

Research Article

Real-Time Return Demand Prediction Based on Multisource Data of One-Way Carsharing Systems

Dongbo Liu^{1,2} Jian Lu¹ and Wanjing Ma^{3,4}

¹School of Transportation, Southeast University, Nanjing, China

²Traffic Management Research Institute of the Ministry of Public Security, Wuxi, China

³Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Road #2, Nanjing, China

⁴Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, China

Correspondence should be addressed to Wanjing Ma; mawanjing@tongji.edu.cn

Received 16 October 2020; Revised 23 March 2021; Accepted 31 March 2021; Published 26 April 2021

Academic Editor: Meaad Saberi

Copyright © 2021 Dongbo Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One-way carsharing system has been widely adopted in the carsharing field due to its flexible services. However, as one of the main limitations of the one-way carsharing system, the imbalance between supply and demand needs to be solved. Predicting pick-up demand has been studied to achieve the goal, but using returned vehicles to reduce unnecessary relocation is also one of the important methods. Nowadays, trajectory data and other data are available for real-time prediction for return demand. Based on the return demand prediction, the relocation response can be more reasonable. Thus, the balance of demand and supply can be largely improved. The multisource data include trajectory data, user application log data, order data, station data, and user characteristic data. Based on these data, a return demand prediction model was used to predict whether the user will return the vehicle in 15 min in real time, and a destination station prediction model was applied to forecast which station the user will park at. Finally, a case study using ten stations' one-week field data was conducted to test the benefit of the dynamic return demand prediction. The results showed that the return demand prediction improves the efficiency of the relocations by mitigating the condition that the station parking space is full or empty. The potential application of this study would effectively reduce unnecessary relocation and further formulate an active operation optimization strategy to reduce the system's operational cost and improve the service quality of the system.

1. Introduction

With the vigorous development of mobile Internet technology and the emergence of the new business model represented by the sharing economy, electric carsharing systems play an increasingly important role in the transportation field. They can improve travel convenience, increase vehicle utilization, and reduce the demand for parking spaces. At the same time, they also have a positive effect on energy saving and environmental protection [1–4].

At present, more and more carsharing systems adopt a one-way mode system rather than a round-way one because of the flexibility and better experience provided by one-way carsharing services [5, 6]. However, there are also limitations to the

one-way carsharing system, such as the imbalance between supply and demand. It is important to use an advanced, dynamic, and real-time management system to improve the relocation efficiency in order to bring higher profit to managers [7–9]. Predicting pick-up demand has been studied to achieve the goal [10–12], but using return demand to reduce unnecessary relocation is also one of the important methods [13]. The methodologies actually are similar. However, the study targets are different. The main goal of user pick-up demand prediction for the carsharing system is to obtain potential user demand and increase the number of orders, and the main goal of user return demand prediction is to make full use of other users' pick-up or return behavior within the system and reduce unnecessary relocation and costs. Therefore, it becomes more

and more significant to demonstrate the importance of necessary relocation in one-way carsharing systems, accurately forecast the return demand for shared cars, and then further guide relocation, thus reducing the system's operational cost and attracting a higher number of users.

Currently, the consideration of demand by research related to vehicle relocation can be mainly classified into two categories. The first one is the vehicle relocation under determinate demand [14, 15]. These studies are mainly based on the one-way carsharing systems with reservation and use mathematic planning models. The second type considers demand uncertainty in the vehicle relocation strategy [16–19]. Four methods are used to solve this kind of problem, i.e., the responsive approach (relocation depending on the threshold that is the number of available vehicles at a station), demand prediction approach (relocation depending on the predicted demand), rolling time-domain approach (relocation depending on the short-term demand), and stochastic programming approach (relocation depending on the stochastic demand). More and more studies start from the perspective of demand prediction because of the popularity of the data-driven environment and advanced deep learning models [20, 21]. Based on the demand prediction model, researchers determine the optimal scheme to relocate cars and rebalance staff. However, existing research about demand prediction mainly focuses on the long-time prediction of a carsharing station at an aggregation level and is not applicable in predicting users' demand at a disaggregation level [22, 23]. Furthermore, research related to vehicle relocation does not use real-time data but historical order data [24–27]. Technically, the historical order data can provide useful information for return demand prediction. However, the characteristics of return demand show obvious randomness and uncertainty. Therefore, only using the order data might not be sufficient for real-time relocation strategy; this paper introduced trajectory data and user app log data as the supplement source of the part which historical order data cannot reflect.

To solve the above problems, this paper aims to establish a real-time dynamic return demand prediction model for a one-way electric carsharing system based on the data of the EVCARD system which is a large carsharing provider and has a fleet of more than 5,000 electric vehicles in Shanghai, China. It does not allow users to make a reservation in advance; thus, users can only use the app called EAPP for accessing vehicles on demand. Additionally, multisource data were used, including vehicle trajectory data, real-time EVCARD application (EAPP) log data, historical order data, real-time station data, and user characteristic data. The collection cycle of the vehicle trajectory data, i.e., the real-time vehicle GPS data, is not a fixed step. For the vehicle trajectory data, the shortest time interval between adjacent trajectory points can reach 5 s under the condition of frequent operation of the vehicle, e.g., twisting the steering wheel frequently, while the general time interval is about 30 s under normal driving and reliable data transmission. The EAPP log data refers to the logs generated by the EVCARD application when users open or click the EAPP, and it records when and where the app is opened or operated. It has been used to reflect users' pick-up demand and can also be applied to return demand [12]. Generally, there is an observation that a

user who has a return demand would open EAPP when he/she is approaching the station. Thus, the EAPP log data can be used as a data source to reflect users' return demand. Meanwhile, even the user has a return demand, the station where the vehicle will be parked is also unknown, and therefore there is a need to discriminate the return demand and further to get the destination station. In this study, based on the collected data, a user return demand prediction model is developed to predict the real-time return demand, and a destination station prediction model is applied to forecast the place where a shared vehicle will be parked. Finally, a case study is conducted to test the benefits of the models proposed by this study.

2. Experiment Design and Methodology

2.1. Experiment Design. Predicting the return demand and the destination station of users is one of the significant methods to reduce unnecessary relocation and improve service quality. Correctly predicting whether the user will return the car in a short time and where to return the car can effectively avoid part of the vehicle relocation tasks carried out by operators. This article does not consider the impact of the cost of returning vehicles to different stations but directly characterizes the user's preference for different stations through the user's historical characteristics. Considering the users' order duration is mainly concentrated on 0–15 min, the prediction models will determine whether a user will return the car in the next 15 minutes and further predict the destination station within a radius of 15 minutes distance with the current track point as the center. After that, the control system will determine if relocating vehicles for stations is needed according to factors, such as station condition and time interval. The schematic diagram is shown in Figure 1.

In the part of building prediction models, this paper used vehicle trajectory data, log information, and other data to establish the return demand prediction model and destination station prediction model. Specifically speaking, for any trajectory point point_i(p_i) in a user's travel trajectory, the two models' judgment results are determined by the current trajectory point and its around station set characteristics. If the current trajectory point is the first point of the user's travel trajectory, that is, $i = 1$, the station set is the stations in the area that covers the circle with point_i as the center and 15 min distance ($R = L_{15\text{ min}}$) as the radius; if the current trajectory point is not the first track point of the trip, the station set is the stations in the area that covers the circle with point_i as the center and 15 min distance as the radius minus the circle with point_{i-1} as the center and 15 min distance ($R = L_{15\text{ min}}$) as the radius. It is worth noting that the 15-minute distance refers to a 15-minute air trip at the user's historical average speed. For a first-time user, it refers to all users' average road trip of 15 minutes before returning the vehicle, and for a repeated user, the average road trip of 15 minutes before he/she returns the vehicle is regarded as his/her mileage of 15 minutes.

