WILEY | Hindawi

Research Article

Inferring Travel Modes from Cellular Signaling Data Based on the Gated Recurrent Unit Neural Network

Yanchen Wang (ID),[1] **Fei Yang** (ID),[1] **Li He,**[2] **Haode Liu,**[3] **Li Tan,**[4] **and Cheng Wang**[4]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[2]*Guiyang Transportation Development Research Center, Guiyang 550081, China*
[3]*Scientific Research Institute of the Ministry of Transport of PRC, Beijing 100029, China*
[4]*China Mobile Communications Group Sichuan Branch, Chengdu 610000, China*

Correspondence should be addressed to Fei Yang; yangfei_traffic@163.com

Cellular signaling data have become increasingly indispensable in analyzing residents' travel characteristic. Especially with the enhancement of positioning quality in 4G-LTE and 5G wireless communication systems, it is expected that the identification accuracy of fine-grained travel modes will achieve an optimal level. However, due to data privacy issues, the empirical evaluation of the performance of different identification methods is not yet sufficient. This paper builds a travel mode identification model that utilizes the gated recurrent unit (GRU) neural network. With 24 features as input, this method can identify four traffic modes, including walking, bicycle, car, and bus. Moreover, in cooperation with the operator, we organized an experiment collecting cellular signaling data, as well as the corresponding GPS data. Using the collected dataset as ground-truth data, the performance of the method presented in this paper and other popular methods is verified and compared. The results indicate that the GRU-based method has a better performance, with a precision, recall, and *F* score of 90.5%. Taking *F* score as an example, the outcome of the GRU-based method is about 6% to 7% higher than methods based on other machine learning algorithms. Considering the identification accuracy and model training time comprehensively, the method suggested in this paper outperforms the other three deep learning-based methods, namely, recurrent neural network (RNN), long short-term memory network (LSTM), and bi-directional long short-term memory network (Bi-LSTM). This study may provide some insights for the application and development of cellular signaling-based travel information collection technology for residents in the future.

1. Introduction

Understanding travel behaviors is important for urban transportation planning and management. A traditional data collection method is to organize residents' travel surveys, which include face-to-face household surveys, or questionnaires completed by telephone, e-mail, and web online. These kinds of approaches often lead to some defects, which typically include the huge implementation costs, an uneven sampling rate, the relatively low response rate, and the poor data quality [1, 2]. Since the early 2000s, scholars have developed several methods to collect residents' travel information based on GPS data. Compared to traditional manual surveys, GPS-based methods can reduce the response burden and investigation cost. GPS-based travel information also trends to be more accurate and with more details [3]. However, this method requires participants to take a specific GPS recording device along or install GPS recording software on their mobile phones, resulting in a series of issues such as data privacy, increasing implementation cost, or higher mobile Internet communication costs. These shortcomings limit the implementation scale of GPS-based methods.

During the past decades, with the growth and spread of wireless communication services, the estimation of residents' mobility and urban travel characteristics according to

cellular signaling data has attracted wide attention. Cellular signaling data are a kind of passive tracking data. They are generated when a cellular phone is linked to a communication base station due to various communication services, such as turning on/off, making calls, sending text messages, or connecting to the mobile Internet. Thus, the data can register the phone user's trajectories in the base station networks. On the basis of the potential mapping relationship between the real trip trajectories and the connection sequences of base stations, inferring the user's trip details using cellular signaling data can be expected [4, 5]. Compared with travel survey and GPS data, cellular signaling data have the advantages of 24-hour uninterrupted collection, wide spatial coverage, and high sampling rate. Meanwhile, the operators need not install additional acquisition equipment, so it can achieve relatively low data collection cost. Based on these technical advantages, scholars have conducted extensive research on travel information identification methods based on cellular signaling data (refer to Literature Review section).

This study proposes a travel mode identification method based on the gated recurrent unit (GRU) neural network. With 24 features as input, the method can identify four common and typical travel modes, containing walking, cycling, cars, and buses. Furthermore, we design and conduct field data collection experiments to collect labeled ground-truth datasets, which are finally provided by the local communication operator. In the experiments, participants' GPS data and trip diaries are simultaneously collected to check and label the cellular signaling data. Finally, we use the collected ground-truth datasets to verify and compare the classification ability of the suggested method and other methods based on machine learning or deep learning algorithms in travel mode recognition, including random forest, support vector machine (SVM), BP neural network, recurrent neural network (RNN), long short-term memory network (LSTM), and bidirectional long short-term memory network (Bi-LSTM).

The paper is arranged as follows: after introducing the identification procedure and GRU neural network model in Sections 2 and 3, this paper describes the data collection experiment and analyzes the temporal-spatial characteristics. Section 4 describes the model parameters and the verification results of the identification methods, while Section 5 describes the discussion and conclusion. The data are used with the permission of the volunteers, and no data privacy issues occur.

2. Literature Review

In the early 2G or 3G environment, only a few studies explored the travel information extraction method. Intel employed a method to estimate the moving speed of a cellular phone by observing the change of GSM signal intensity or the frequency of cell area transitions. Based on the assumption that the speed falls within a certain range for a specific transportation mode, they could infer the phone user's travel mode. However, their method was limited to classifying the travel modes that are easy to detect, including

stationary, walking, or making a drive [6]. Wang et al. utilized a k-means clustering method to separate the travel time of all trips into several groups for a given OD based on anonymous cellular signaling data, in order to estimate the percentage of travelers using different travel modes [7]. Overall, in an early 2G or 3G wireless communication environment, the location frequency of the cellular signaling data is fairly low, so it is difficult to recognize the individual traffic modes.

Since 2015, scholars have realized that cellular signaling data have good potential to distinguish between road trips and rail trips. Wireless communication base stations are generally arranged along both sides of roads or tracks, so there is a mapping relationship between phone users' base station connection sequences and traffic facility networks. Larijani et al. proposed a rail trip identification method based on rule-based heuristics (RBH), which can identify the inbound and outbound stations and travel paths. They also developed an APP to help passengers plan travel routes [8]. Thomas Holleczek et al. suggested a similar procedure and organized a manual survey at Orchard Station in Singapore to validate the accuracy. The outcomes indicated that the proposed approach's identification errors for the number of people entering or leaving the station per hour are both approximately 9.5% [9]. Horn et al. proposed a method to extract railway travel mode and departure time based on cellular signaling data. After comparing with the train operating data of the railway department, the identification error of the train departure time using this method is less than 5 minutes [10]. Hasan Poonawala et al. presented a model to identify road trips or rail trips, combining the hidden Markov model with the topological properties of different traffic networks [11]. Yamada et al. of Osaka University also focused on using the speed characteristics of cellular signaling data and traffic facility network data to distinguish road trips or rail trips. They further proposed a simulation model to verify the identification accuracy [12]. These methods rely on speed as the main indicator, which makes it difficult to differentiate between transportation modes that have similar speed profiles, such as buses and cars. Moreover, the current frequency and accuracy of cellular signaling data are insufficient to capture the subtle variations in speed that could help distinguish between these modes.

In recent years, with improvements in location frequency and accuracy, some researchers have begun to extract residents' multiclass travel modes from mobile phone data. Combining cellular signaling data with traffic facility network data, Qu et al. put forward a mode split model applying RBH with logit model to identify walking, cars, and buses [13]. However, the study only compared the percentage of different travel modes with the real mode shared data obtained from the US census, which cannot fully explain how accurately the model can identify individual travel modes. Danafar et al. uses the Bayesian probability method to identify walking, cycling, cars, and public transportation (bus, subway, and tram) [14]. However, this study has not verified the accuracy of the method for individual travel mode recognition. On the basis of 4G cellular signaling data,

Kimberley et al. proposed two supervised methods, RBH with random forest (RF) and RBH with a fuzzy logic model, and an unsupervised method combining RBH with k-medoid clustering, to identify multiclass transport modes, involving walking, cycling, car, metro, train, and tramcar. To verify the accuracy of the algorithms, two simultaneous collection experiments of cellular signaling data and GPS data were conducted in Switzerland. The evaluation results indicate that the complex model that combines RBH and RF outperforms the other methods, achieving a differentiating accuracy of 73% [15]. This research is a rare empirical study on the identification accuracy of fine-grained travel modes.

