case of an individual’s ROI identification. Different clusters are distinguished by color and each cluster represents an ROI. The change in the ROIs from (a) to (b) is shown.
where \( N'(r_{\text{old}(i)} \cap r_{\text{new}(j)}) \) refers to the number of same locations in the two ROIs, and \( N'(r_{\text{old}(i)} \cup r_{\text{new}(j)}) \) denotes the total number of locations. If \( S(r_{\text{old}(i)}, r_{\text{new}(j)}) \) reaches a threshold, the two ROIs are the same, and their locations are assigned to the same label. Then, the updated ROIs are defined as \( r'_{\text{old}(i)} \) and \( r'_{\text{new}(j)} \) and the updated ROI sets are \( R'_{\text{old}} \) and \( R'_{\text{new}} \). The similarity between \( R'_{\text{old}} \) and \( R'_{\text{new}} \) is defined as follows:

\[
S(r'_{\text{old}}, r'_{\text{new}}) = \frac{N(r'_{\text{old}} \cap r'_{\text{new}})}{N(r'_{\text{old}} \cup r'_{\text{new}})} \tag{2}
\]

where \( N(r'_{\text{old}} \cap r'_{\text{new}}) \) refers to the number of the common locations in the two sets, and \( N(r'_{\text{old}} \cup r'_{\text{new}}) \) refers to the total number of locations. Here, if the labels of a location in two ROI sets are the same, then this location is common. If \( S(r'_{\text{old}}, r'_{\text{new}}) \) reaches the similarity threshold, the ROI sets are unchanged after clustering again, and the prediction model does not need to be rebuilt. The similarity thresholds of ROIs and ROI sets are both set at 0.99 to detect as many changes as possible and allow minimal changes that may be produced by statistical errors.

The average rate of changes in ROIs per new data (RCP) is used to analyze the changes in ROIs, and can be calculated as follows:

\[
RCP = \frac{N_{\text{detection}}}{\sum_{i=1}^{N_{\text{newdata}}} S(r_{\text{old}}, r_{\text{new}})} \tag{3}
\]

where \( N_{\text{detection}} \) is the number of detected times, \( N_{\text{newdata}} \) is the number of new data at the ith change detection, and \( S(r_{\text{old}}, r_{\text{new}}) \) is the similarity between the old and new ROI sets.

4.2. PMDR model construction

4.2.1. Mobility features

As mentioned earlier, in this study, we integrated two classification models to predict future ROIs; therefore, several features must be defined for classification. We focused on spatiotemporal data and defined three different features: previous visit, time of the day, and day of the week.

Previous visit (pv) is an important feature for describing the human mobility pattern and is commonly used in predictions (Chauhan, Tejwani, and Toshniwal 2016; Huang 2017). For example, if mobility behavior is from \( r_i \) to \( r_j \), then \( pv \) refers to \( r_j \).

Time of the day (td) refers to the time when an individual visits \( r_j \) during the day. In this study, we utilized the division method used by Comito (2018), wherein a day is divided into six parts: night (N, 0:00–5:59), early morning (EM, 6:00–9:59), morning (M, 10:00–13:59), afternoon (A, 14:00–17:59), early evening (EE, 18:00–20:59), and evening (E, 21:00–23:59).

Day of the week (dw) refers to the day when an individual visits \( r_j \) during the week. It was used in this study because individuals usually exhibit clearly different mobility patterns on weekdays and weekends (Long, Jin, and Joshi 2012), and a certain degree of difference is observed even on different weekdays (Zhou et al. 2018; Zhang and Liu 2018).

4.2.2. PMAU model and PMDF model construction

Both PMAU and PMDF models are based on the naïve Bayes algorithm (NB). The NB in this study can be written as

\[
\hat{r}_{NB} = \underset{r \in R}{\text{argmax}} \ P(r = r_i|pv, td, dw) \\
\propto \underset{r \in R}{\text{argmax}} \ P(pv|r_i)P(td|r_i)P(dw|r_i)P(r_i)
\]

where \( \hat{r}_{NB} \) is the predicted ROI by NB and \( r_i \) is the ROI of the ROI set.

To solve the problem of gradual changes in the movement regularity, the PMAU model is proposed and defined as follows:

\[
\hat{r}_{\text{PMAU}} = \underset{r \in R}{\text{argmax}} \ \frac{P(pv \land r_i)^{w_{pv}}}{P(r_i)} \frac{P(td \land r_i)^{w_{td}}}{P(r_i)} \frac{P(dw \land r_i)^{w_{dw}}}{P(r_i)} P(r_i)^{w_r} \tag{5}
\]