In the relocation method part, a case study is conducted. During the process of driving, the vehicle's real-time trajectory point will be uploaded to the control system, and for any users' driving trajectory point, this prediction will be triggered to figure out when and where the user will return the car. The

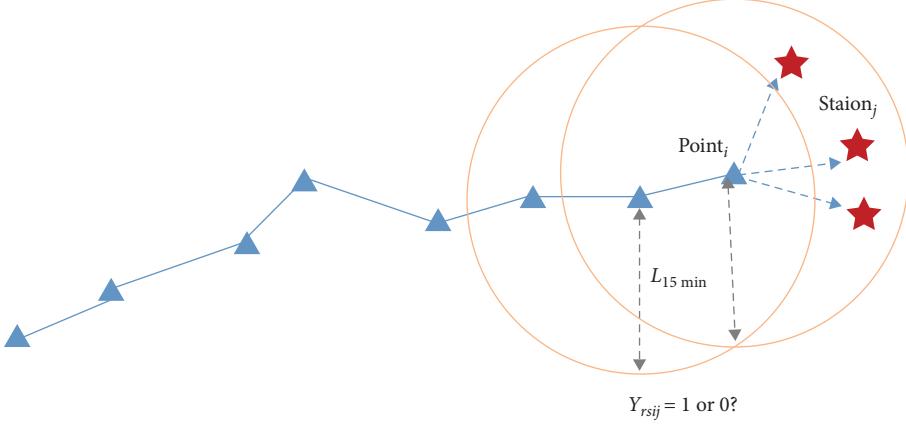


FIGURE 1: Diagram of return demand prediction and destination station prediction.

thresholds used to determine whether there is a return demand for the user within 15 min and where the vehicle will be parked are needed to be determined. If the thresholds are too high, the number of unnecessary relocations will increase, thereby reducing the profit. Accordingly, if the thresholds are too low, the system may misjudge a vehicle that will be returned to the station by its user, resulting in insufficient vehicle supply. Then a process that determines whether the relocation is needed was conducted, which will be introduced in detail in 5.2. The framework of the study is shown in Figure 2.

2.2. Prediction Models

2.2.1. Return Demand Prediction. For any trajectory point $point_i$ in the users' trajectory, if the timestamp of the point i is within 15 minutes before returning the car, $Y_{ri} = 1$; if not, $Y_{ri} = 0$. Since the dependent variable is binary, a Logistic model is used. The model is shown as follows:

$$\text{Log}\left(\frac{P(Y_{ri} = 1)}{1 - P(Y_{ri} = 1)}\right) = b_0 + b_1 x_{1i} + \dots + b_k x_{ki} + \varepsilon_i, \quad (1)$$

where $P(Y_{ri} = 1)$ represents the probability that the trajectory point p_i has return demand within 15 min. If several trajectory points have return demand, the earliest one is used. For the other parameters, b_0 is the intercept, and b_k is the coefficient of the k^{th} independent variable (x_{ki}), for example, the user's historical average mileage. ε_i is the error and it follows a standard logistic distribution.

2.2.2. Destination Station Prediction. To predict the destination station where the vehicle will be parked, a Logistic model is used. For any users' current trajectory point $point_i$ in trajectory T_u , if in the station set S_r there is a station s_j that can be matched with the predicted destination station D_u , $Y_{rsij} = 1$; if not, $Y_{rsij} = 0$. The model is as follows:

$$\text{Log}\left(\frac{P(Y_{rsij} = 1)}{1 - P(Y_{rsij} = 1)}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad (2)$$

where $P(Y_{rsij} = 1)$ is the probability that the trajectory point $point_i$ indicates the return of the car at the station s_j . If several destination stations are matching with the current trajectory, the one with the highest probability will be chosen. Besides, β_0 is the intercept, and β_k is the regression coefficient of the k^{th} independent variable (x_{ki}). ε_i follows a standard logistic distribution.

3. Data Collection

3.1. Data Sources. The data sources of this study are from an electric vehicle sharing system, i.e., the EVCARD system which was established in 2013, and up to March 2017, there were 1,739 stations, about 90,000 users, and more than 5,000 electric vehicles in Shanghai, China. The data contains static data and dynamic data. The static data includes historical order data, static station data, and historical trajectory data, while dynamic data includes real-time station data, real-time trajectory data, and real-time EAPP log information. Among them, the order data is from December 1st, 2016, to March 15th, 2017, which contains about 750,000 orders. The vehicle trajectory data is from February 15th, 2017, to March 15th, 2017, about 101,000 trips. The EAPP log information is from March 1st, 2017, to March 15th, 2017, which contains about 1.5 million logs which contain time, user ID, and locations. The common time of these data is from March 1st, 2017, to March 15th, 2017, so the data during this period is used to build real-time characteristics, but historical characteristics, such as whether a user used the EAPP before and the number of orders a user makes on a particular station, are extracted from earlier data. The detailed fields in each data source and the connections between them are shown in Figure 3.

3.2. Collected Variables. In this paper, 38 variables were collected and they were divided into six categories, as shown in Table 1. The station set characteristics are used to explore the relationship between the station set and historical route OD (origin and destination) pair. The driving mileage and driving duration characteristics are used to describe a user's driving habits which contain two aspects, i.e., the user's all

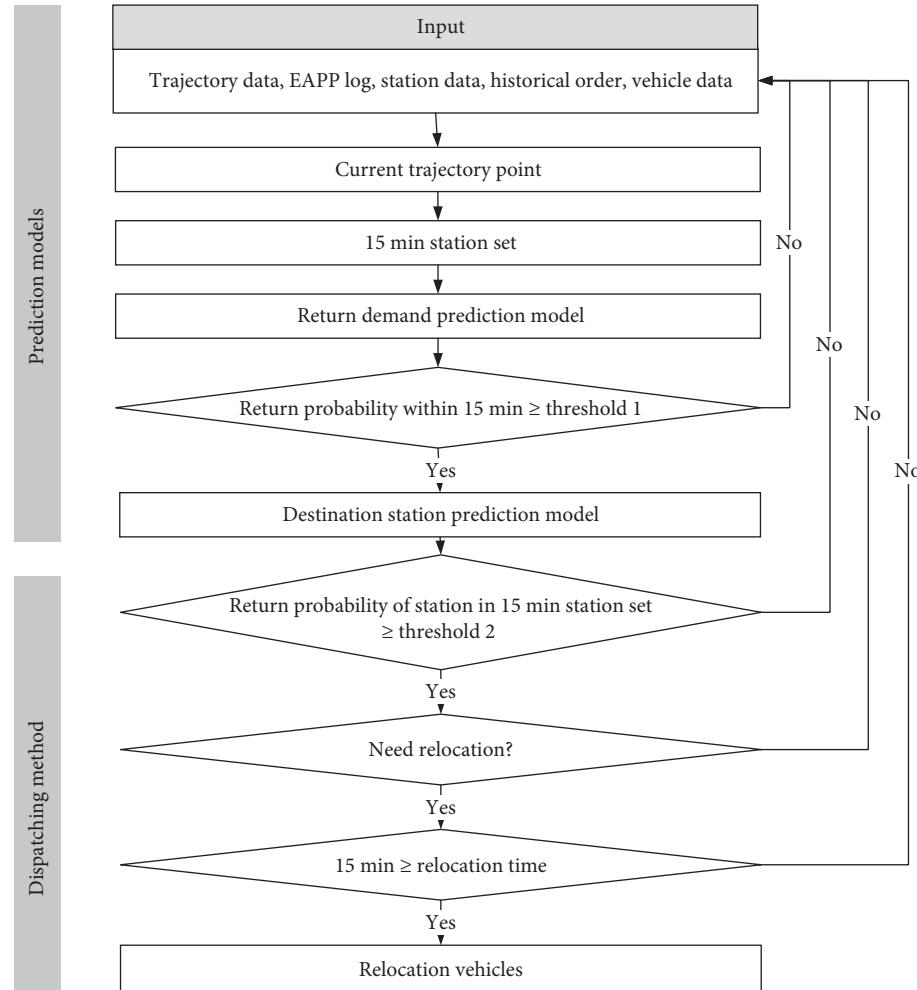


FIGURE 2: Framework of the study.