Despite the evidence from previous studies on the feasibility of extracting fine-grained travel modes from cellular signaling data, some challenges remain unresolved. First, the evolution of mobile communication technology has significantly improved the positioning quality of cellular signaling data. As described in subsequent sections, the current average location frequency of cellular signaling data can reach a level of less than 60 seconds, which is much higher than that of cellular signaling data in the early 2G or 3G era. Thus, more studies are required to ascertain the extent to which the accuracy of detecting fine-grained travel modes can be enhanced by this high-frequency cellular signaling data. Second, limited by privacy policy, it is difficult to obtain personal cellular signaling data, which poses great difficulties for technical verification. Therefore, the identification accuracy of fine-grained travel mode using different types of methods in real 4G-LTE or 5G wireless communication environment remains to be fully verified.

Furthermore, a deep learning method has already been broadly and successfully employed in the field of travel information extraction or prediction. Petersen et al. merged a convolutional layer and a long short-term memory layer into a new deep neural network to predict bus travel time. The model outperformed other methods the authors compared with, including historical average model, pure LSTM, or Google Traffic, and could find the complicated patterns not discovered by the compared models [16]. Kim et al. proposed a long-term recurrent convolutional network to extract transportation modes utilizing GPS data. The modes are divided into walk, bike, driving, train, bus, and electric mobility scooter. The validation results displayed that the proposed method has a better performance than other methods from existing studies [17]. Wang et al. presented a transportation mode recognition model based on a residual and LSTM recurrent networks, which utilized several kinds of light-weight sensors internally installed in smartphones. The model introduced the residual units to improve the model's learning efficiency and enhance the detection performance of different transportation modes. The recognition model has been extensively validated and found to achieve the highest recognition accuracy for eight transportation modes [18]. In summary, existing research has demonstrated that deep learning algorithms can achieve high accuracy and robustness in travel information recognition or prediction fields based on GPS data or other high spatial-temporal granularity data. However, in the field of trip information recognition based on cellular signaling data, which location quality is relatively irregular and whether the deep learning algorithms can maintain this advantage need to be further proved.

3. Methods

3.1. Overview. When the phone users travel in the city and keep connections with communication base stations, their trajectories can be recorded by the wireless communication network completely through the cellular signaling data. Figure 1 illustrates the overview of the main steps for travel mode identification and the key objectives of this paper. For raw data, a preprocessing procedure should be first conducted to reduce the impact of the noise data. Second, through a trip and identification method, the data of each user are segmented into several single trips, each of which represent a moving trip between a pair of OD and contain a single transport. That means we identify the main travel mode of each trip. For example, if a trip is completed "walking-bus-walking," we consider this trip's corresponding mode to be the bus. The trip end identification method designed with the same datasets using in this paper was described in the literature [19]. Thus, this paper focuses on the following steps: identifying the travel mode of each single trip segment. Third, for each moving trip, a variety of temporal and spatial features are extracted from the trajectory and trip characteristics, which are used as input for travel mode identification. Finally, taking advantage of GRU neural network in processing data with time series and indefinite length, a deep learning-based model is established to identify the travel mode corresponding to each trip, including walking, bicycle, car, and bus. Furthermore, a ground-truth dataset is used to validate the classification ability of the proposed method and to compare the differences in accuracy and efficiency between the proposed model and other machine learning-based or deep learning-based models.

3.2. Preprocessing: Data Cleaning. Raw cellular signaling data usually contain noise data due to wireless communication disturbances or data transmission errors. Drifting data and oscillation data are the most common types of noise data.

(1) Drifting data. Drift is a phenomenon, where a mobile phone abruptly switches to connect to a faraway base station during its continuous connection to the wireless communication network. For drifting data, we used a speed-based method to eliminate noise data. Initially, the shifting speed between two successive data l_i and l_{i+1} is computed. When the speed exceeds a threshold V_d, data l_{i+1} is marked as a possible outlier. Then, we compare the Euclidean distance of data l_i to l_{i+1}, d_{i-i+1}, and the distance of data l_i to l_{i+2}, d_{i-i+2}. If the d_{i-i+1} is greater than d_{i-i+2}, the data l_{i+1} is removed as a drifting data.

(2) Oscillation data. Oscillation, or the ping-pong effect, refers to the phenomenon of a mobile phone signal switching frequently among several base stations, leading to adjacent data exhibiting a handoff pattern such as "1-2-1" or "1-2-3-1." For oscillation data, a pattern-based method is
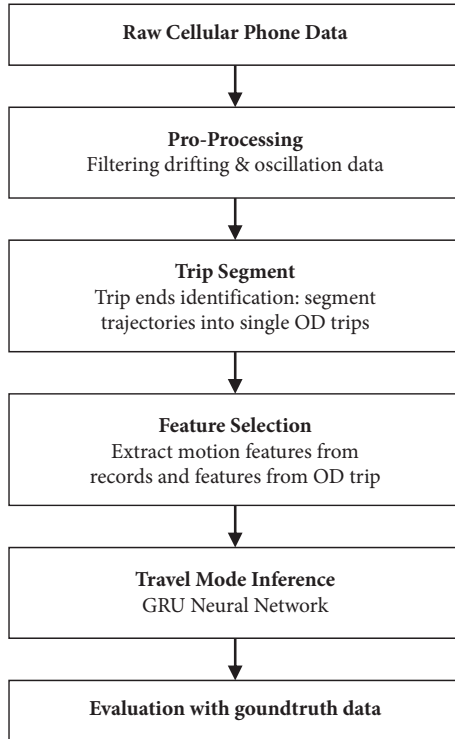
Figure 1: Overview of the methodology.

introduced to remove noise data. When the adjacent data show the oscillation pattern "1-2-1" or "1-2-3-1" and the time interval between the first and the last data is shorter than the threshold T_o, only the first and the last data are kept and the rest are deleted. On the basis of repeated tests, V_d and T_o are separately set to 200 km/h and 150 s.

After removing the drifting and oscillation data, we further processed the duplicated data. When the phone user generates intensive communication behavior, several cellular signaling data may be continuously generated on the same base station. These data have the same coordinates and the handover speed is all 0 km/h, which may interfere with the model's performance for distinguishing various travel modes. Thus, for duplicated data, we retained the first and the last data and removed the other ones.

3.3. Trip Segment. After preprocessing the individual cellular signaling data, the next step is to identify its trip ends. For this purpose, scholars have proposed a variety of identification methods, which can be primarily classified into two types of categories:

(1) *Rule-Based Methods.* These methods usually detect trip endpoints by comparing the spatial-temporal features of the cellular signaling trajectories, which can be different in the two states of staying or moving. As a large-scale dataset, the most direct and efficient method to process the cellular signaling data is setting some simple filtering rules, including distance threshold or time threshold. Calabrese et al. suggested that the virtual central location formed by

consecutive points is the trip ends of the phone user when the coverage radius of the consecutive points is less than 1 km [20]. Wang et al. considered that if a phone user stays in a certain area for more than 15 minutes, the user is considered to be in a stay state [21]. Schlaich et al. set a similar time threshold as 60 minutes [22]. Ni et al. regarded a group of continuous trajectory points that satisfy the spatial distance less than 200 m and the duration longer than 30 minutes as stay points [23]. The values of the thresholds should consider the communication network characteristics of the research area as much as possible, and their rationality relies heavily on the subjective experience of the researchers. Therefore, the parameters proposed in one study may not be easily applicable to another city, which hinders the widespread promotion and application.

(2) *Clustering-Based Methods.* These methods mainly use the differences in the shape, volume, or density of the cellular trajectory clusters in the moving or staying state to identify trip ends. Clustering algorithms are usually unsupervised algorithms, which can avoid the influence of researchers' subjective experience to a certain extent. Chen et al. employed a clustering method based on a statistic model for extracting clusters, which does not require a pre-specified number of clusters. Subsequently, to distinguish between true activity locations and stay points during movement (traffic jams and waiting for buses), they used a logistic regression model with two explanatory variables (a shape variable and a volume variable) to extract the true activity locations [24]. Jiang et al. proposed an improved DBSCAN method to identify trip ends. First, they used a genetic algorithm to optimize the clustering radius under different base station densities, and obtained a series of optimal parameters related to the base station densities. Second, when using DBSCAN to process cellular phone points into clusters, the proper clustering radius is selected according to points' surrounding BS densities, thereby reducing the identification error that may be caused by the fixed parameters [25]. However, clustering-based methods suffer from low model efficiency due to the large amount of distance calculation between trajectory points during execution. Moreover, these methods face challenges in deploying on distributed computing servers, which limits their applicability to large-scale (such as city-wide) datasets.