where the symbol \( \land \) means the left and right things of \( \land \) happen together, \( w_{pv}, w_{td}, w_{dw}, w_r \) are the weights of the four objects, including three features and one prediction object. To calculate the probabilities in the equation, the adaptive window (ADWIN) was built for each combination of different values of features and prediction objects. The window stores an unfixed length of data, and when the window obtains new data, the data in the window are repeatedly divided into old and new parts from the first data point to the last. The mean values of the two parts are calculated to check if a change occurred. If the data change, the old part is deleted, and a new part is used to calculate the probability. Then, a method to adaptively adjust the weights according to the real-time prediction result was
developed. In Equation (4), four objects need to be weighted; the larger the weight, the more important is the object. If a prediction result tends toward the true value of the ROI (r_{true}) due to an object, the object is more important. Here, r_{true} refers to the ROI that an individual visits given the three mobility features. Therefore, the probabilities of four objects with r_{i} (r_{i} \in R) were tested to find the ROI with the largest probability, \hat{r}_{object}, where object \in \{pv, td, dw, r\}. For example, to weigh the mobility feature pv, \hat{r}_{pv} can be calculated as follows:

\[
\hat{r}_{pv} = \arg \max_{r_{i} \in R} \left( \frac{P(pv \cap r_{i})}{P(r_{i})} \right)
\]  

(6)

If \hat{r}_{object} is the same as r_{true}, the object is considered to have a positive role in the equation, and its weight can increase. The weights can be updated as follows:

\[
w_{object} = \begin{cases} 
W_{object} & \text{if } \hat{r}_{object} = r_{true} \\
W_{object} \times c & \hat{r}_{object} \neq r_{true}
\end{cases}
\]

(7)

where object \in \{pv, td, dw, r\}, c is a coefficient that ranges from 0 to 1, and p is the probability of \hat{r}_{object}. When p is greater than 0.5, the result is more certain, and if it is true, its weight is increased further. When p is less than 0.5, even if it is true, the result is not reliable, and the weight is increased in smaller increments. If the result is incorrect, regardless of whether the probability is greater or less than 0.5, it is considered unreliable, and the weight is decreased. The value of c represents the degree of change in the mobility pattern, and can be determined based on a grid search. The training data were divided into two parts (80% and 20%). The proposed model with different c (ranging from 0 to 1 and 0.05 increase each time) is trained on one part (80%) and tested on another (20%) to find the value of c that can obtain the best performance.

Although we propose the PMAU model for rapid changes, it is not always stable. In a window, when the data changes slowly, there is no clear split point to divide the data into two parts: data before a change and data after a change, which is a significant problem. Additionally, the uncertainty in human mobility can influence the performance of the PMAU. For example, if on one day an individual stays at home and does not go to work due to a minor illness, PMAU may detect the change and delete all the data before the illness. When the individual returns to work, the PMAU model needs to learn the pattern again. Therefore, another stable method that can weaken the influence of outdated data is required. In this study, we propose an additional model, PMDF, that combines NB with decay factors for spatiotemporal location prediction. The probability of r_{i} can be calculated as follows:

\[
P(r_{i} | pv, td, dw) \propto \frac{N(pv \cap r_{i})N(td \cap r_{i})N(dw \cap r_{i})N(r_{i})}{N(r_{i}) N(r_{i}) N(r_{i}) N(r) N(r) N(r)}
\]

(8)

where N(pv \cap r_{i}) is the number of visits to r_{i} from pv. N(td \cap r_{i}) and N(dw \cap r_{i}) denote the number of visits to r_{i} at the time represented by td and dw, respectively. N(r_{i}) is the number of visits to r_{i} and N(r) is the total number of visits to any ROI of an individual. In summary, all variables in the equation represent the number of occurrences of different incidents, including pv \cap r_{i}, td \cap r_{i}, dw \cap r_{i}, r_{i}, and r. Then t_{io} denotes the time when an incident occurs (io) and S_{io} denotes the set of occurrences of this incident. For example, when the incident refers to td \cap r_{i}, io means the individual visit r_{i} at td once; when it refers to r, io means the individual visits any ROI once. From the decay factor, the number of occurrences of an incident at t, N_{t}(io), can be defined as follows:

\[
N_{t}(io) = \sum_{io \in S_{io}} c^{-t_{io}}
\]

(9)

where c is the decay factor value, which is equal to that in Equation (7).

Based on Equations (8) and (9), the PMDF model can be extended with weighted features as follows:

\[
\hat{r}_{PMDF} = \arg \max_{r_{i} \in R} \frac{N_{t}(pv \cap r_{i})^{w_{pv}} N_{t}(td \cap r_{i})^{w_{td}}}{N_{t}(r_{i})} \frac{N_{t}(dw \cap r_{i})^{w_{dw}} N_{t}(r_{i})^{w_{r}}}{N_{t}(r_{i})^{w_{r}}}
\]

(10)

The weights in Equation (10) are updated by Equation (7) when getting a new data point.