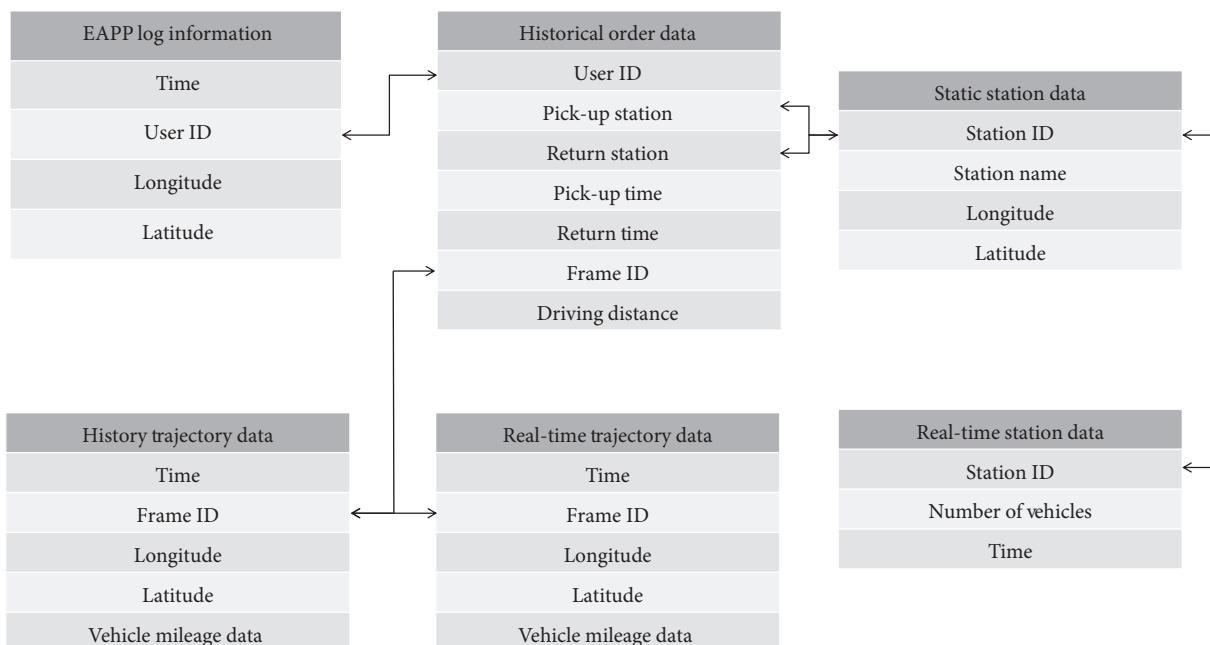


FIGURE 3: Data structure.

TABLE 1: Characteristic variables and symbols.

Category	Variables	Symbols	Examples
Station set	Whether the station (s_j) in the current trajectory point's station set (S_r) belongs to the stations where the user has picked up vehicles	whetherO	1
	Whether the station (s_j) in the current trajectory point's station set (S_r) belongs to the stations where the user has returned vehicles	whetherD	1
Driving mileage	Current track length + 15 min mileage > user's historical minimum mileage	whetherLumin	1
	Current track length + 15 min mileage < user's historical maximum mileage	whetherLumax	1
	User historical average mileage	Luave	27.26 km
	User historical median mileage	Lumedian	25.53 km
	Current track length+15 min mileage > minimum mileage of the user's historical order with the same starting point	whetherLuomin	1
	Current track length+15 min mileage < maximum mileage of the user's historical order with the same starting point	whetherLuomax	1
	The historical average mileage of user's order with the same starting point	Luoave	27.56 km
	Historical median mileage of user's order with the same starting point	Luomedian	26.77 km
	Current track length + 15 min mileage	LaddR	20.15 km
	Current travel duration+15 min > user's historical minimum travel duration	whetherTumin	1
Driving duration	Current travel duration+15min < user's historical maximum travel duration	whetherTumax	1
	Current travel duration+15 min > the 15th percentile of the user's historical travel duration	whetherTuq15	1
	Current travel duration+15min < the 85th percentile of the user's historical travel duration	whetherTuq85	1
	User historical average driving duration	Tuave	15.10 min
	User historical median driving duration	Tumedian	14.70 min
	Current travel duration+15 min > minimum travel duration of the user's historical order with the same starting point	whetherTuomin	1
	Current travel duration+15 min < maximum travel duration of the user's historical order with the same starting point	whetherTuomax	1
	Current travel duration+15 min > the 15th percentile of travel duration of the user's historical order with the same starting point	whetherTuoq15	1
	Current travel duration+15 min < the 85th percentile of travel duration of the user's historical order with the same starting point	whetherTuoq85	1
	User average driving duration of the user's historical order with the same starting point	Tuoave	14.32 min
EAPP	User median driving duration of the user's historical order with the same starting point	Tuomedian	15.22 min
	Current travel duration+15 min	Tadd15	16 min
Destination station	A user opened the APP within 15 minutes before returning the vehicle in history	APP	0
	The number of all users' historical orders whose destination station D can be matched with station (s_j) in the station set (S_r)	Dordernum	1
	The proportion of all users' historical orders whose destination station D can be matched with station (s_j) in the station set (S_r)	Dorderfre	0.2
	The minimum time difference between the order duration of all users' historical orders whose destination station D can be matched with station (s_j) in the station set (S_r) and the current trajectory duration (t_{pi})	DdetminT	3.3 min
	The number of the user's historical orders whose destination station D can be matched with station (s_j) in the station set (S_r)	UDordernum	4
	The proportion of the user' historical orders whose destination station D can be matched with station (s_j) in the station set (S_r)	UDorderfre	0.25
	The minimum time difference between the order duration of the user's historical orders whose destination station D can be matched with station (s_j) in the station set (S_r) and the current trajectory duration (t_{pi})	UDdetminT	4.1 min

TABLE 1: Continued.

Category	Variables	Symbols	Examples
Pick-up station and destination station	The number of all users' historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) The proportion of all users' historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) The minimum time difference between the order duration of all users' historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) and the current trajectory duration (t_{pi}) The number of the user's historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) The proportion of the user's historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) The minimum time difference between the order duration of the user's historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) and the current trajectory duration (t_{pi}) The trajectory difference between the order duration of the user's historical orders whose starting station is the same as the current trajectory's starting point and destination station D can be matched with a station (s_j) in the station set (S_r) and the current trajectory duration (t_{pi})	ODordernum ODorderfre ODdetminT UODordernum UODorderfre UODdetminT DiffOD	2 0.1 6.3 min 2 0.05 5.5 min 0.58

orders characteristics and the characteristics of the user's order whose starting station is the same with the current trajectory. The EAPP characteristics are used to represent the habit that a user checks the APP before returning the vehicle. The pick-up station and destination station characteristics are used to explore the similarity between the current trajectory and historical trajectory to find the destination station in the station set.

It should be noted that, in the process of model building, these variables are extracted from each set of matched data; in the process of simulation, these variables are all updated every five minutes to achieve a real-time judgment. All variables are calculated based on the actual condition.

