Construction and Evaluation of a Return Prediction Model for One-Way Car Sharing

Ryota Saze[1][0000−0002−7311−6966], Manato Fujimoto[2,3][0000−0002−6171−5697], Hirohiko Suwa[1,3][0000−0002−8519−3352], and Keiichi Yasumoto[1,3][0000−0003−1579−3237]

[1] Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
{saze.ryota.sl4, h-suwa, yasumoto}@is.naist.jp
[2] Osaka Metropolitan University, Osaka, Osaka 558-8585, Japan
manato@omu.ac.jp
[3] RIKEN, Tokyo, Japan.

Abstract. One-way ECS (Electric Car Sharing Service) is attracting attention as a new sustainable mobility option in urban areas. On the other hand, the vehicle uneven distribution problem occurs in one-way ECSs due to their usage patterns. In this paper, we propose a vehicle return prediction model for vehicle relocation to solve this problem. In the proposed method, two machine learning models are created to predict where and when a user will return a vehicle using static information such as departure time and location, and dynamic information such as the vehicle's current location and direction of movement. The model is used to continually update the prediction results of vehicle returns during use, aiming for more accurate predictions. The proposed method has been evaluated using actual data from a one-way ECS and has achieved an accuracy of 0.93 for the prediction of stations to be returned. The method also achieved a MAE of 42.3 min. and MAPE of 47% for the prediction of return times.

Keywords: Electric Car Sharing · Machine Learning · Return Prediction · Vehicle Uneven Distribution Problem.

1 Introduction

In recent years, one-way electric car sharing services (hereinafter referred to as "one-way ECS") have become popular as a form of mobility in urban areas. In fact, in Japan, TIMES MOBILITY CO., LTD. has launched a service that can be used in Tokyo[4]. In addition to being flexible for short-distance travel in urban areas, one-way ECSs are expected to reduce CO2 emissions by reducing gasoline consumption [1]. On the other hand, in one-way ECS, there is a "vehicle uneven distribution problem" in which the demand for vehicles is unevenly distributed at certain times and places. If this situation occurs, the system's ability to meet the demand from users will be compromised, resulting in a decline in the rate

[4] https://share.timescar.jp/roadway/

at which the system meets usage requirements. A solution to this problem is "vehicle relocation" in which vehicles are relocated by the system operator or by the users themselves [6]. In order to properly perform this vehicle relocation, it is necessary to predict the future vehicle deployment and vehicle relocation in advance in such a way as to prevent the occurrence of vehicle uneven distribution problem.

There are some studies that predict vehicle demand for each station, but few studies predict future vehicle deployment by predicting the return station and return time for each vehicle in use.

In this paper, we propose a method for constructing a return prediction model that predicts the return station and return time of vehicles in real time by capturing the movement direction vector of each vehicle during use as a feature, and evaluate the constructed model by using data obtained from a car sharing service that is actually in operation. As a result, it is confirmed that the proposed method improves the return prediction results while the vehicle is moving. The contribution of this study is three folds:

1. First, we propose a return prediction model to predict the return station and return time for a one-way ECS.
2. Second, we evaluate using data obtained from a one-way ECS in real world.
3. Finally, we confirm that the prediction results are improved over time by using real-time vehicle movement direction vectors and other features as explanatory variables.

The rest of this paper is organized as follows. In Section II, we briefly review the existing studies related to our study. We then describe the challenges of a one-way ECS in Section III. Our proposed method for vehicle return prediction in one-way ECS is described in Section IV. We then present the evaluation method of our method and the results in Section V. Finally, Section VI concludes this paper.

2 Related Work

One-way ECSs have been the subject of various studies, including comparisons with the conventional round-trip car-sharing method, analysis of user usage characteristics, and research to gain insight into actual operations [3, 10, 14].