To further enhance the accuracy and robustness of the trip end identification method, we developed a model based on the random forest algorithm, which leveraged the powerful performance of machine learning algorithms in pattern recognition. The model details and procedures were reported by Yang et al. [19]. First, we enriched each cellular signaling data with four types of feature attributes, and incorporated external data (POI) to increase the distinction of feature attributes between the "moving" and "staying"

states. Then, we built a random forest model and optimized the model parameters using methods such as cross-validation. Finally, we validated the precision and recall of our proposed model utilizing the same ground-truth data used in this research. The results showed that our model outperformed rule-based methods, clustering-based methods, and three other machine learning algorithms in terms of overall identification performance. Moreover, the proposed method could continuously adapt to the identification objects and improve the identification accuracy as more input data were available. Furthermore, our method could be implemented in a distributed computing environment, which made it suitable for analyzing travel information and urban travel characteristics from a large-scale dataset.

After the identification of the trip ends, each user's cellular trajectory can be segmented into several single trips, each of which represent a moving trip between a pair of ODs and contain a single travel mode. The following step, as well as the focus of this paper, is to identify the corresponding travel mode of each single trip.

3.4. Feature Selection.
Features can be used to describe the differences in the trajectories of cellular signaling data between different travel modes, which are usually calculated by the physical characteristics of the trajectories. The choice of the feature parameters has a significant influence on the model identification performance. Based on the generation principle, we select two types of features: motion features from the cellular records and features from OD trips, which contain 24 specific feature parameters.

3.4.1. Motion Features of Cellular Records.
First, 21 motion features are calculated directly from the adjacent cellular signaling data records of users or records in specific time windows, such as average distance, speed, or time. These features reflect the motion and trajectory differences in the wireless communication network within the same time range caused by the different travel modes' moving speed. The location coordinates of cellular signaling data are approximately replaced by the coordinates of the communication base station, which means that it cannot directly reflect the user's activity trajectory. However, the differences in switching rate and frequency between communication base stations are strongly related to the actual differences in moving speed or frequency when the phone users adopts different travel modes. Table 1 shows nine types of features extracted from the cellular signaling data records, including 21 specific features. Figure 2 visually shows the difference between linear distance ZD_T and cumulative distance LD_T calculation.

3.4.2. Features from OD Trip.
In the daily travel of residents, the distance and time information from the origin to the destination has a major influence on the choice behavior of travel modes. When the trip distance is long, residents are usually more likely to choose cars or buses. When the travel

distance is short, residents tend to prefer convenient transportation modes such as walking or cycling. Even if some transportation modes may have very similar speed characteristics in some congested sections, there are still significant differences in total travel distance, travel time, or travel speed from a comprehensive perspective. Therefore, when identifying transportation modes for each trip, adding travel information between the origins to the destination is expected to increase the accuracy of travel mode identification. Therefore, this paper selects three characteristics between ODs for each trip, including the Euclidean distance D_{OD}, Euclidean distance D_{OD}, and the average speed V_{OD}.

In summary, this paper selects 24 features as input parameters of the GRU neural network model. These features include physical quantities such as distance, speed, and time with different dimensions or units. In order to prevent different dimensions or orders of magnitude from affecting the accuracy of model training, this paper employs a Z score standardization method to normalize all features. The Z score standardization method utilizes the mean and standard deviation of the original data for standardization processing, and the processed data follows a standard normal distribution. The Z score standardisation method is shown in (1), where X^* denotes the normalized characteristic value, X represents the original characteristic value, and μ and σ represent the mean and standard deviation of all samples, respectively.

$$X^* = \frac{X - \mu}{\sigma}. \tag{1}$$

3.5. Gated Recurrent Unit Networks.
From the perspective of machine learning, the detection of travel modes from cellular signaling data has the following characteristics: (1) Basically, it is a typical "many-to-one" classification problem in the domain of pattern recognition, that is, judging which travel mode belongs to a single trip (including several pieces of data). (2) The cellular signaling data sequence corresponding to a trip clearly has time series features and the length of the sequence is uncertain. (3) Trajectories generated by different travel modes have significant differences in speed, distance, base station connection frequency, and other features. These characteristics have some similarities with the characteristics of pattern recognition problems such as speech recognition and text classification. In deep learning algorithms, the GRU neural network is a typical neural network structure that is able to process data with time series features or serialized data. It can also process data types with indefinite length and has achieved successful applications in complex pattern recognition fields, for instance, the computer vision or the natural language processing [26, 27]. Therefore, drawing on the successful experience of GRU in the above fields, we attempt to introduce it to solve the problem of traffic mode identification according to cellular signaling data.

In 2014, Cho Kyunghyun from New York University introduced a neural network model called GRU (gated recurrent unit). The GRU neural network can be regarded as

TABLE 1: Motion features from cellular signaling data records.

| Motion features	Description
D_b	Euclidean distance between the coordinates of two adjacent data
T_b	The difference between two adjacent data timestamps
V_b	Euclidean distance of two adjacent data divided by the time difference
LD_T	For data l_i generated at t_i, the distance of adjacent data is calculated and summed as the cumulative distance LD_T in the time window $[t_i - 0.5T, t_i + 0.5T]$. T is set to 5 min, 7 min, 9 min, and 11 min, respectively, so we obtain 4 features
ZD_T	For data l_i generated at t_i, the distance between the first and last data is calculated as the linear distance ZD_T in time window $[t_i - 0.5T, t_i + 0.5T]$. T is set to 5 min, 7 min, 9 min, and 11 min
V_{aveLDT}	Cumulative distance LD_T divided by the time interval from the first data to the last data in time window $[t_i - 0.5T, t_i + 0.5T]$. T is set to 5 min, 7 min, 9 min, and 11 min, respectively
V_{aveZDT}	Linear distance ZD_T divided by the time interval from the first data to the last data in time window $[t_i - 0.5T, t_i + 0.5T]$. T is set to 5 min, 7 min, 9 min, and 11 min
n	For a trip between a pair of OD, n represents the amount of distinct base stations connected in the whole trip
f	The amount of distinct communication base stations connected, n, divided by the total travel time of the trip
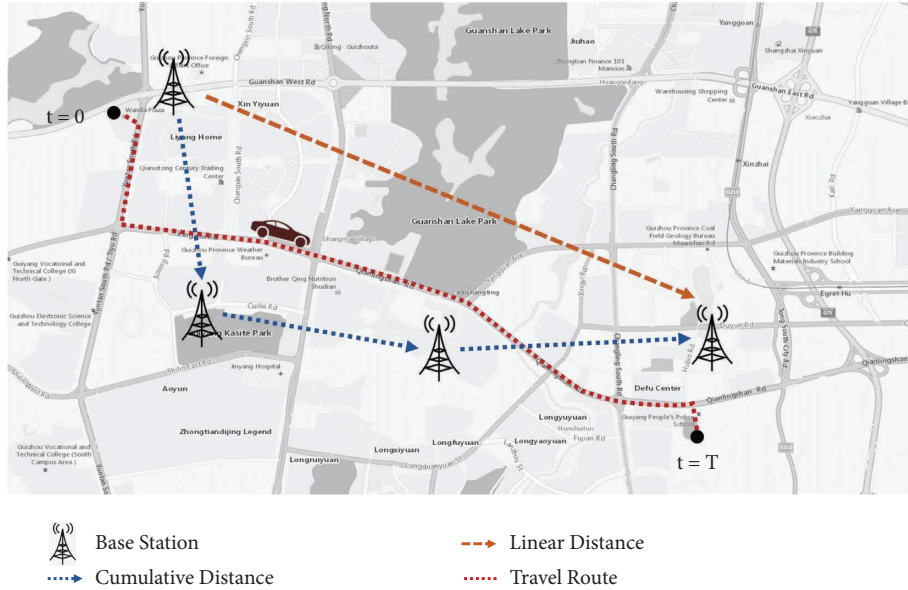
FIGURE 2: Illustration of linear distance and cumulative distance.

a simplified model of LSTM, which preserves the ability of LSTM to integrate long-term and short-term memory, but reduces the complexity of the cell structure, the amount of parameters, and the training time. The main simplification of GRU is to merge the forget gate and the input gate in LSTM into a new update gate. Figure 3 illustrates the structure of the GRU neural network [28].