4.2.3. Integrating models

The build of the PMDR model is based on the two prediction models. As the PMAU and PMDF models are both Naïve Bayes-based models and use the same mobility features described in Section 4.1, we can implement the same data input, storage, and management on the two models. The main difference between the two models is the method used to calculate the probability: Equations (5) and (10). Therefore, we set a weight for each model to evaluate the reliability of the two models’ probabilities; for predictions, the model with the higher weight was selected. The integration process is defined as follows:
\[
\hat{r}_{\text{PMDR}} = \begin{cases} 
\hat{r}_{\text{PMAU}} & \text{if } w_{\text{PMAU}} \geq w_{\text{PMDF}} \\
\hat{r}_{\text{PMDF}} & \text{if } w_{\text{PMAU}} < w_{\text{PMDF}}
\end{cases}
\]

where \(\hat{r}_{\text{PMDR}}\) is the prediction result of the integration model, \(w_{\text{PMAU}}\) and \(w_{\text{PMDF}}\) are the weights of the two models. The same updating strategy of the features was used to update the weights of the models. The same updating strategy of the features was used to update the weights of the models. In Equation (6), let \(\text{object} \in \{\text{PMAU, PMDF}\}\). Then, \(w_{\text{PMAU}}\) and \(w_{\text{PMDF}}\) are the weights of the two models, and \(r_{\text{object}}\) is the ROI predicted by the two classification models. This process yields a larger model weight with higher prediction accuracy. The upper bound of the weights was set to 1, and the lower bound was set to 0.1, instead of 0, to ensure that the weights reached the lower bound in a limited time and increased more quickly. Most existing studies train the weights of different models on the training data, and the weights are fixed; but our method updates the weights with new data to adjust the most suitable model for the new data.

Based on Equations (5), (10), and (11), the PMDR model is defined by Equation (12).

\[
\hat{r}_{\text{PMDR}} = \begin{cases} 
\arg \max_{\theta \in \Theta} \left( w_{\text{PMAU}} \cdot \frac{N_{\text{PMAU}}(r_i^\theta)}{N_{\text{PMAU}}(r_i)} + (1 - w_{\text{PMAU}}) \cdot \frac{N_{\text{PMDF}}(r_i^\theta)}{N_{\text{PMDF}}(r_i)} \cdot \hat{r}_{\text{PMAU}} \right) & \text{if } w_{\text{PMAU}} \geq w_{\text{PMDF}} \\
\arg \max_{\theta \in \Theta} \left( w_{\text{PMDF}} \cdot \frac{N_{\text{PMAU}}(r_i^\theta)}{N_{\text{PMDF}}(r_i)} + (1 - w_{\text{PMDF}}) \cdot \frac{N_{\text{PMDF}}(r_i^\theta)}{N_{\text{PMDF}}(r_i)} \cdot \hat{r}_{\text{PMDF}} \right) & \text{if } w_{\text{PMAU}} < w_{\text{PMDF}}
\end{cases}
\]

5. Experiment

5.1. Efficiency in change detection of ROIs

The raw coordinates of individuals were clustered to extract ROIs, and the details are shown in Table 2. The average number of ROIs per user in the Geolife dataset is maximal, which means that this dataset recorded the most detailed movements of individuals. The New York dataset has the largest number of visits to each ROI. It is reasonable that the New York dataset has the largest average records per individual in the social media datasets and the visits to the Geolife dataset are calculated by stay points. In addition, the New York dataset has the largest average area of ROIs, which means that the individuals in this dataset may post in several main ROIs. To analyze the changes in ROIs, the similarities calculated by the change detection method were recorded, and their distributions are shown in Figure 6. Most values of the similarity were in the range 0.9–1, and the proportion of other segmentations was significantly less than this range, which means that the changes in ROIs were mostly minimal. Additionally, the similarity distributions of the three Weibo datasets were comparable, and the similarities of the New York dataset were smaller than those of the other datasets. This implies that the ROIs of individuals in New York changed more than individuals in other cities for every change detection. Figure 7 shows the rate of detected changes to all detected times, as well as the average rate of changes in ROIs per new data (RCP), which is defined in Section 4.1. Additionally, for the Geolife dataset, a new stay point is used in the RCP instead of each new GPS record. From Figure 7(a), it is evident that changes in the ROIs were widespread, and the maximum rate was obtained for the Geolife datasets. According to Figure 7(b), the RCP of the New York dataset is the lowest, which means that the ROIs of the other four datasets changed more quickly than those of the New York dataset. The RCP of the Shanghai dataset is maximal, which may be due to the maximum time coverage of COVID-19. For the Geolife dataset, the large value of RCP may be caused by the detailed records, which means that the stay points can represent much more activities of individuals than social media data. In summary, changes in ROIs were widespread and minimal at most times in all five datasets.

The Shanghai dataset was selected to further analyze the influence of COVID-19 on the changes in the ROIs. Figure 8 shows the time distribution of the similarity between the old and new ROIs. An obvious minimum value of similarity appears around the end of January 2020, during the COVID-19 outbreak and beginning of the Spring Festival. Then, from mid-February to early April 2020, the similarity is very high because many individuals reduced outdoor activities and stay at home. Finally, the similarity dropped for two months, which shows that individuals returned to work and school. In addition, individuals may travel on May Day holiday and students start summer vacation. Overall, the demand for
outdoor activities rebounded and led to changes in the ROIs. This shows that the similarity measure method captures the changes in crowd activities, so the change detection method can ensure a more realistic extraction of ROIs than others without change detection.