Huang *et al.* [4] compared two relocation methods, operator-based and user-based. Senju *et al.* [11] proposed a method of user-based vehicle relocation and confirmed that it improves the utilization demand fulfillment rate in simulations assuming a small-scale one-way ECS. Yang *et al.* [19] proposed an integrated model for operator-based vehicle relocation to minimize daily operational costs by optimizing request acceptance, relocation tasks, and relocation staff travel routes. Wang *et al.* [17] proposed a combined user-based and operator-based vehicle relocation method that considers the impact of dynamic constraints such as user demand, vehicle state of charge, and operating profit. Wang *et al.* [18] also proposed an adaptive co-location model that combines ECS and cycle share to increase ECS utilization. In this research, they achieved a 70% relocation cost reduction compared to staff-based relocation. Luo *et al.* [9] used multi-agent reinforcement learning to model and treat vehicle relocation tasks, and confirm that demand fulfillment and net benefits are improved using actual operational data.

In order to implement vehicle relocation, some studies have been made to predict the demand from users, and determine where to relocate. Luo *et al.* [8] extracted features from a graph representing various correlations (distance, point of interest, etc.) between stations of a one-way ECS using multi-graph convolution, and used them to predict the demand for vehicles per station. Yu *et al.* [20] performed LSTM-based vehicle demand forecasting using temporal features such as time of day, day of week, and weather. Huo *et al.* [5] used real-time data as input and built a data-driven optimization model that combines a stochastic expectation model and a linear programming problem to deal with the vehicle relocation problem. Wang *et al.* [16] used application log data of ECS users, historical usage data, as well as real-time station data and user's personal data. Wang *et al.* [13] proposed a method to improve the operational efficiency of ECS by dynamically changing the frequency of demand forecasting in ECS according to the amount of demand at each time of day.

On the other hand, there are several previous studies on predicting vehicle returns, which is another necessary element for vehicle relocation. Wang *et al.* [15] constructed a model to predict which stations are likely to return vehicles by calculating trajectories with high similarity by matching real-time GPS trajectory data with historical trajectory data. Liu *et al.* [7] utilized real-time vehicle trajectory data and application log data in addition to historical usage data to calculate the probability that a user will return a vehicle within 15 minutes and the probability of returning a vehicle to stations that can be reached within 15 minutes. The method proposed in their paper calculates the probability that a user will return a vehicle within 15 minutes and the probability of returning a vehicle to each station that can be reached within 15 minutes.

However, while there have been studies on vehicle return prediction based on real-time vehicle data, to the best of the authors' knowledge, there have been no studies that predict the return station and return time from the time a user starts using a vehicle, and that continue to improve the prediction accuracy over time by updating the prediction results during the user's use of the vehicle. In this study, we propose a method for predicting vehicle returns at regular intervals using static information such as the start time of use and the starting station, and dynamic information such as the real-time movement direction vector and location of the vehicle.

3 Challenges of a one-way ECS

3.1 One-way ECS

A one-way ECS is a type of ECS that allows vehicles to be returned to a station different from where they were rented. As described in Section 1, the vehicle

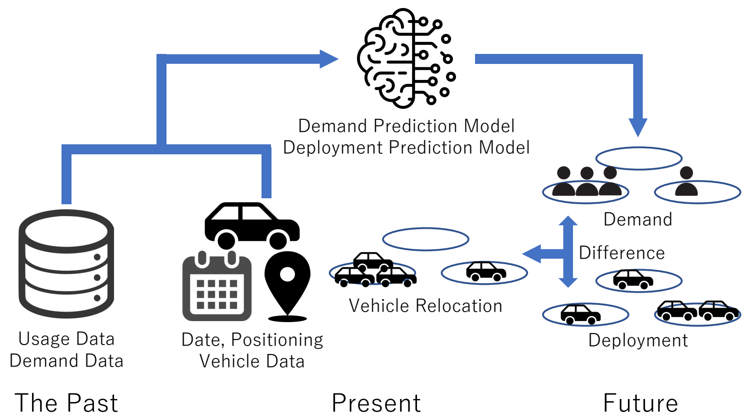

Fig. 1. Conceptual diagram of vehicle relocation in a one-way ECS.

uneven distribution problem occurs in one-way ECSs. This is due to the fact that the demand for vehicles is unbalanced temporally and spatially, resulting in stations with no vehicles or overflowing with vehicles. There are two ways to solve this uneven distribution problem: operator-based vehicle relocation, in which personnel hired by the system operator relocate vehicles, and user-based vehicle relocation, in which users are requested to reallocate their vehicles.