From an external structure perspective, the input and output structures of the GRU neural network are similar to those of the ordinary RNN model. Each unit inputs two variables and outputs two variables. In Figure 3, x_t donates the input at the current time, C_t is the hidden layer state, y_t represents the output at the current time, and C donates the GRU structure. It can be seen that the hidden state C_t at time t relies not only on its corresponding input data x_t but also on the hidden state C_{t-1} at the prior time, as shown in (2). U and W are the weight coefficients between different network components.

$$C_t = f(Ux_t + Wc_{t-1}). \quad (2)$$

The internal structure of the GRU neural network is displayed in Figure 4 [29]. The GRU model simplifies the internal neurons into two gate structures: update gate and reset gate. In this figure, x_t donates the input of the neuron, y_t is the output of the neuron, Z_t donates the GRU output of the update gate, r_t is the output of the reset gate of GRU, and $\overline{h_t}$ represents the candidate hidden state at the current time. σ represents the sigmoid activation function.
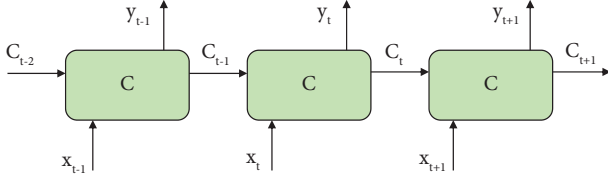
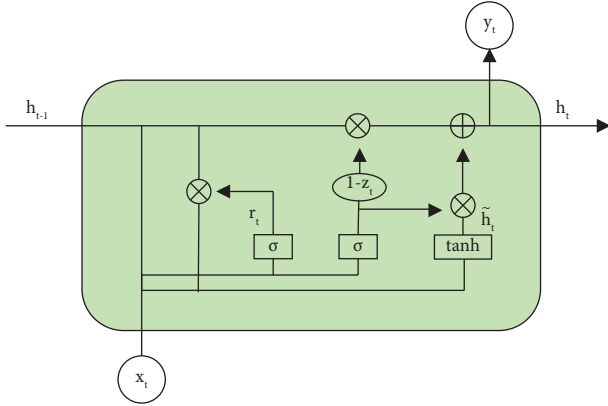FIGURE 3: The structure of the GRU neural network.

FIGURE 4: The GRU neuron structure.

As shown in equations (3)–(6), the update gate Z_t is formed by multiplying a weight matrix with a concatenation vector of the prior hidden state h_{t-1} and x_t. Then, the sigmoid activation function is applied to transform the elements in this vector into real numbers in the [0, 1] range, and this vector serves as the gate control state of the update gate. The reset gate r_t is similar to the update gate, but uses the parameter weight W_r of the reset gate. The candidate hidden state multiplies the result of applying the reset gate state value to the h_{t-1} vector with x_t and concatenates it with x_t. The concatenated vector is converted into a vector of real numbers between −1 and 1 using the tanh function. When outputting information, GRU applies update gates to h_{t-1} and candidate hidden states, respectively, and sums them up. Then the result is used as output information for the current state. The abovementioned analysis shows that each neuron in GRU participates in the decision-making process for each information output, creating dependencies among the neurons. In general, reset gates are more active for short-term dependencies, while update gates are more active for long-term dependencies [25].

$$z_t = \sigma\left(W_z \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_z\right), \tag{3}$$

$$r_t = \sigma\left(W_r \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_r\right), \tag{4}$$

$$\widetilde{h}_t = \tan h\left(W \cdot \begin{bmatrix} r_t \otimes h_{t-1} \\ x_t \end{bmatrix} + b_h\right), \tag{5}$$

$$h_t = (1 - z_t) \otimes h_{t-1} \oplus z_t \otimes \widetilde{h}_t. \tag{6}$$

3.6. Model Construction. Figure 5 illustrates the travel mode identification model based on GRU, which consists of an input layer, a GRU layer, a fully connected layer, and an output layer. First, the features of each trip segment are computed based on the cellular trajectory points and fed into the deep learning model as inputs. For a trip composed of n cellular signaling data points, 24 corresponding features are computed for each point, transforming an n-dimensional vector into an $n \times 24$ matrix. The $n \times 24$ feature matrix is employed as the input of the neural network and trained on the GRU layer. The GRU layer can not only perform model classification based on the input attribute values at a single time point, but also effectively capture the correlation between longer sequence feature values prior to the current time point, which can better handle data with temporal dependencies. The output of the GRU layer is served as the input for two fully connected neural networks, and finally, the output of the fully connected layer is taken as the input for the output layer. The output layer converts all output values from the fully connected layer into probability values between 0 and 1 through a sigmoid function and outputs the model results at the last node. The output layer selects the travel mode with the longest cumulative time as the final mode for the trip.

The GRU model is trained by minimizing the loss function in the training set. Travel mode identification is a typical multiclassification problem, so it is more appropriate to choose a multiclass cross-entropy loss function, as shown in equation (7). L represents the loss function, and we take minimizing the loss function as the model training objective. Specifically, the calculated loss value in the equation represents the error value between the probability distribution of the neural network output and the real probability distribution of the label. The model is trained by minimizing this error value. X represents the input sample, and Y represents the result output by the neural network. $P_{i,m}$ represents the probability that the i-th input sample is predicted as the m-th category. In this paper, m represents four travel modes. $Y_{i,m}$ represents whether the m-th category is the true category of the input sample x_i. If it is true, it is 1; otherwise, it is 0.

$$L(Y, P(Y \mid X)) = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{m=0}^{M-1} y_{i,m} \log p_{i,m}. \tag{7}$$

4. Experiments and Data

4.1. Data Collection. To obtain cellular signaling data and their corresponding real travel information, the research team conducted a synchronized data collection experiment in a southwestern city of China. The experiment was carried out from September 7 to September 12, September 20 to September 25, and December 15 to December 22 in 2019, totaling 19 days. More than 25 volunteers carried out purposeful activities and travelled throughout the city according to the planned route and plan. To minimize the sample bias of the collected data, the design of travel plan considers multiple factors, including purpose of activity, destination
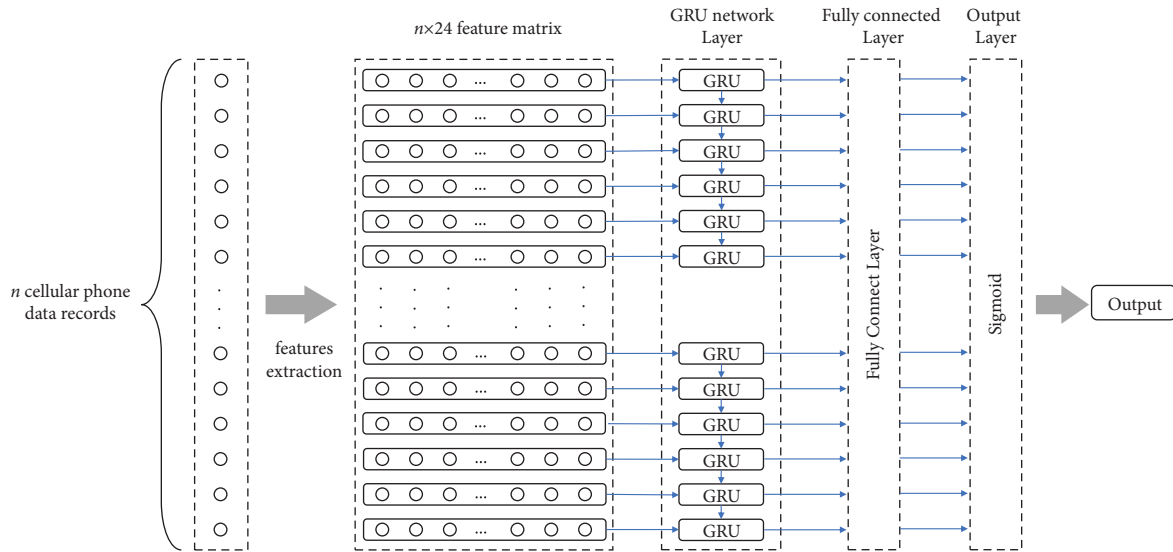
FIGURE 5: The structure of the GRU-based model for recognizing travel mode.

type, as well as travel mode. Travel purposes include going to work, going to school, seeking medical treatment, dining, entertainment, shopping, leisure, and returning home. The stay areas cover different areas with different base station densities, such as urban areas and suburban areas. Travel modes include walking, nonmotorized vehicles, cars, and buses, which are commonly used in this city.