5.2. **Accuracy evaluation**

To evaluate the performance of the proposed model, each individual’s data were divided into two parts: the first (70%) and second (30%) parts were used to train and test the model, respectively. The performance is
described by the prediction accuracy, which is the proportion of correct predictions to the total predictions, and three popular classification indicators, including recall, precision, and F1-score. In addition to the proposed PMDR model, a multi-feature weighted Bayesian model (MWBM) proposed by Chen et al. (2019), a sparse mobility Markov chain model (SMMC) proposed by Huang (2017), spatial temporal recurrent neural networks (ST-RNN) proposed by Q. Liu et al. (2016), and the baseline model NB were tested. The parameters of MWBM were set to the recommended value in the study, and the features of NB were extracted in the same way as in this study. Compared to previous studies, the proposed method updates and detects changes in ROIs, which provides more complex mobility patterns and potentially influences the prediction accuracy. The proposed method is more suitable for real-world applications.

Figure 9 shows the accuracy of the five models for the datasets, and Figure 10 presents the detailed distribution of the accuracy. According to Figure 9, PMDR demonstrated the best performance among the five models on all datasets, which proves the validity of considering the dynamic ROIs. Additionally, Figure 10 showed that almost no individuals’ accuracies of PMDR distributed between 0 and 0.1, but for the other four models there were many examples whose accuracies ranged from 0 to 0.1, which shows the better stability of PMDR. The performance of MWBM was second only to PMDR,
benefiting from the fixed-length window. The prediction accuracies of the SMMC, NB, and ST-RNN models were poor in all cases because the models retained the outdated data, which may deviate from the latest mobility patterns. The SMMC can train the weights for the length of time each user stays in each region and the importance of a region, and the NB model assigns the same weight to all features; thus the NB model obtained a lower prediction accuracy. The ST-RNN model utilizes temporal information by extracting the time interval between two ROIs but ignores the importance of individuals’ specific activities at a specific time, which is used in all the other four models. As a result, the ST-RNN model obtained the lowest accuracy for most datasets. Totally, the performance of PMDR was better and more stable compared with the other models.

It is evident from Table 3 that PMDR had the best performance in Shanghai, and the average improvements by PMDR above other models were all over 10% and some even reached 35%. The performance of PMDR on the three datasets (Beijing, Shenzhen, and New York) was similar and worse than that of the Shanghai and Geolife datasets. This can be explained by the larger time span of the Shanghai and Geolife datasets than

**Table 3.** Detailed performance comparison of the five models. In addition to the prediction accuracy, three popular classification indicators are used to evaluate the comprehensive performance.

|                   | PMDR  | MWBM  | SMMC  | ST-RNN | NB    |
|-------------------|-------|-------|-------|--------|-------|
| **Beijing dataset** |       |       |       |        |       |
| Accuracy          | 60.42%| 48.06%| 42.32%| 32.58% | 40.15%|
| Recall            | 41.67%| 30.44%| 23.07%| 11.43% | 26.97%|
| Precision         | 41.12%| 26.77%| 26.70%| 21.09% | 25.24%|
| F1-score          | 41.40%| 28.49%| 24.58%| 14.83% | 26.07%|
| **Shenzhen dataset** |       |       |       |        |       |
| Accuracy          | 58.01%| 46.02%| 40.00%| 27.97% | 38.05%|
| Recall            | 39.49%| 28.20%| 21.59%| 8.23%  | 24.38%|
| Precision         | 38.88%| 25.29%| 25.34%| 19.94% | 23.75%|
| F1-score          | 39.19%| 26.66%| 23.40%| 11.65% | 24.06%|
| **Shanghai dataset** |       |       |       |        |       |
| Accuracy          | 70.84%| 52.95%| 48.41%| 26.11% | 34.36%|
| Recall            | 57.56%| 43.69%| 35.39%| 11.50% | 25.93%|
| Precision         | 49.16%| 31.10%| 35.42%| 21.23% | 21.20%|
| F1-score          | 53.03%| 36.34%| 35.41%| 14.92% | 23.33%|
| **New York dataset** |       |       |       |        |       |
| Accuracy          | 56.89%| 44.02%| 28.45%| 27.58% | 32.36%|
| Recall            | 34.98%| 25.10%| 10.03%| 6.72%  | 16.18%|
| Precision         | 36.54%| 22.01%| 14.14%| 15.56% | 17.16%|
| F1-score          | 35.74%| 23.45%| 11.73%| 9.39%  | 16.66%|
| **Geolife dataset** |       |       |       |        |       |
| Accuracy          | 64.10%| 35.31%| 26.78%| 20.02% | 31.43%|
| Recall            | 42.42%| 24.07%| 14.88%| 4.66%  | 23.29%|
| Precision         | 41.99%| 17.34%| 17.67%| 13.44% | 21.13%|
| F1-score          | 42.20%| 20.15%| 16.16%| 6.92%  | 22.16%|
| **Average improvement** |     |     |     |        |       |
| Accuracy          | \    | 16.74%| 24.86%| 35.20% | 26.78%|
| Recall            | \    | 12.93%| 22.24%| 34.72% | 19.88%|
| Precision         | \    | 17.04%| 17.72%| 23.29% | 19.84%|
| F1-score          | \    | 15.29%| 20.06%| 30.77% | 19.86%|
the others, which allows the models to extract mobility patterns better. However, the other four models, except PMDR, all show worse performance on the Geolife dataset than those on other datasets. This indicates that for a dense dataset, our proposed model PMDR can still obtain a better performance based on the updated weights and change detection both in the ROIs and the movement regularity. In conclusion, the PMDR model performed better than the other models for all four indicators of the five datasets.