Additionally, the trajectory difference has to be explained in detail [28]. First, the trajectory T is composed of trajectory points p . The current trajectory is $T_u = \{p_i\}$, $i = 1, \dots, n_u$, and the historical trajectory is $T_d = \{p_j\}$, $j = 1, \dots, n_d$. Secondly, for any trajectory point p_i belonging to a trajectory T_u , the minimum Euclidean distance from it to another trajectory T_d can be calculated as the minimum of all distances between points of T_d and the trajectory point p_i .

$$d_p(p_i, T_d) = \min_j \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad \forall i = 1, \dots, n_u. \quad (3)$$

Similarly, the minimum distance between points of the current trajectory T_u and the historical trajectory point p_j are as follows:

$$d_p(p_j, T_u) = \min_i \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad \forall j = 1, \dots, n_d. \quad (4)$$

At last, the trajectory difference which refers to the difference between the current trajectory and the historical trajectory can be calculated as follows:

$$d(T_u, T_d) = \frac{G}{L_u} \cdot \frac{1}{n_u + n_d} \cdot \left(\sum_{i=1}^{n_u} d_p(p_i, T_d) + \sum_{j=1}^{n_d} d_p(p_j, T_u) \right), \quad \forall i = 1, \dots, n_u, \forall j = 1, \dots, n_d, \quad (5)$$

where L_u is the current travel distance of the user's track, which is used to unify the result to reduce the impact of trajectory difference calculation brought by the trajectory length; since there are fewer alternative paths for short-distance trajectories than long-distance ones, the matching degree of short-distance trajectory is high; and G is the approximate linear conversion factor from the geographic coordinate to the rectangular coordinate based on linear geometry in the EVCARD operation area, which is 111 km [28].

Additionally, considering the trajectory to be predicted is not the complete travel route, it cannot directly match the completed historical trajectory, so the variable n_c is introduced, as shown in the following formula:

$$n_c = \begin{cases} n_d, & n_d < n_u, \\ n_u, & n_d \geq n_u. \end{cases} \quad (6)$$

For the destination $D = D_1, \dots, D_l$ from the same starting point, where l is the number of historical destinations. For all the trajectories that may appear in a pair of OD,

the most similar trajectory is selected to calculate the trajectory difference of destination.

$$\text{DiffOD} = \min_k d(T_u, T_d^k), \quad (7)$$

where T_d^k represents the k^{th} historical trajectory of destination D , $k = 1, \dots, N_D$. N_D represents the number of trajectories of destination D , and $D = D_1, \dots, D_l$.

In order to illustrate this research clearly, the authors explain the calculation and application of the trajectory difference. The current blue cross trajectory starts from the starting point s_1 . From Figure 4, it can be seen that in the early driving period the historical trajectory with the orange cross has the smallest trajectory difference 0.12 with the current trajectory; in the middle period the yellow circle historical trajectory has a difference of 0.19 with the current trajectory; in the later period the purple diamond-shaped historical trajectory has a degree of difference of 0.19. The return points for the three historical tracks are all s_7 . The return point for the dark blue cross historical trajectory is s_{1288} , and its difference from the current trajectory is 0.85; the return point for the dark red minus sign historical trajectory is s_{2170} , and the difference between it and the current trajectory is 0.27. The difference between the trajectory where the return point is s_7 and the current trajectory is the smallest, indicating that the return point is most likely to be s_7 .

4. Model Results Analysis and Verification

4.1. Return Demand Prediction Model. The sample which contains about 50,000 trajectory data, EAPP log information, and other data from March 1st, 2017, to March 15th, 2017, has been randomly divided into training and validation datasets with a ratio of 7:3. The study uses logistic regression to build the prediction model, which can directly and quantitatively obtain the relationship between dependent variables and independent variables.

Beginning with all variables considered, the insignificant variables are eliminated from the following model building step. Later, to select the most significant and not highly correlated variables, firstly the correlation tests between dependent and independent variables are conducted to find the insignificant independent variables, and these variables are all significantly correlated with the dependent variable. Secondly, to prevent a high correlation between independent variables for the two models in this study, a correlation test is performed. Considering the dataset size of each characteristic variable is about 50,000 which is a large enough number for the parametric correlation tests [29, 30], Pearson test between continuous variables, Chi-test between categorical variables, and T -test between a continuous variable and a categorical variable are used to analyze the correlation of two variables. For example, for the correlation calculation between *Luoave* and *whetherLumin*, *Luoave* is a continuous variable and *whetherLumin* is a categorical variable; thus a T -test is used to perform correlation test and its correlation coefficient value is less than 0.4, and the two variables are both included in the next step. Also, for the correlation

between whether*Lumin* and whether*Tuoq15*, the Chi-test is used and it is found that whether*Lumin* is positively related to whether*Tuoq15* and its absolute of the correlation coefficient value is 0.908. In this study, the variable which provides a lower AIC than another variable is chosen.

Finally, the return demand prediction model is shown in Table 2. For the Logistic model, the Area under ROC Curve (AUC) is usually used as an evaluation index to evaluate the accuracy and reliability of the model. The AUC value is between 0.5 and 1.0, and the higher, the better. The AUC for training and validation are 0.712 and 0.715, respectively. It indicates the model performs similarly for training and validation datasets and demonstrates that the model's predictive accuracy for predicting return demand is good.

The return demand prediction model shows that as the user's current driving length (*LaddR*) and travel duration (*Tadd15*) increase, the probability of the user returning the vehicle in 15 minutes increases. Meanwhile, when the user's historical driving mileage (*Luave*, *Luoave*) and the historical driving time (*Tumedian*, *Tuomedian*) increase, the probability of the user returning the vehicle within the next 15 minutes decreases because the longer historical driving shows that this user prefers long-distance driving. For discriminant variables, if the current mileage or duration is greater than the minimum value or the 15% quantile of the user's historical data (*whetherLuomin*, *whetherTumin*, *whetherTuoq15*, *whetherTuoq15*), the probability of the user returning the car in the next 15 minutes will increase slightly; besides, if the current travel mileage or duration is greater than the user's historical 85% quantile or the maximum value (*whetherLuomax*, *whetherTumax*, *whetherTuomax*, *whetherTuq85*), the probability of the user's demand to return the car will decrease. Because as the user's historical order travel mileage or duration increases, users are more likely to use the carsharing system for long-distance travel in the future, resulting in the probability that the demand for returning cars does not increase but decreases. It can be seen from the figure below that as the average historical mileage of users increases, users are more inclined to travel long distances. Therefore, in the aggregate results of each section, the proportion of users returning cars in the next 15 minutes gradually decreases (Figure 5).

4.2. Destination Station Prediction Model. Similar to the process of building a return demand prediction model, the destination station prediction model first tests the correlation between each independent variable and the dependent variable. Then, a correlation test is conducted between the significant independent variables, i.e., the Pearson test between continuous variables, Chi-test between categorical variables, and T -test between a continuous variable and a categorical variable. If the correlation level between two variables is higher than 0.4, the variable which could provide a lower AIC was kept. The final model is shown in Table 3. The AUC for training and validation are 0.959 and 0.954, respectively. It indicates the model performs similarly for training and validation datasets and demonstrates that the

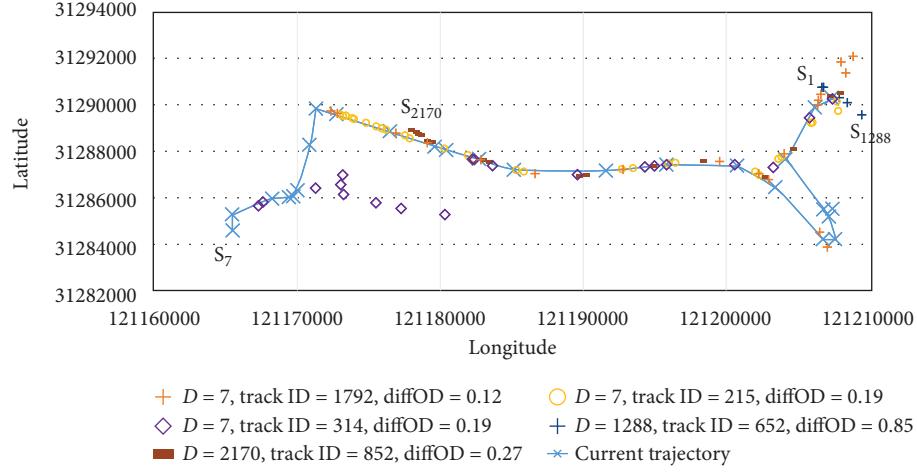


FIGURE 4: The schematic diagram of the trajectory difference calculation.