All volunteers received formal training and participate in a re-experiment after the training to ensure that they can proficiently complete the data collection tasks according to the plan in the formal experiment. In the formal experiment, each volunteer carried a mobile phone with a SIM card from the local operator, and the phone had a GPS data collection application installed. During the experiment, the GPS data collection application remained on, and the volunteers recorded detailed travel logs, including activity locations, arrival/departure times, and travel modes. In future research, GPS data and travel logs can be used to determine the real travel status corresponding to each mobile phone signal data. The GPS data recording APP and samples of travel logs are shown in Figure 6.

With the consent of volunteers who signed confidentiality and authorization agreements, the operators provided cellular signaling data for all volunteers during the experiment, which provided a rare opportunity for this study to obtain users' cellular signaling data and corresponding real travel information. During the experiment, a total of 179377 pieces of cellular signaling data were generated by all volunteers who collected more than 200 pieces of real travel chain information. Table 2 displays the typical fields of cellular signaling data. Global Identifier and User ID can both act as the unique identity code for each phone user. By combining LAC (location area code) and CI (Cell ID), the identity code of each base station can be determined. Based on the start time and end time, the duration of the communication service can be calculated easily. The longitude and latitude in the table represent the location of the base

station that was connected when the communication service was generated.

Based on cellular signaling data, this paper focuses on identifying four kinds of travel modes: walking, cycling, car, and bus. After further extraction and screening of the dataset for these four modes, 620 travel segments were collected in this experiment, including 69,059 cellular signaling data. After preprocessing, the total sample size was 62,020. Considering that deep learning algorithms require a higher number of data samples, a sliding time window-based sample construction method is used to process trips with more than 60 cellular location records. The sliding window length is set to a minimum of 50 cellular location records with an increment of 10, while a moving step size is set to 10 location records. The final data sample set was obtained by randomly oversampling the minority class samples to improve the balance of the data samples. Table 3 displays the sample size corresponding to each travel mode used for model training. In one-hot encoding form, we added a travel mode label to the cellular signaling data corresponding to each trip.

4.2. Data Characteristics. The location quality of cellular signaling data is influenced by the location frequency, which is determined by the frequency of communication services generated by users. Using the cellular signaling dataset collected in the synchronized data collection experiment, we conducted a statistical analysis of location frequency. Overall, each volunteer generated an average of 1425 cellular signaling data per day. As shown in Figure 7, the probability that the time interval between adjacent data is less than 30 seconds exceeds 70%, with an average time interval of 48 seconds and a median of 20 seconds. Compared with early signaling data [30], the location frequency of cellular signaling data in the 4G environment has increased significantly, providing possibilities for inferring multicategory and fine-grained travel modes.

Status	Condition	Time
Stay	Leave Home	9:02:55
Bus	Walk to bus station	9:07:30
Bus	Get on	9:17:28
Bus	Get off	9:28:17
Metro	Get on	9:35:51
Metro	Get off	10:06:45
Stay	Walk to Museum	10:23:16
Stay	Leave Museum	11:52:34
Car	Get on	11:59:21
Car	Get off	12:05:56
Stay	Arrived at restaurant	12:06:07
Stay	Leave restaurant	14:08:45

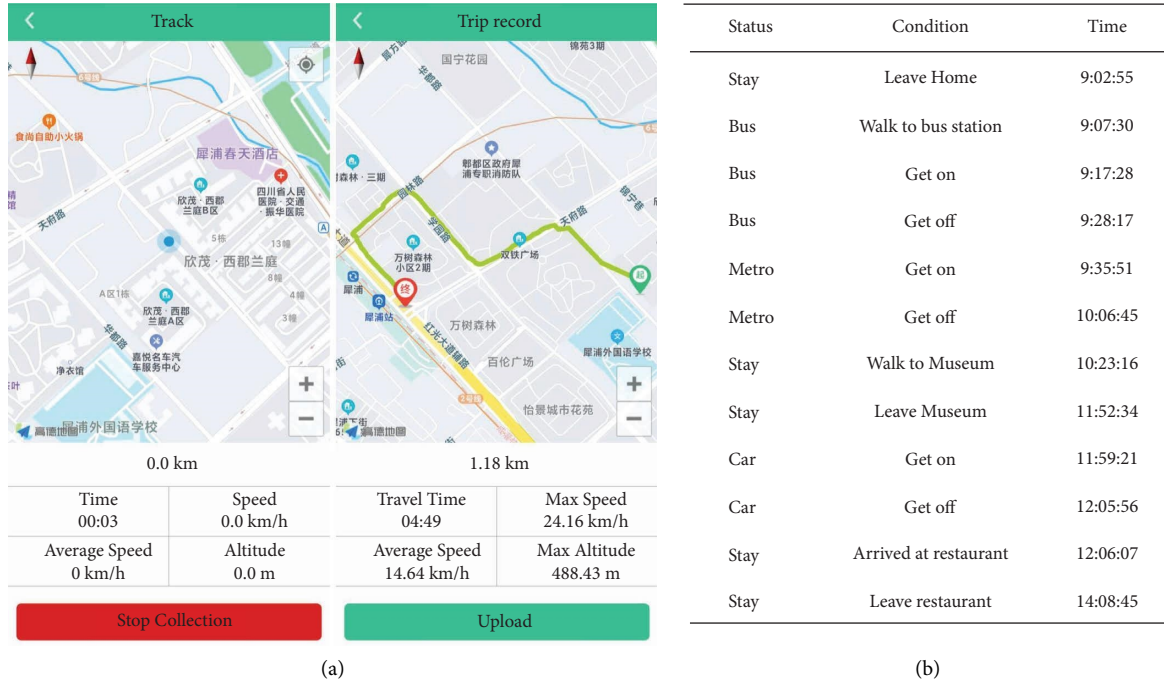(a)   (b)

FIGURE 6: The illustration of GPS data recording APP and travel log data. (a) GPS data recorded by an APP. (b) Travel log data recorded manually.

TABLE 2: Samples of cellular signaling data.

Global identifier	User ID	LAC	CI
460****71	130****4926	34051	167939598
460****71	130****4926	34050	168004374
460****71	130****4926	34051	167936011

Start time	End time	Longitude (°)	Latitude (°)
12:54:18	12:54:18	106.713497	26.601376
12:54:20	12:54:20	106.743359	26.592555
12:54:21	12:54:21	106.716586	26.613950

TABLE 3: The number of trips for training and testing the recognition model.

Mode	Total number	Training set	Test set	One-hot encoding
Walking	654	458	196	[1000]
Cycling	628	440	188	[0100]
Cars	627	439	188	[0010]
Buses	660	462	198	[0001]

The accuracy is another factor that affects the location quality of cellular signaling data, which can be described by the distance error between the real coordinates and the coordinates of the base station. In the data collection experiment, cellular signaling data and GPS data are collected simultaneously. When matched to the data generation time, a total of 97,267 cellular signaling data were successfully matched to GPS locations at the same time. GPS data can represent the user's real location coordinates, while cellular signaling data can reflect the location coordinates of the corresponding communication base station. A statistical analysis of the distance error between these two types of posited data is conducted. As shown in Figure 8, more than 53% of the location errors are within 300 m and more than 73% of the location errors are within 500 m. The average value of location error is 357 m and the median is 278 m.