Two indicators, the rate of detected changes to all detected times and RCP, were used to analyze the relationship between the prediction accuracy and change in the mobility pattern of an individual. For a more intuitive display, the RCP values were normalized and divided into ten intervals. In addition, while the value of RCP after normalization was greater than 0.4, there was very little data to be statistically significant in the intervals. Therefore, the data in these intervals were deleted, and normalization was performed on RCP again. Figure 11"
shows a plot of the relationship between the two indicators and prediction results, including the prediction accuracies of the five models and improvement of PMDR compared with other models. According to Figure 11(a,c), the overall prediction accuracy of the five models tended to be lower when the two indicators increased. This is because for large indicators, the mobility uncertainty of an individual is high, and the models cannot accurately fit the data. In addition, the prediction accuracies quickly reduced initially, before slowing, which implies that the prediction accuracy is greatly influenced when the changes in the mobility pattern increase from zero. In addition, the proposed PMDR model was less affected compared to the other models; the prediction accuracy almost not changes when the RRC increased, as shown in Figure 11(c). As a result, PMDR performed better than other four models, as shown in Figure 11(b,d). It can also be observed that with an increase in two indicators, the improvement of PMDR increased, compared with the other four models.
The Shanghai dataset was selected to show the influence of COVID-19 on prediction accuracy. Figure 12 shows the variation in the prediction accuracies and the improvements with time. The prediction accuracies of the five models remained steady at first and increased around February 2020 because individuals reduced their outdoor activities and stayed at home for epidemic prevention. PMDR still shows good performance from March to September 2020, but the prediction accuracies of other models decrease when individuals start going out. As a result, noticeably better improvement was observed in the PMDR compared with other models, as shown in Figure 12(b). Thus, it can be concluded that the proposed PMDR model can better fit data with ROI changes and has a better stability than the other four models.

5.3. Analysis of the prediction results

To further analyze the spatial and temporal characteristics of the prediction results, one spatial feature referring to the number of ROIs of individuals and

![Figure 12](image-url) - Comparison of the prediction results at different times. The variations in the prediction accuracies and the improvements are displayed to analyze the influence of COVID-19 on the prediction results.
another temporal feature referring to the time interval
in a day, which is described in Section 4.2.1, was
extracted. Figure 13 shows the relationship between
the number of ROIs and prediction accuracy of PMDR
using a kernel density estimation. The prediction
accuracy tends to decrease with increasing number
of ROIs in both datasets. This is because individuals
with fewer ROIs present a simple mobility pattern,
which can be easily fitted by the prediction model.
Figure 14 shows the prediction accuracies at different

![Diagram](image_url)

**Figure 13.** Relationship between the number of ROIs and prediction accuracy. The top and right parts of the figures show the distributions, and the middle parts show the relationship by a two-dimensional kernel density estimation.
times per day. From the figure, it is evident that the prediction accuracy exhibits certain regularities with different characteristics of the five datasets. The prediction accuracy decreased from N to A, and then increased from A to E for the Beijing and Shenzhen datasets. A similar observation is made in the Shanghai dataset, but it reached the lowest value in the time interval EE. Compared with the three Weibo datasets, the accuracy on the New York and Geolife datasets showed substantial differences. The lowest value was obtained in M, and the accuracies in other intervals were similar for the New York dataset; the maximum value was obtained in EE for the Geolife dataset. This can be attributed to the different activity patterns in the groups of individuals in the four datasets. In conclusion, the individuals showed different spatial and temporal characteristics, which greatly influenced the prediction performance.

Figure 14. Prediction accuracies performed on the five datasets at different times in a day. The accuracy of a time interval is the mean of all prediction accuracies in the time interval.
6. Conclusion

In this research, we proposed an integration model, PMDR, that includes the dynamic ROIs for movement prediction; the proposed model was evaluated on five datasets. An updating strategy was used to ensure that the ROIs conform to the real activity areas, and a spatial similarity measure method was designed to determine the changes in the ROIs; accordingly, changes in the activity areas were considered. The PMDR model integrates two newly proposed models: PMAU and PMDF. This led to improvements in the prediction accuracy when the mobility patterns changed, and an experiment was performed using three Weibo datasets under the COVID-19 pandemic, one Twitter dataset and one GPS dataset. The changes in the ROIs indicated that variations existed in individuals’ activity areas, and a change detection method is needed. The best prediction accuracy of the proposed model was 70.84% for the Shanghai dataset. Compared to state-of-the-art models, the proposed model performed better, especially when the human mobility patterns changed greatly. On average, the prediction accuracies for the MWBM, SMMC, and ST-RNN increased by 16.74%, 24.86%, and 35.20%, respectively.