TABLE 2: Return demand prediction model.

Variable	Coefficient	Std.	Z value	Pr ($> z $)	
Intercept	-0.8713	0.1060	-8.2230	$<2e-16$	***
Luave	-0.0115	0.0031	-3.6980	0.0002	***
whetherLuomin	0.5241	0.065	8.0670	0	***
whetherLuomax	-0.4051	0.0356	-11.3760	$<2e-16$	***
Luoave	-0.0086	0.0032	-2.6670	0.0076	***
LaddR	0.0073	0.0016	4.6790	0	***
whetherTumin	0.5221	0.1043	5.0080	0	***
whetherTumax	-0.3653	0.0685	-5.3340	0	***
whetherTuq15	0.5689	0.0948	6.0010	0	***
whetherTuq85	-0.1178	0.0442	-2.6650	0.0077	**
Tumedian	-0.0031	0.0006	-5.6120	0	***
whetherTuomax	-0.2841	0.057	-4.9840	0	***
whetherTuoq15	0.3848	0.0623	6.1800	0	***
Tuomedian	-0.0007	0.0001	-4.7190	0	***
Tadd15	0.0021	0.0001	20.5680	$<2e-16$	***
APP	2.1035	0.138	15.2400	$<2e-16$	***
Training AUC:	0.712	validation AUC:	0.715		
AIC:	34701				

Symbol ‘*’ of different numbers means the significant variable in different values, symbol ‘***’ means close to zero, symbol ‘**’ means 0.001–0.01, and symbol ‘*’ means 0.01–0.05.

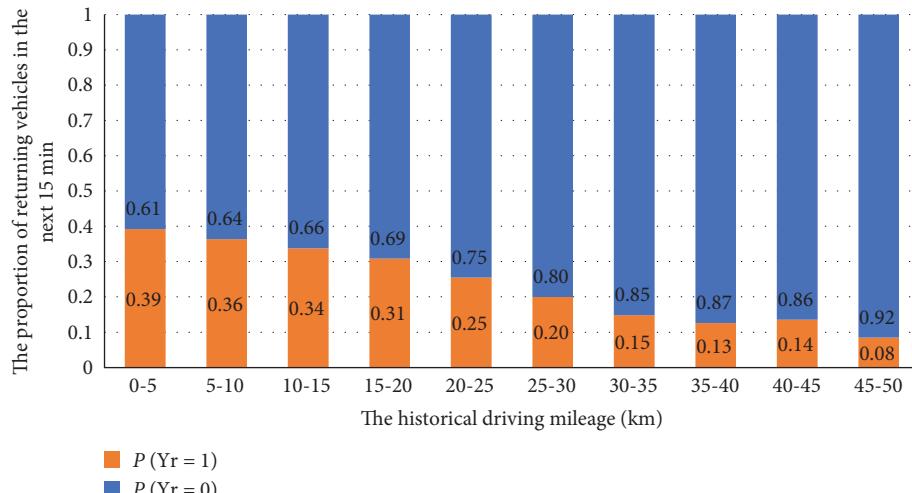


FIGURE 5: The percentage of users returning vehicles in 15 min under different historical mileages.

model's predictive accuracy for predicting return station is good.

The destination station prediction model shows that the proportion of all users' historical orders whose destination station can be matched with a station in the station set (Dorderfre) has a major influence on the destination station prediction, which means that if the station s_j is used more frequently as a return station in all historical orders, so for the current user's trajectory point, the station s_j within 15 minutes station set is more likely to be the destination station for this trip. Similarly, UDordernum, UDorderfre, and UODordernum also illustrate that the users were more likely to use the most frequent return station. At the same time, the model also contains variables negatively related to Y_{rsij} . For the variables of detminT type which contains DdetminT, UDdetminT, ODDetminT, and UODdetminT, the smaller the time difference between the user's historical time returned to s_j and the current driving time, the greater the possibility of returning to s_j during this trip. Additionally, for the trajectory difference (DiffOD), if the trajectory difference is smaller, the possibility that the current trajectory will return to s_j within 15 minutes is greater.

5. Case Study

5.1. Experiment Design. To verify the effectiveness of the return demand prediction model and destination station prediction model, this study builds a simulation based on the trajectory data from March 8th, 2017, to March 14th, 2017, for a carsharing system consisting of ten stations that are all located in Jiading, Shanghai. These stations are located in areas where the type of use (residential/business/education) and the demography of the population are quite different (see Figure 6). The information, i.e., station location, the number of parking spaces, and the initial number of vehicles at each station (at 0:00 on March 8th, 2017), from ten field

stations serves as input to the simulation, and the users' trips are between these stations. Additionally, the station distance is calculated by road distance between stations. Besides, two staff are assigned at the No.262 and No.27 station which is with the most traffic volume and the second most traffic volume, respectively, among the ten stations, and the speed of relocation trips is 30 km/h. In the return demand prediction model, the threshold that is used to estimate whether the user will return a vehicle in 15 min is set as 0.25, and the threshold of the destination station prediction model to discriminate which station the vehicle will be returned to is set as 0.30 (Figure 6).

This study also adopts four indexes, i.e., full ratio, empty ratio, waste ratio, and profit to evaluate the performance of the relocation strategies under various demand assumptions. The full ratio is the proportion that a station is full. The empty ratio is the proportion that a station is empty. The waste ratio is the proportion that the relocation is redundant, and the redundant relocation is defined as a phenomenon in which a user wants to return a vehicle to a certain station or wants to borrow a vehicle from a certain station, which could have been solved by borrowing and returning vehicles from other users in the system, but a relocation staff is still arranged for relocating. The profit refers to the seven days' income from the operation of the above carsharing system consisting of ten stations, i.e., the average benefit multiplied with the number of trips. In this simulation, a complete trip is defined by both the pick-up demand and the return demand being met. Among them, each order's benefit, each staff's employment expense per day, and the electric charge per mileage need to be predetermined. According to the field data, it can be found that the average benefit of each order is 33 Yuan. Besides, the staff's employment expense per day is 267 Yuan and the electric charge per mileage is 1 Yuan [31]:

$$\begin{aligned}
 \text{waste ratio} &= \frac{\text{the number of redundant relocation tasks}}{\text{the total number of relocation tasks}}, \\
 \text{full ratio} &= \frac{\text{the number of intervals when a station is full}}{\text{the total number of intervals during the operation period}}, \\
 \text{empty ratio} &= \frac{\text{the number of intervals when a station is empty}}{\text{the total number of intervals during the operation period}}, \\
 \text{profit} &= \text{average benefit} \times \text{the number of complete trips} - \text{each staff's daily salary} \\
 &\quad \times \text{days of operation period} - \text{electric charge per mileage} \times \text{total relocation mileage}.
 \end{aligned} \tag{8}$$

5.2. Relocation Method. Under different assumptions, three relocation methods, i.e., relocation with predictions, relocation with no prediction, and no relocation, are tested to explore how the relocation strategy and demand prediction affect the system performance. Each of them is explained in detail as follows [32–34].