5. Result Analysis and Discussion

5.1. Model Specification. The GRU neural network model training was completed under the TensorFlow 2.2.0 deep learning framework installed in Python 3.9.7. The processor used for the model training environment is Intel® Core i5-7200U @2.5 GHz, with a memory capacity of 4G and an operating system of Window 10. The NVIDIA GeForce 940MX graphics card with 2G video memory was used for training.

As a complex deep learning algorithm, the GRU neural network contains a large amount of hyperparameters that affect the deep learning-based models' classification accuracy. To enhance the recognition and generalization ability of the model, we introduce multiple parameter optimization strategies during model training and testing. First, we start by dividing the dataset into two parts: a training set and a test set. 70% of the data samples were randomly selected as the training set, and the remaining 30% were used as the test set. The training set is applied to train the deep learning model and adjust its parameters, while the test set is only employed to test the generalization ability of the final model. Second, during the model training process, we introduce a five-fold cross-validation strategy. The dataset used for training the model is randomly divided to five parts, with four parts served as a training set and one part served as a validation
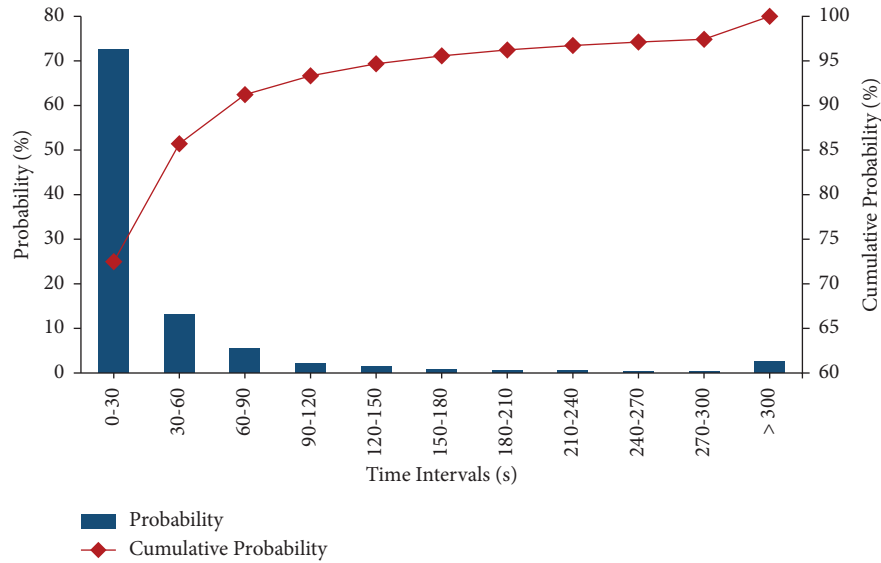
FIGURE 7: Probability distribution of the time intervals of adjacent cellular signaling data.
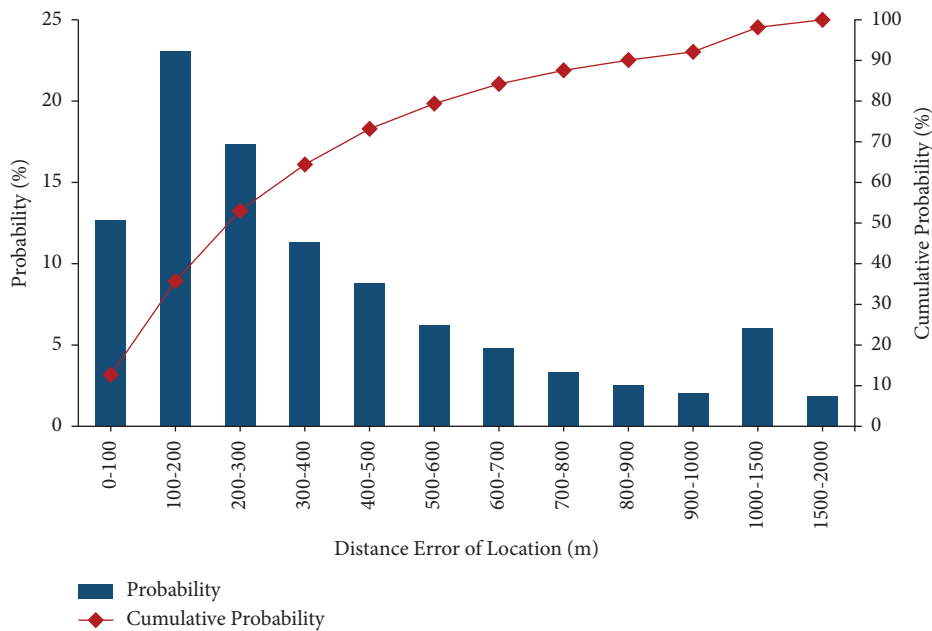

FIGURE 8: Probability distribution of locations' distance error.

set. This results in five well-trained models. After all models' loss functions converge, we select the model with the lowest loss value as the best model. Moreover, to prevent overfitting during the training process, we add a dropout strategy to the fully connected layer of the model. During training, the model randomly ignores some neuron information so that it does not rely too much on local features, thereby making the model's generalization ability stronger. The dropout rate value is set to 0.5. The final parameter settings of the model are displayed in Table 4.

The training set was applied to construct the GRU neural network model with these parameters, and the model loss curve during training is demonstrated in Figure 9.

TABLE 4: The model parameters of the GRU neural network.

Parameter name	Values
Number of layers	3
Number of units	[128, 64, 128]
Batch_size	32
Epochs	500
Initial learning rate	0.001

The model loss decreases rapidly as the number of training times increases, and reaches a minimum value of 0.59 when the training rounds are 60. The model has some degree of overfitting at this time. Then, as the model continues to train,
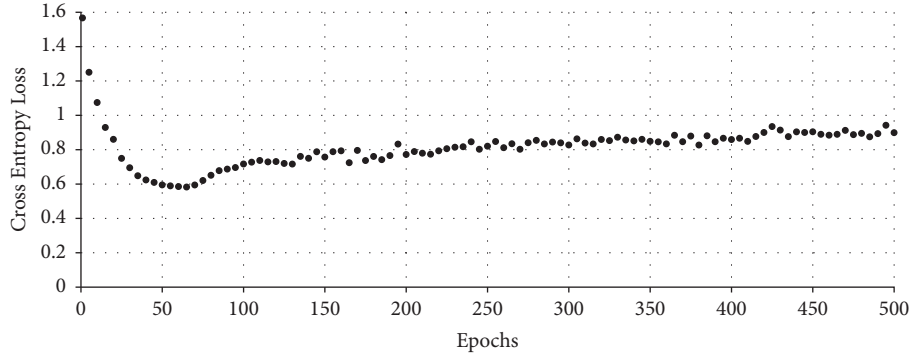
FIGURE 9: The loss reduction process in model training.

the model loss slightly increases and stabilizes around 0.9, which indicates the best overall accuracy and generalization ability of the model. The model training took 18 minutes and 1 second in total.

5.2. The Performance of Identifying Travel Modes. Travel mode identification is a complex multiclassification problem. To evaluate the model's classification ability, the identification results were categorized into three groups: True Positive (TP), False Negative (FN), and False Positive (FP). TP represents the right part of all identified travel modes, while FN denotes the real travel modes that were not detected, which can be viewed as the missed part. Similarly, FP refers to the travel modes that were found but did not match with real samples, which can be viewed as the incorrect part. Subsequently, for the purpose of comparing the overall performance of different recognition methods, three indicators of precision, recall, and *F* score were introduced as model assessment indicators. As shown in equations (8)–(10), precision is the correctly recognized samples of a certain travel mode to the entire quantity of samples recognized as that mode. Recall is calculated as the ratio of the correctly recognized samples of a certain travel mode by the model to the number of actual samples of that mode. *F* score is a weighted harmonic mean of precision and recall and can more comprehensively reflect the model's classification ability.