Social media data and GPS data were used in this research, covering sparse and dense data. Social media data are easily accessible for research and obtained from location-based services, which contributes to the practical application of the proposed model. In addition, social media data are actively released by users and are mostly sparse, which reduces the severity of privacy issues compared with other datasets. As a result, our proposed model based on social media data can be better used in areas such as tourism and information services. However, the sparse data hardly represents individuals’ detailed mobility; the GPS data, which are dense data, were used to compensate for this shortcoming. The raw locations of the datasets are coordinates and are then represented by ROIs. The areas of ROIs have been shown in Table 2 and the ROIs can be schools, business centers, stadiums, etc. However, there may be some large ROIs that can influence the usefulness of the proposed model. Benefiting from the change detection method, some of the large ROIs can be updated and divided into several small parts.

This work can be useful in real applications for the protection of privacy and the description of the activity space. The use of ROIs can blur real positions, like home numbers, to protect the privacy of individuals. Many data providers reduce the accuracy of locations to prevent privacy leaks. This is fatal for the models based on accurate locations but has less impact on the ROI-based models. Moreover, the activities of individuals are dynamic, and the accurate positions at time points cannot describe them. ROIs can better represent the individual activity space around specific time points. Therefore, many services can provide better experiences based on the predicted ROIs. Many kinds of information, such as traffic, weather, epidemic situation, can be obtained according to the time points and predicted ROIs. Then, individuals can be informed of the related status of the predicted ROIs in advance to get better activity experiences. After the ROI prediction of individuals, governments can also facilitate the reasonable arrangement of public services, such as buses, subways, and taxis.

The proposed model is scalable and can be enhanced in the future. Textual social media data can be used as a model feature to extract semantic information. Other related data, such as points of interest and land use data, can enrich semantic information using methods such as map matching. In addition, our model currently runs on a computer, whereas most social media-based applications are used by mobile phones. In such situations, efficient routing, resource allocation, and energy management become major challenges, and methods such as an energy-aware clustering scheme may be necessary (Comito, Talia, and Trunfio 2011). Additionally, Twitter data have been proved helpful in analyzing epidemic outbreaks (Comito 2021). Further, we may try to combine Weibo datasets with COVID-19 data to propose a risk forecast method based on mobility prediction and apply it to mobile phones.

Data and codes availability statement

The data and codes that support the findings of this study are available in figshare.com with the identifier https://figshare.com/s/6854e58547fe0124c033.
Disclosure statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by National Key Research and Development Program of China under Grant Number [2019YFB2103102] and Otto Poon Charitable Foundation Smart Cities Research Institute, The Hong Kong Polytechnic University [Work Program: CD03].

ORCID

Xiaoqi Shen http://orcid.org/0000-0002-6156-8906
Wenzhong Shi http://orcid.org/0000-0002-3886-7027
Pengfei Chen http://orcid.org/0000-0002-0856-7234
Zhewei Liu http://orcid.org/0000-0002-4023-9142

References

Ahas, R., S. Silm, O. Järv, E. Saluveer, and M. Tiru. 2010. “Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones.” Journal of Urban Technology 17 (1): 3–27. Taylor & Francis. doi:10.1080/10630731003597306.

Bifet, A., and R. Gavalda. 2007. “Learning from Time-Changing Data with Adaptive Windowing.” Proceedings of the 7th SIAM International Conference on Data Mining. 443–448. doi:10.1137/1.9781611972771.42.

Bonaccorsi, G., F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. Lucia Schmidt, et al. 2020. “Economic and Social Consequences of Human Mobility Restrictions under COVID-19.” Proceedings of the National Academy of Sciences of the United States of America 117 (27): 15530–15535. doi:10.1073/pnas.2007658117.

Caminha, C., V. Furtado, T. H. C. Pequeno, C. Ponte, H. P. M. Melo, E. A. Oliveira, and J. S. Andrade. 2017. “Human Mobility in Large Cities as a Proxy for Crime.” PLoS ONE 12 (2): 1–13. doi:10.1371/journal.pone.0171609.

Cesario, B., A. Cimino, and M. den Hertog. 2017. “Understanding the Impacts of Human Mobility on Accessibility Using Massive Mobile Phone Tracking Data.” Annals of the American Association of Geographers 108 (4): 1115–1133. doi:10.1080/24694452.2017.1411244.

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. Taylor & Francis. doi:10.1080/13658816.2019.1630630.

Chen, Y., and L. Tu. 2007. “Density-Based Clustering for Real-Time Stream Data.” Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 133–142. doi:10.1145/1281192.1281210.

Cho, S. B. 2016. “Exploiting Machine Learning Techniques for Location Recognition and Prediction with Smartphone Logs.” Neurocomputing 176: 98–106. Elsevier. doi:10.1016/j.neucom.2015.02.079.