- (1) *Relocation with Predictions.* Firstly, March 8th, 2017, to March 14th, 2017, vehicle trajectory data is updated at each five-minute interval, and the return demand prediction probability greater than 0.25 predicted by the return demand prediction model is filtered out. If the user has return demand in 15 min, the return

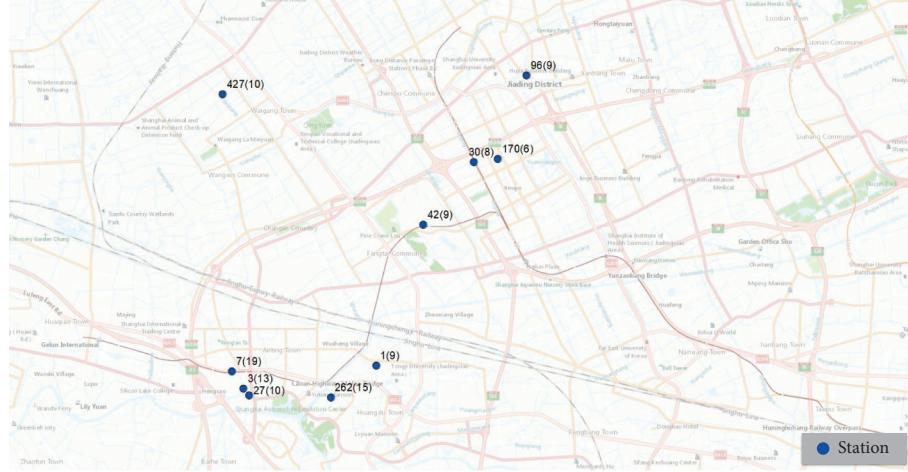


FIGURE 6: The distribution of selected stations.

TABLE 3: Demand time prediction model.

Variable	Coefficient	Std.	Z	Pr (> z)
Intercept	0.5814	0.2531	2.2970	0.0216 *
Dorderfre	370.0000	13.0200	28.4250	<2e - 16 ***
DdetminT	-0.1071	0.0360	-2.9780	0.0029 **
UDordernum	0.0663	0.0269	2.4640	0.0137 *
UDorderfre	6.9810	0.7335	9.5170	<2e - 16 ***
UDdetminT	-0.0776	0.0082	-9.4210	<2e - 16 ***
ODdetminT	-0.0577	0.0054	-10.7570	<2e - 16 ***
UODordernum	0.1646	0.0853	1.9310	0.0435 *
UODdetminT	-0.0187	0.0023	-2.5200	0.0284 *
DiffOD	-0.0133	0.0008	-16.1790	<2e - 16 ***

Training AUC: 0.959
Validation AUC: 0.954
AIC: 8470

Symbol “*” of different numbers means the significant variable in different values, symbol “***” means close to zero, symbol “**” means 0.001–0.01, and symbol “*” means 0.01–0.05.

station of the user will be predicted by the destination station prediction model. Secondly, considering the return prediction results and the number of existing vehicles at each station, whether the station needs relocated-in vehicles is determined by comparing the number of predicted available vehicles at each station at the next interval with the predetermined minimum number of available vehicles which is one in this research. If there are enough available vehicles, vehicles are not needed to be relocated to the station. Thirdly, similar to the decision of relocating in vehicles is filtering the oversupplied stations which need operators to relocate vehicles out by comparing the number of predicted available vehicles at each station at the next interval with the predetermined maximum number of available vehicles which is the number of each station’s parking spots minus one in this paper. Fourth step is

selecting the closest station to each staff from the stations where vehicles need to be relocated out as the alternative relocate-out station and selecting the closest station to each alternative relocate-out station, respectively, from the stations where vehicles need to be relocated in as the alternative relocate-in station. Then, selecting the station pairs with the smallest distance as the final relocate-out station and the final relocate-in station respectively. Fifthly, through comparing relocation time between the final relocate-out station and final relocate-in station with 15 minutes, whether the relocation task needs to be carried out will be determined, i.e., if relocation time is less than 15 minutes, relocating vehicle from the final relocate-out station to the final relocate-in station. Thus, in every step of the simulation, one relocation can be conducted at most. The relocation process is shown in Figure 7.

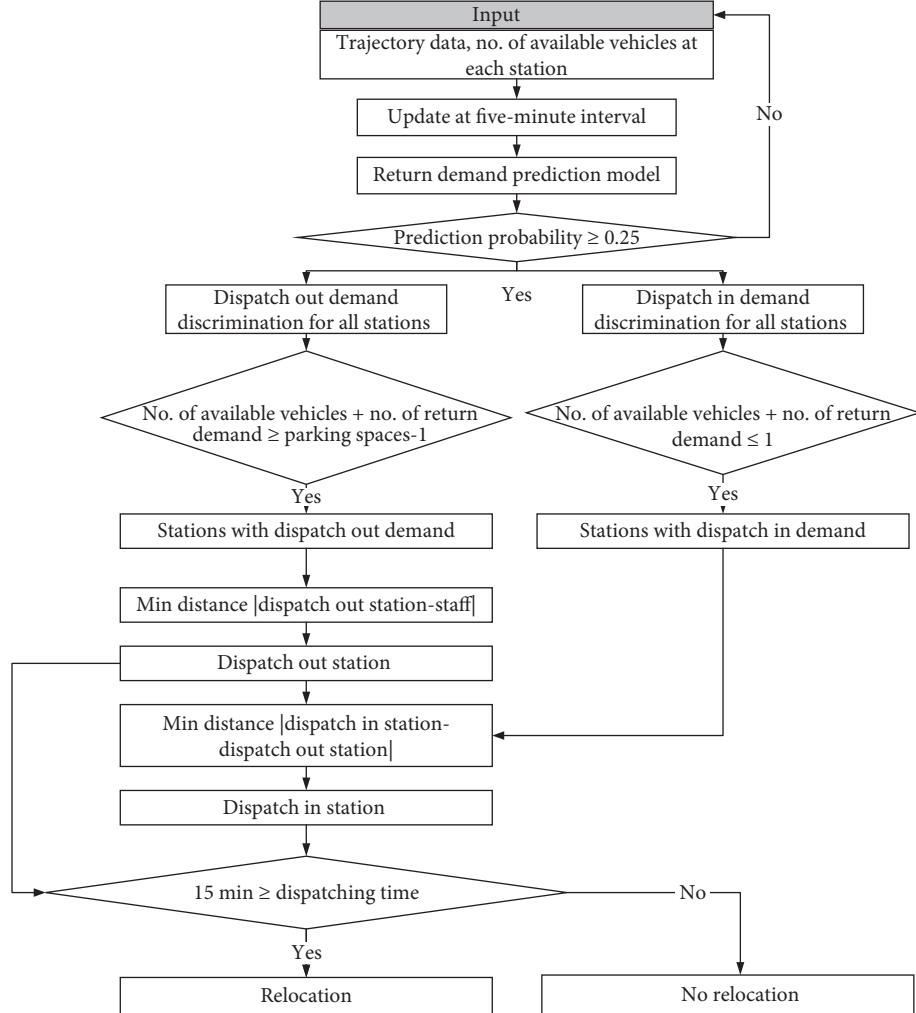


FIGURE 7: Relocation with prediction.

- (2) *Relocation with No Prediction*. According directly to the number of available vehicles at each station without considering the return prediction results, the relocation process is the same as that of the first relocation method. When there are oversupplied (the number of available vehicles at each station is larger than the number of each station's parking spots minus one) or undersupplied (the number of predicted available vehicles at each station is less than one) vehicles, relocating vehicle from the final relocate-out station to the final relocate-in station is conducted. The relocation process is shown in Figure 8.
- (3) *No Relocation*. No relocation means that there are no professional operators employed to relocate vehicles between stations. The March 8th, 2017, to March 14th, 2017, order data is updated at each five-minute interval to represent user demand.