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{9}$$

$$F - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{10}$$

Table 5 demonstrates the travel mode identification results of the test set. The test set contains 770 trips, corresponding to four travel modes. Walking has the highest precision and recall, which are 97.9% and 95.9%, respectively. This is mainly due to the lowest moving speed and achievable travel distance of walking, which leads to more obvious differences in most features from the other three

travel modes, making it easier to be recognized. Similarly, with higher average moving speed and longer travel distance, car has the second highest precision and recall, which are 94.4% and 93.4%, respectively. In contrast, the recognition performance of bicycles and buses is relatively poor, and these two modes are most likely to be misidentified. The main reason is that most roads in the city have dedicated lanes for nonmotor vehicles, and the average travel speed of bicycles under the exclusive road rights is close to that of buses. At the same time, both nonmotor vehicles and buses can cover short and medium distance trips in the city. These similarities lead to overlapping intervals in the calculation results of features such as distance and speed for cellular signaling data generated by nonmotor vehicles and buses, which in turn causes the model to easily confuse these two modes. In addition, among the motorized travel modes, there is a situation of misidentification between buses and cars. One possible reason for this is that the roads are more congested during peak hours in the morning and evening. Therefore, the speed and distance differences between the two travel modes are not obvious. This leads to errors in the recognition results. Due to the high misidentification rate between buses and nonmotor vehicles and cars, the precision and recall of bus mode are both 83.5%, which are the lowest among the four modes of transportation. Overall, the recognition model constructed in this paper has a positive performance for the four modes of transportation, and the precision, recall, and *F* score can reach 90.5%.

5.3. Comparison of Different Algorithms. We first compare the travel mode identification performance of the model based on the GRU neural network and models based on other classical machine learning algorithms, such as random forest, support vector machine (SVM), and BP neural network. Figure 10 displays the comparison result. It indicates that the recognition performance of the three machine learning algorithms is relatively close, with their *F* scores ranging from 83.1% to 85.2%. In comparison, the method based on the GRU neural network has a better recognition performance, and its *F* score is about 6% to 7% higher than that of machine learning methods. As a deep learning model, the GRU neural network has advantages such as more neurons, more complex hidden layers, and the

TABLE 5: Statistics of travel mode identification performance.

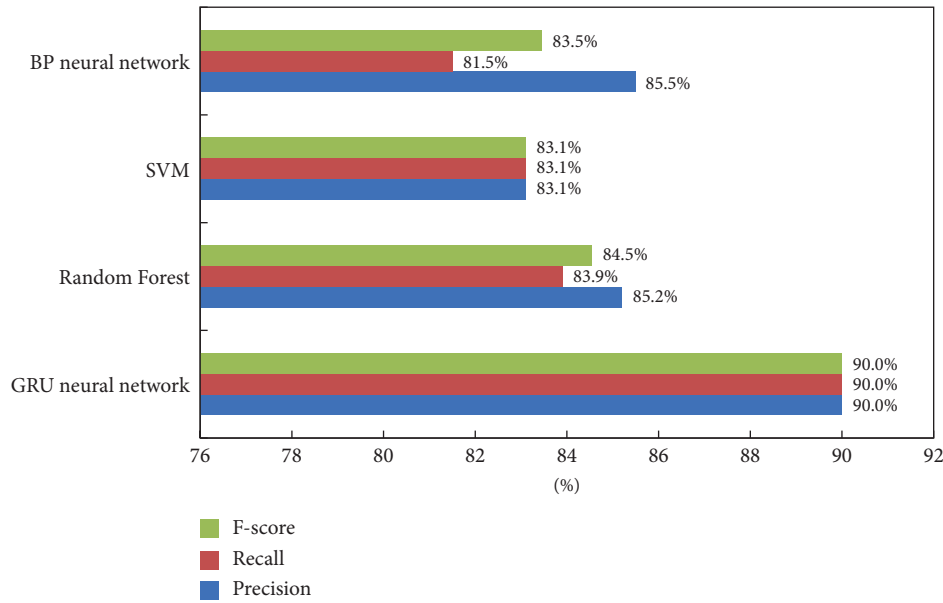| Mode	Number of trips	Identification results				Precision (%)	Recall (%)	F scores (%)
		Walking	Bicycle	Bus	Car			
Walking	196	186	5	5	0	94.9	94.9	94.9
Bicycle	188	6	165	17	0	87.3	87.8	87.5
Bus	188	4	16	157	11	83.1	83.5	83.3
Car	198	0	3	10	185	94.4	93.4	93.9
Total	770	196	189	189	196	90.0	90.0	90.0

FIGURE 10: Identification accuracy of different types of machine learning methods.
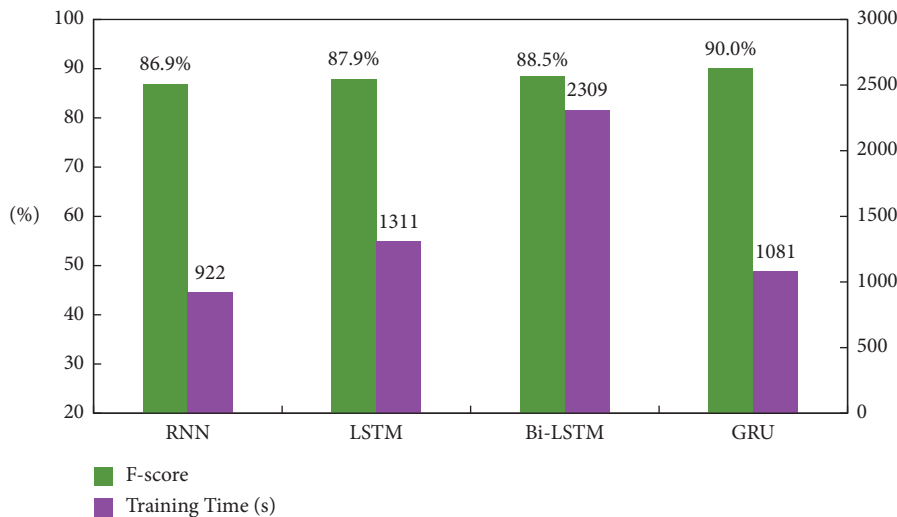

FIGURE 11: Identification accuracy and training time of different types of deep learning methods.

ability to use long-term or short-term features. Therefore, it has shown stronger fine-grained travel mode recognition capabilities.

In this paper, we further compare the accuracy and efficiency of various deep learning-based identification models in the fine-grained travel mode recognition task, including recurrent neural network (RNN), long short-term memory (LSTM), bidirectional long short-term memory (Bi-LSTM), and the GRU neural network proposed. Four models used the same training and test sets. During the model training process, similar parameter optimization strategies were adopted for all four models to ensure that

each model's parameters were fully optimized. The results are provided in Figure 11. It displays that the recognition ability of the four methods based on deep learning algorithms is not significantly different. Among them, the F score of the model based on GRU is the highest, reaching 90%, and the F score of the model based on RNN is the lowest, reaching 86.9%. The error between the two models is only 3.1%. However, the difference in model training time between the four methods is more obvious. The training time of the model based on RNN is 922 seconds, which is the shortest. The training time of the model based on Bi-LSTM is 2309 seconds, which is the longest and 2.5 times that of the model based on RNN. The training time of the model based on GRU is 1081 seconds, which is the second shortest among the four models and about 17% more than the shortest model training time. Therefore, in terms of model recognition accuracy and model efficiency, the model recognition model based on the GRU neural network has the best overall performance among the four deep learning models.

6. Conclusions

Using large-scale cellular signaling data to extract residents' travel information offers a potential opportunity for comprehensive, real-time, wide-area analysis, and monitoring of urban travel activities. Building an efficient, accurate, and robust method for travel mode identification is one of the key steps in this process. The existing travel mode identification methods have some limitations, such as the unsatisfactory performance for fine-grained travel mode identification, and the lack of sufficient empirical evidence for the existing identification technology. Deep learning algorithms have demonstrated their powerful ability to solve complex classification problems across domains such as natural language processing and text sentiment analysis. This paper makes two contributions. First, it proposes a travel mode identification method utilizing a GRU neural network model. Using 24 features as model input, this method can identify fine-grained travel modes, including walking, cycling, car, and bus. Second, with the support of mobile communication operators, this paper designs and conducts synchronized data collection experiments, obtains individual detailed cellular signaling data, and empirically assesses the identification performance of the method in this paper and other existing models.