Comito, C., D. Talia, and P. Trunfio. 2011. “An Energy-Aware Clustering Scheme for Mobile Applications.” Proceedings - 11th IEEE International Conference on Computer and Information Technology, CIT 2011: 15–22. IEEE. doi:10.1109/CIT.2011.119.

Comito, C. 2018. “Exploiting Twitter for Next-Place Prediction.” International Conference on Information Society, ISociety 2017 2018-Janu: 143–148. doi:10.23919/ISociety.2017.8354690.

Comito, C. 2020. “NextF: A Framework for next-Place Prediction on Location Based Social Networks.” Knowledge-Based Systems 204: 106205. Elsevier B.V. doi:10.1016/j.knosys.2020.106205.

Comito, C. 2021. “How COVID-19 Information Spread in US the Role of Twitter as Early Indicator of Epidemics.” IEEE Transactions on Services Computing 1374 (c). IEEE. doi:10.1109/TSC.2021.3091281.

Cui, Y., X. Xie, and Y. Liu. 2018. “Social Media and Mobility Landscape: Uncovering Spatial Patterns of Urban Human Mobility with Multi Source Data.” Frontiers of Environmental Science & Engineering 12 (5): 7. Springer. doi:10.1007/s11783-018-1068-1.

Do, T. M. T., and D. Gatica-Perez. 2012. “Contextual Conditional Models for Smartphone-Based Human Mobility Prediction.” Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 163–172. doi:10.1145/2370216.2370242.

Fukumura, Y. E., J. M. Schott, G. M. Lucas, B. Becerik-Gerber, and S. C. Roll. 2021. “Negotiating Time and Space When Working from Home: Experiences during COVID-19.” OTR: Occupation, Participation and Health, 41, no. 4 (October 2021): 223–231.

González, M. C., C. A. Hidalgo, and A. Lázslo Barabási. 2008. “Understanding Individual Human Mobility Patterns.” Nature 453 (7196): 779–782. doi:10.1038/nature06958.

Harkins, C., B. C. McDonald, D. K. Henze, and C. Wiedinmyer. 2021. “A Fuel-Based Method for Updating Mobile Source Emissions during the COVID-19 Pandemic.” Environmental Research Letters 16 (4): 045018. IOP Publishing. doi:10.1088/1748-9326/ac0660.

Hasan, M., M. A. Orgun, and R. Schwitter. 2018. “A Survey on Real-Time Event Detection from the Twitter Data Stream.” Journal of Information Science 44 (4): 443–463. SAGE Publications Sage UK: London, England. doi:10.1177/0165551517698564.

Huang, Q., and D. W. S. Wong. 2015. “Modeling and Visualizing Regular Human Mobility Patterns with Uncertainty: An Example Using Twitter Data.” Annals of the Association of American Geographers 105 (6): 1179–1197. doi:10.1080/00045608.2015.1081120.
Huang, Q. 2017. “Mining Online Footprints to Predict User’s Next Location.” International Journal of Geographical Information Science 31 (3): 523–541. Taylor & Francis. doi:10.1080/14699601.2016.1209506.

Huang, W., S. Li, X. Liu, and Y. Ban. 2015. “Predicting Human Mobility with Activity Changes.” International Journal of Geographical Information Science 29 (9): 1569–1587. Taylor & Francis. doi:10.1080/14699601.2015.1033421.

Jiang, S., J. Ferreira, and M. C. Gonzalez. 2016. “Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore.” IEEE Transactions on Big Data 3 (2): 208–219. doi:10.1109/tbdta.2016.2631141.

Karatzoglou, A., H. Sentürk, A. Jablonski, and M. Beigl. 2017. “Applying Artificial Neural Networks on Two-Layer Semantic Trajectories for Predicting the Next Semantic Location.” Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10614: 233–241. LNCS. doi:10.1007/978-3-319-68612-7_27.

Kautsar, G., and S. Akbar. 2017. “Trajectory Pattern Mining Using Sequential Pattern Mining and K-Means for Predicting Future Location.” Journal of Physics: Conference Series 801 (1). doi:10.1088/1742-6596/801/1/012017.

Li, C., and Z. Su. 2018. “Digital in 2018: World’s Internet Users Pass the 4 Billion Mark.” We Are Social. https://wearsocial.com/uk/blog/2018/01/global-digital-report-2018/.

Liu, J., X. Yu, Z. Xu, K. K. R. Choo, L. Hong, and X. Cui. 2017. “A Cloud-Based Taxi Trace Mining Framework for Smart City.” Software - Practice and Experience 47 (8): 1081–1094. doi:10.1002/spe.2435.

Liu, Q., S. Wu, L. Wang, and T. Tan. 2016. “Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts.” Proceedings of the AAAI Conference on Artificial Intelligence 30.

Long, X., L. Jin, and J. Joshi. 2012. “Exploring Trajectory-Driven Local Geographic Topics in Foursquare.” Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 927–934. doi:10.1145/2370216.2370423.