5.3. Results of Case Study. In the no relocation scenario, the system that operated without any relocation tasks can realize about 16,895 Yuan profits and over 3% full ratio and empty

ratio. If relocation tasks are carried out with no demand prediction, the profits can be enhanced, but it leads to a certain waste because the redundant relocation not only causes the labor costs and electric charge waste but also occupies available vehicles. When using the relocation strategy with demand prediction, there are 82 return demands which can be used to be relocated between the 10 stations per week. And the results show that the waste ratio is decreased to zero and profit is increased by 7837 Yuan compared to the no relocation method, which has a great significance. What is more, the improvement of the full ratio and empty ratio indicates that the service level can be enhanced from the original base. Therefore, if these above two models are applied in an actual project, the system can formulate an active operation optimization strategy, which can reduce unnecessary relocation, improve the service quality of the system, and bring about more profit (Table 4).

6. Conclusion

It is important to discriminate the return demand and further to obtain the destination station quickly and

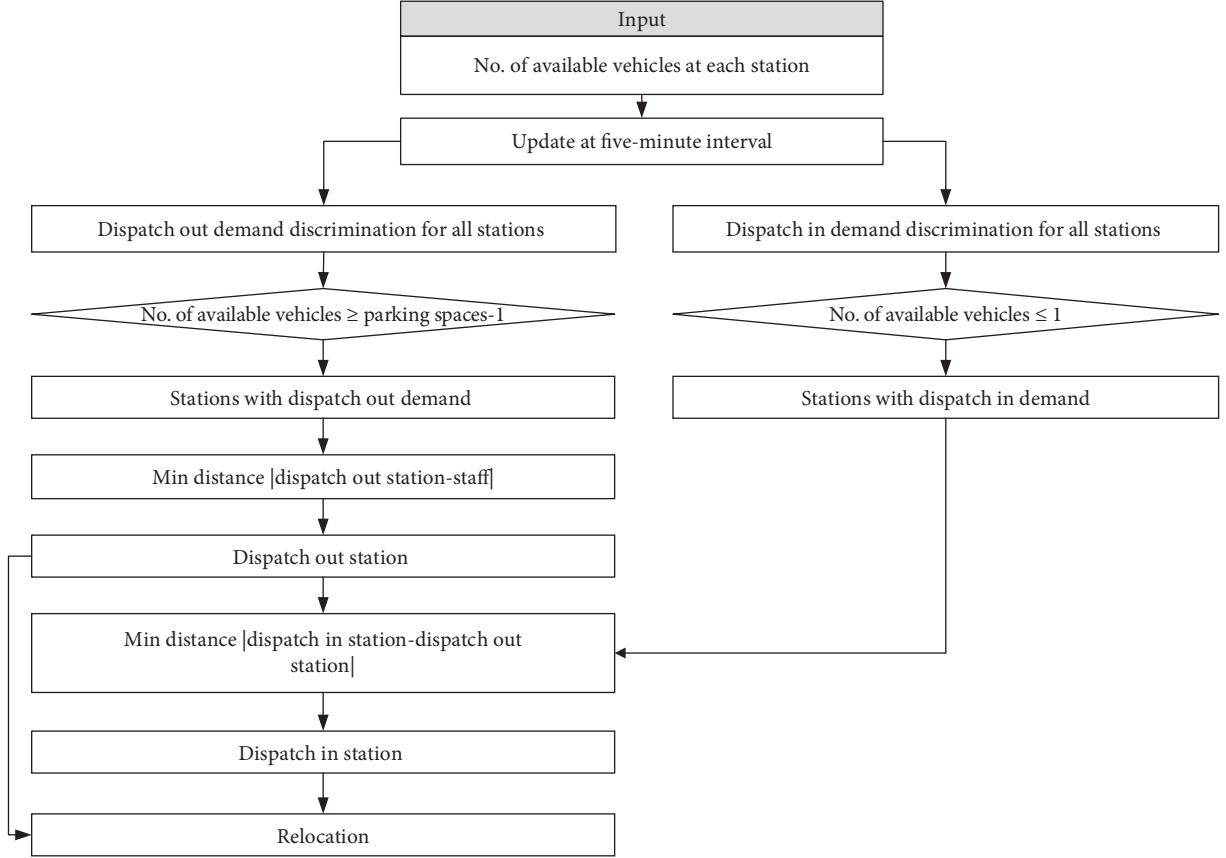


FIGURE 8: Relocation with no prediction.

TABLE 4: Results of three ways of relocation.

Factor scenario	Profit (Yuan)	Waste ratio (%)	Full ratio (%)	Empty ratio (%)
No dispatch	16895.00	0.00	3.03	3.32
Dispatch with no prediction	21356.00	16.49	1.54	2.48
Dispatch with prediction	24732.00	0.00	0.69	2.13

accurately so that to realize the real-time return demand forecast of one-way electric carsharing systems. Previous studies often model the characteristic variables based on historical order data, but they cannot draw the randomness and uncertainty of users' real-time demand. This paper introduced trajectory data and EAPP log data as the supplement source of the part which historical order data cannot reflect to explore the return demand of users.

A user return demand prediction model and a destination station prediction model are established to realize the dynamic real-time return demand prediction for the one-way electric carsharing system. The first model shows that as the user's current driving length or travel duration increases, the probability of the user returning the vehicle in 15 min increases. Meanwhile, when the user's historical driving mileage or travel duration increases, the probability decreases because the longer historical driving shows that this user prefers long-distance driving. Taken as a whole, the user's demand to return the vehicle first increases and then decreases with duration or mileage. If the current mileage or

duration is greater than the minimum value of 15% quantile of the user's historical data, the probability of the user returning the car in 15 minutes will increase slightly; if the current travel mileage or duration is greater than the user's historical 85% quantile or the maximum value, the probability of the user's demand to return the car will decrease.

The second model shows that the match of users' destination station is mainly decided by all users' destination station habits, which means that when the station s_j is used more frequently as a return station in all historical orders, for the current user's trajectory point, the station s_j within 15 minutes station set is more likely to be the return station for this trip. Similarly, the user is more likely to use his/her most frequent return station. From the perspective of driving duration, the smaller the time difference between the user's historical time returned to s_j and the current driving time, the greater the possibility of returning to s_j during this trip. Additionally, the trajectory difference has a great impact on the prediction of the destination station.

Finally, a case study with ten stations during a one-week operation period is conducted to test the benefit of the user return demand prediction model and the destination station prediction model. According to the compared results, relocation based on the above two models shows a great significance in reducing unnecessary relocation and increasing profit. What is more, the improvement of the full ratio and empty ratio indicates that the service level can be enhanced from the original base.

In brief, the real-time return demand prediction for the one-way electric carsharing system could effectively reduce unnecessary relocation and improve system profit. In practical applications, for each user in the system, the user real-time return demand prediction model is used to predict whether he/she will return the vehicle in the next 15 minutes. If there is a demand, the return station prediction model is further used to determine the most likely station for the user to return to. These models can be used to predict the return demand of a one-way carsharing system from the perspective of users, so as to improve the service quality of the system, reduce the system's operational cost, and attract a higher number of users.

Additionally, for different users, the demand characteristics show larger volatility and randomness over the time of day and weekday, and their return demand can be influenced by different factors, e.g., available parking sharing service [35]. Thus, the impact of time characteristics and other types of characteristics on vehicle return demand might be explored by advanced deep learning methods based on comprehensive analysis [36] in future studies. Moreover, this study used EV CARD in Shanghai as the study case. The patterns of the carsharing system in different cities might be different. This would be another future research direction.

Data Availability

Data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study has been supported by the National Key Research and Development Program of China (2018YFB1601000). Meanwhile, it was also supported by the Shanghai Science and Technology Innovation Action Plan Project (19DZ1209004 and 19DZ1208800).