The empirical results indicate that the identification model suggested in this paper has a favorable performance for four modes of transportation, with a precision, recall, and F score of 90.5%. This performance is better than other identification models based on machine learning, including random forest, support vector machine, and BP neural network. Moreover, considering both the model recognition accuracy and the model training efficiency, the model based on a GRU neural network also outperforms the other three recognition models based on deep learning algorithms, including recurrent neural network (RNN), long short-term memory network (LSTM), and bidirectional long short-term memory network (Bi-LSTM).

The method presented in this paper also has some aspects for optimization and validation. First, due to factors as experimental cost, the cellular signaling data used for training and validating the travel mode identification model consist of about 62,000 records. To increase the size of the dataset for model training, we use the sliding time window method to generate more datasets. This is reasonable in the theoretical research stage of the model. However, before applying this method to the big-data platform, more real data, instead of synthetic data, are required to conduct adequate model performance validation. Second, as shown in Figures 10 and 11, the models based on deep learning algorithms have higher recognition accuracy. However, deep learning models need more training data and use more computing resources because of a larger number of parameters, which implies that the implementation cost of deep learning-based models is higher. Therefore, in the implementation process of the big-data platform, the choice of travel information identification model ultimately depends on a comprehensive evaluation of two major factors: recognition accuracy and computing efficiency.

The current recognition accuracy of cellular signaling data is not sufficient to achieve the identification of travel mode chains. For example, for the combination of travel mode walking-bus-walking, it is difficult to identify the walking trips after departure or before arrival based on cellular signaling data. In the future, based on the high-precision positioning technology in the 5G environment, combined with more source data information, such as vibration, temperature, sound, and other built-in information of mobile phones, it is expected to further explore and realize the identification method of the abovementioned travel mode chains.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest with respect to the publication of this paper.

Acknowledgments

References

[1] C. Pan, J. Lu, S. Di, and B. Ran, "Cellular-based data-extracting method for trip distribution," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1945, no. 1, pp. 33–39, 2006.

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

[3] F. Yang, Z. Yao, Y. Cheng, B. Ran, and D. Yang, "Multimode trip information detection using personal trajectory data," *Journal of Intelligent Transportation Systems*, vol. 20, no. 5, pp. 449–460, 2016.

[4] N. Caceres, J. P. Wideberg, and F. G. Benitez, "Deriving origin–destination data from a mobile phone network," *IET Intelligent Transport Systems*, vol. 1, no. 1, p. 15, 2007.

[5] F. Wang and C. Chen, "On data processing required to derive mobility patterns from passively-generated mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 58–74, 2018.

[6] T. Sohn, A. Varshavsky, A. LaMarca, M. Y. Chen, T. Choudhury, and I. Smith, "Mobility detection using everyday gsm traces," in *Proceedings of the UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006 Orange County*, pp. 212–224, Springer Berlin Heidelberg, Orange County, CA, USA, September 2006.

[7] H. Wang, F. Calabrese, G. Di Lorenzo, and C. Ratti, "Transportation mode inference from anonymized and aggregated mobile phone call detail records," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems*, pp. 318–323, IEEE, Funchal, Portugal, September 2010.

[8] A. N. Larijani, A. M. Olteanu-Raimond, J. Perret, M. Brédif, and C. Ziemlicki, "Investigating the mobile phone data to estimate the origin destination flow and analysis; case study: Paris region," *Transportation Research Procedia*, vol. 6, pp. 64–78, 2015.

[9] T. Holleczek, D. The Anh, S. Yin, Y. Jin, and S. Antonatos, "Traffic measurement and route recommendation system for mass rapid transit (mrt)," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1859–1868, Sydney, Australia, August 2015.

[10] C. Horn and R. Kern, "Deriving public transportation timetables with large-scale cell phone data," *Procedia Computer Science*, vol. 52, pp. 67–74, 2015.

[11] H. Poonawala, V. Kolar, S. Blandin, L. Wynter, and S. Sahu, "Singapore in motion: insights on public transport service level through farecard and mobile data analytics," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and data mining*, pp. 589–598, San Francisco, CA, USA, August 2016.

[12] Y. Yamada, A. Uchiyama, A. Hiromori, H. Yamaguchi, and T. Higashino, "Travel estimation using Control Signal Records in cellular networks and geographical information," in *Proceedings of the 2016 9th IFIP Wireless and Mobile Networking Conference (WMNC)*, pp. 138–144, IEEE, Colmar, France, July 2016.

[13] Y. Qu, H. Gong, and P. Wang, "Transportation mode split with mobile phone data," in */Proceeding of the IEEE 18th International Conference on Intelligent Transportation Systems (ITSC 2015)*, pp. 285–289, IEEE, Gran Canaria, Spain, September 2015.

[14] S. Danafar, M. Piorkowski, and K. Kryszczuk, "Bayesian framework for mobility pattern discovery using mobile network events," in *Proceeding of the 2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 1070–1074, IEEE, Kos, Greece, August 2017.

[15] K. Chin, H. Huang, C. Horn, I. Kasanicky, and R. Weibel, "Inferring fine-grained transport modes from mobile phone cellular signaling data," *Computers, Environment and Urban Systems*, vol. 77, Article ID 101348, 2019.

[16] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Systems with Applications*, vol. 120, pp. 426–435, 2019.

[17] J. Kim, J. H. Kim, and G. Lee, "GPS data-based mobility mode inference model using long-term recurrent convolutional networks," *Transportation Research Part C: Emerging Technologies*, vol. 135, Article ID 103523, 2022.

[18] C. Wang, H. Luo, F. Zhao, and Y. Qin, "Combining residual and LSTM recurrent networks for transportation mode detection using multimodal sensors integrated in smartphones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5473–5485, 2021.

[19] F. Yang, Y. Wang, P. J. Jin, D. Li, and Z. Yao, "Random forest model for trip end identification using cellular phone and points of interest data," *Transportation Research Record*, vol. 2675, no. 7, pp. 454–466, 2021.

[20] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 36–44, 2011.

[21] M.-H. Wang, S. D. Schrock, N. Vander Broek, and T. Mulinazzi, "Estimating dynamic origin-destination data and travel demand using cell phone network data," *International Journal of Intelligent Transportation Systems Research*, vol. 11, no. 2, pp. 76–86, 2013.

[22] J. Schlaich, T. Otterstätter, and M. Friedrich, "Generating trajectories from mobile phone data," in *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2010.

[23] L. Ni, X. C. Wang, and X. M. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 510–526, 2018.

[24] C. Chen, L. Bian, and J. Ma, "From traces to trajectories: how well can we guess activity locations from mobile phone traces?" *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 326–337, 2014.

[25] H. Jiang, F. Yang, X. Zhu, Z. Yao, and T. Zhou, "Improved F-DBSCAN for trip end identification using mobile phone data in combination with base station density," *Journal of Advanced Transportation*, vol. 2022, Article ID 3099721, 17 pages, 2022.

[26] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, and F. Bougares, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, https://arxiv.org/abs/1406.1078.

[27] Y. Wu, Y. Fang, C. Liu, Z. Fan, and C. Wang, "Gated recurrent unit based frequency-dependent hysteresis modeling and end-to-end compensation," *Mechanical Systems and Signal Processing*, vol. 136, Article ID 106501, 2020.

[28] H. Zhao, C. Hou, H. Alrobassy, and X. Zeng, "Recognition of transportation state by smartphone sensors using deep Bi-LSTM neural network," *Journal of Computer Networks and*

Communications, vol. 2019, Article ID 4967261, 11 pages, 2019.

[29] W. Wu, W. Liao, J. Miao, and G. Du, "Using gated recurrent unit network to forecast short-term load considering impact of electricity price," *Energy Procedia*, vol. 158, pp. 3369–3374, 2019.

[30] H. Huang, Y. Cheng, and R. Weibel, "Transport mode detection based on mobile phone network data: a systematic review," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 297–312, 2019.