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. IEEE. doi:10.1109/TVT.2016.2611654.

Ma, Z., A. Sun, Q. Yuan, G. Cong, and K. Zhao. 2015. “Who, Where, When, and What.” ACM Transactions on Information Systems 33 (1): 1–33. doi:10.1145/2699667.

Martí, P., L. Serrano-Estrada, and A. Nolasco-Cirugeda. 2019. “Social Media Data: Challenges, Opportunities and Limitations in Urban Studies.” Computers, Environment and Urban Systems 74 (September 2018): 161–174. Elsevier. doi:10.1016/j.comenvurbsys.2018.11.001.

Nizetic, I., K. Fertalj, and D. Kalpic. 2009. “A Prototype for the Short-Term Prediction of Moving Object’s Movement Using Markov Chains.” Proceedings of the 21st International Conference on Information Technology Interfaces 559–564.

Papadopoulos, S., Y. Kompatsiaris, A. Vakali, and P. Spyridonos. 2012. “Community Detection in Social Media Performance and Application Considerations.” Data Mining and Knowledge Discovery 24 (3): 515–554. doi:10.1007/s10618-011-0224-2.

Rojas-Rueda, D., and E. Morales-Zamora. 2021. “Built Environment, Transport, and COVID-19: A Review.” Current Environmental Health Reports 8: 1–8. Springer. doi:10.1007/s40572-020-00303-3.

Shtyuk, A., and C. Ratti. 2010. “Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks.” Journal of Urban Technology 17 (1): 41–60. doi:10.1063/1.3607310.53597322.

Song, C., Z. Qu, N. Blumm, and A. L. Barabási. 2010. “Limits of Predictability in Human Mobility.” Science 327 (5968): 1018–1021. doi:10.1126/science.1177170.

Thomason, A., N. Griffiths, and V. Sanchez. 2016. “Identifying Locations from Geospatial Trajectories.” Journal of Computer and System Sciences 82 (4): 566–581. Elsevier Inc. doi:10.1016/j.jcss.2015.10.005.

Wu, L., Y. Zhi, Z. Sui, and Y. Liu. 2014. “Intra-Urban Human Mobility and Activity Transition: Evidence from Social Media Check-in Data.” PLoS ONE 9 (5). doi:10.1371/journal.pone.0907010.

Xu, Y., A. Belyi, I. Bojic, and C. Ratti. 2018. “Human Mobility and Socioeconomic Status: Analysis of Singapore and Boston.” Computers, Environment and Urban Systems 72 (April): 51–67. Elsevier. doi:10.1016/j.compenvurbys.2018.04.001.

Yang, C., M. Xiao, X. Ding, W. Tian, Y. Zhai, J. Chen, L. Liu, and X. Ye. 2019. “Exploring Human Mobility Patterns Using Geo-Tagged Social Media Data at the Group Level.” Journal of Spatial Science 64 (2): 221–238. Taylor & Francis. doi:10.1080/14498596.2017.1421487.

Yang, M., C. Han, Y. Cui, and Y. Zhao. 2021. “COVID-19 and Mobility in Tourism Cities: A Statistical Change-Point Detection Approach.” Journal of Hospitality and Tourism Management 47: 256–261. Elsevier. doi:10.1016/j.jhtm.2021.03.014.

Yao, D., C. Zhang, J. Huang, and J. Bi. 2017. “SERM: A Recurrent Model for Next Location Prediction in Semantic Trajectories.” International Conference on Information and Knowledge Management, Proceedings Part F1318: 2411–2414. doi:10.1145/3132847.3133056.

Ying, J. J. C., W. C. Lee, and V. S. Tseng. 2013. “Mining Geographic-Temporal-Semantic Patterns in Trajectories for Location Prediction.” ACM Transactions on Intelligent Systems and Technology 5 (1): 1–33. doi:10.1145/2542182.2542184.

Zhang, Y., and L. Liu. 2018. “Understanding Temporal Pattern of Human Activities Using Temporal Areas of Interest.” Applied Geography 94 (March): 95–106. Elsevier Ltd. doi:10.1016/j.apgeog.2018.03.009.

Zheng, Y., L. Zhang, X. Xie, and W.-Y. Ma. 2009. “Mining Interesting Locations and Travel Sequences from GPS Trajectories.” Proceedings of the 18th International Conference on World Wide Web - WWW 9: 791. doi:10.1145/1526709.1526816.

Zhong, C., M. Batty, E. Manley, J. Wang, Z. Wang, F. Chen, and G. Schmitt. 2016. “Variability in Regularity: Mining Temporal Mobility Patterns in London, Singapore and Beijing Using Smart-Card Data.” PLoS ONE 11 (2): 1–17. doi:10.1371/journal.pone.0149222.

Zhou, Y., B. Pik Lek Lau, C. Yuen, B. Tucker, and E. Wilhelm. 2018. “Understanding Urban Human Mobility through Crowdsensed Data.” IEEE Communications Magazine 56 (11): 52–59. doi:10.1109/MCOM.2018.1700569.