References

- [1] T. D. Chen and K. M. Kockelman, "Carsharing's life-cycle impacts on energy use and greenhouse gas emissions," *Transportation Research Part D: Transport and Environment*, vol. 47, pp. 276–284, 2016.
- [2] A. S. Vasconcelos, L. M. Martinez, G. H. A. Correia, D. C. Guimarães, and T. L. Farias, "Environmental and financial impacts of adopting alternative vehicle technologies and relocation strategies in station-based one-way carsharing: an application in the city of Lisbon, Portugal," *Transportation Research Part D: Transport and Environment*, vol. 57, pp. 350–362, 2017.
- [3] S. Illgen and M. Höck, "Electric vehicles in car sharing networks - challenges and simulation model analysis," *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 377–387, 2018.
- [4] R. F. F. Lemme, E. F. Arruda, and L. Bahiense, "Optimization model to assess electric vehicles as an alternative for fleet composition in station-based car sharing systems," *Transportation Research Part D: Transport and Environment*, vol. 67, pp. 173–196, 2019.
- [5] J. Firnkorn and M. Müller, "What will be the environmental effects of new free-floating car-sharing systems? The case of car2go in Ulm," *Ecological Economics*, vol. 70, no. 8, pp. 1519–1528, 2011.
- [6] G. H. D. A. Correia, D. R. Jorge, and D. M. Antunes, "The added value of accounting for users' flexibility and information on the potential of a station-based one-way carsharing system: an application in lisbon, Portugal," *Journal of Intelligent Transportation Systems*, vol. 18, no. 3, pp. 299–308, 2014.
- [7] S. Le Vine, M. Lee-Gosselin, A. Sivakumar, and J. Polak, "A new approach to predict the market and impacts of round-trip and point-to-point carsharing systems: case study of London," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 218–229, 2014.
- [8] D. Jorge, C. Barnhart, and G. H. de Almeida Correia, "Assessing the viability of enabling a round-trip carsharing system to accept one-way trips: application to Logan Airport in Boston," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 359–372, 2015.
- [9] B. Boyaci, K. G. Zografos, and N. Geroliminis, "An integrated optimization-simulation framework for vehicle and personnel relocations of electric carsharing systems with reservations," *Transportation Research Part B: Methodological*, vol. 95, pp. 214–237, 2017.
- [10] X. Lu, Q. Zhang, Z. Peng, Z. Shao, H. Song, and W. Wang, "Charging and relocating optimization for electric vehicle carsharing: an event-based strategy improvement approach," *Energy*, vol. 207, pp. 118–285, 2020.
- [11] A. Angelopoulos, D. Gavalas, C. Konstantopoulos, D. Kyriadi, and G. Pantziou, "Incentivized vehicle relocation in vehicle sharing systems," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 175–193, 2018.
- [12] L. Wang, H. Zhong, W. Ma, Y. Zhong, and L. Wang, "Multi-source data-driven prediction for the dynamic pickup demand of one-way carsharing systems," *Transportmetrica B: Transport Dynamics*, vol. 8, no. 1, pp. 90–107, 2020.
- [13] D. Yu, Z. Li, Q. Zhong, Ai Yi, and W. Chen, "Demand management of station-based car sharing system based on deep learning forecasting," *Journal of Advanced Transportation*, vol. 2020, Article ID 8935857, 15 pages, 2020.
- [14] A. Carlier, A. Munier-Kordon, and W. Klaudel, "Mathematical model for the study of relocation strategies in one-way carsharing systems," *Transportation Research Procedia*, vol. 10, pp. 374–383, 2015.
- [15] D. Jorge, G. H. A. Correia, and C. Barnhart, "Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1667–1675, 2014.

- [16] S. Weikl and K. Bogenberger, "Relocation strategies and algorithms for free-floating car sharing systems," *IEEE Intelligent Transportation Systems Magazine*, vol. 5, no. 4, pp. 100–111, 2013.
- [17] M. Repoux, B. Boyaci, and N. Geroliminis, *Simulation and Optimization of One-Way Car-Sharing Systems with Variant Relocation Policies*, in *Proceedings of the 94th Annual meeting of Transportation Research Board*, pp. 1–19, Washington DC, USA, December 2015.
- [18] G. Santos and G. Correia, "A MIP model to optimize real time maintenance and relocation operations in one-way carsharing systems," *Transportation Research Procedia*, vol. 10, pp. 384–392, 2015.
- [19] P. Briest and C. Raupach, "The Car Sharing problem," in *Proceedings of the Twenty-Third Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pp. 167–176, ACM, San Jose, CA, USA, June 2011.
- [20] C. Li, L. Bai, W. Liu, L. Yao, and S. T. Waller, "Graph neural network for robust public transit demand prediction," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–13, 2020, In press.
- [21] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: a graph convolutional neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 258–276, 2018.
- [22] W. Herbawi, M. Knoll, M. Kaiser, and W. Gruel, "An evolutionary algorithm for the vehicle relocation problem in free floating carsharing," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2873–2879, IEEE, Vancouver, BC, USA, July 2016.
- [23] K. Huang, G. H. d. A. Correia, and K. An, "Solving the station-based one-way carsharing network planning problem with relocations and non-linear demand," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 1–17, 2018.
- [24] C. Morency, M. Trépanier, and B. Agard, "Typology of carsharing members," in *Proceedings of the Paper presented at the Transportation Research Board 90th Annual Meeting*, Washington DC, USA, January 2011.
- [25] T. D. Chen, K. M. Kockelman, and J. P. Hanna, "Operations of a shared, autonomous, electric vehicle fleet: implications of vehicle & charging infrastructure decisions," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 243–254, 2016.
- [26] G. Cao, L. Wang, Y. Jin, J. Yu, W. Ma, and Q. Liu, "Determination of the vehicle relocation triggering threshold in electric car-sharing system," in *Proceeding of the Chinese Intelligent Systems Conference*, pp. 11–22, Springer, Singapore, 2016.
- [27] X. Chen, H. Zheng, Z. Wang, and X. Chen, *Exploring Impacts of On-Demand Ridesplitting on Mobility via Real-World Ridesourcing Data and Questionnaires*, pp. 1–21, Transportation, Washington, DC, USA, 2018.
- [28] L. Wang, Y. Zhong, and W. Ma, "GPS-Data-Driven dynamic destination prediction for on-demand one-way carsharing system," in *Proceedings of the Transportation Research Board 97th Annual Meeting*, Washington, DC, USA, January 2018.
- [29] S. E. Edgell and S. M. Noon, "Effect of violation of normality on the t test of the correlation coefficient," *Psychological Bulletin*, vol. 95, no. 3, p. 576, 1984.
- [30] J. Algina and S. Olejnik, "Sample size tables for correlation analysis with applications in partial correlation and multiple regression analysis," *Multivariate Behavioral Research*, vol. 38, no. 3, pp. 309–323, 2003.
- [31] N. Wang, W. Zhang, X. Liu, and J. Zuo, "Inter-site-vehicle artificial scheduling strategy design for electric vehicle sharing," *Journal of Tongji University (Natural Science)*, vol. 46, no. 8, pp. 1064–1071, 2018.
- [32] L. Wang, Q. Liu, and W. Ma, "Optimization of dynamic relocation operations for one-way electric carsharing systems," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 55–69, 2019.
- [33] S. Hu, P. Chen, H. Lin, C. Xie, and X. Chen, "Promoting carsharing attractiveness and efficiency: an exploratory analysis," *Transportation Research Part D: Transport and Environment*, vol. 65, pp. 229–243, 2018.
- [34] G. Molnar and G. H. d. A. Correia, "Long-term vehicle reservations in one-way free-floating carsharing systems: a variable quality of service model," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 298–322, 2019.
- [35] S. Jian, W. Liu, X. Wang, H. Yang, and S. T. Waller, "On integrating carsharing and parking sharing services," *Transportation Research Part B: Methodological*, vol. 142, pp. 19–44, 2020.
- [36] K. Jintao, H. Zheng, H. Yang, and M. C. Xiqun, "Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 591–608, 2017.