3.2 Vehicle Relocation

In order to perform appropriate vehicle relocation, it is necessary to predict the future utilization demand and vehicle deployment in the ECS. Fig. 1 shows the relationship between vehicle relocation and the prediction of utilization demand and vehicle deployment in a one-way ECS. As shown in the figure, vehicle relocation is possible by predicting both vehicle demand and vehicle deployment.

The demand prediction is necessary to determine in advance which stations will have a high concentration of demand in the future and to determine the necessity of relocation. On the other hand, vehicle deployment prediction is necessary to reduce unnecessary relocation of vehicles in combination with future demand predictions. For example, suppose a situation where a large demand is predicted for a certain station in the future. If the future vehicle deployment is not predicted, it is not known how many vehicles currently in use will be returned to that station, and unnecessary relocation of vehicles may occur. This would result in extra operational costs, and thus reduce the operational efficiency of the ECS.

3.3 Vehicle Deployment Prediction

Vehicle deployment prediction can be achieved by predicting when and where a vehicle in use will be returned to a station based on historical and real-time vehicle usage data. In this paper, we define this as “Vehicle Return Prediction”, and in particular, we define the task of predicting where a vehicle in use will be returned as “Return Station Prediction”, and the task of predicting when a vehicle will be returned as “Return Time Prediction”. Even if the return station prediction can be done with high accuracy, if the time of return is not known accurately, the future deployment of vehicles will be ambiguous. In the opposite case that the place of return is unclear and the time of return is accurate, it will be assumed that the future placement of vehicles will be incorrect. Therefore, both of these prediction tasks must be performed with high accuracy for proper vehicle return prediction.

4 Proposed method

We assume that for accurate vehicle return prediction, it is necessary to take into account the real-time direction of vehicle movement and the time since the vehicle was first boarded. Thus, we propose a return prediction model that predicts vehicle returns using real-time vehicle data, aiming to realize optimal vehicle relocation for solving the vehicle uneven distribution problem.

4.1 Summary of Proposed Methodology

The proposed method predicts the return of a user’s vehicle every 5 minutes from the start of the user’s use of a vehicle. The vehicle return prediction is based on static information such as the user’s departure station and departure time, and dynamic information such as the vehicle’s location and direction of movement, using machine learning. Note that the return station prediction and return time prediction are performed using different models, and therefore two machine learning models are used in the proposed method.

4.2 Features Used in The Model

Table 1 shows the information used to create the return prediction model. The *Start slot* indicates the time slot in which the user used the vehicle. In this study, one day is divided into 20-minutes time slots, and 72 slots are assumed per day. The *User ID* is an ID that identifies each user of the ECS. The *Ori station* and *Dst station* are the stations where the user started and finished using the ECS vehicle. In this study, each station is encoded as a number. The *Day of week* is the day of the week when used. We encoded *Day of week* as integers from 0 to 6, e.g., Monday is encoded as 0. The *Elapsed time* indicates the time elapsed since the start of use. The unit of Elapsed time is minutes. The *Latitude* and *Longitude* are the GPS position information at the time the data was saved. The *Movement direction* is the calculated azimuth angle moved from 5 minutes ago when the north direction is 0[°]. The *Total usage time* is the total time a user uses the ECS.

Table 1. Variables in return prediction models.

Column	Variable	Unit
Total usage time	Objective	Minutes
Dst station	Objective	Integer
Start slot	Explanatory	Integer
User ID	Explanatory	Integer
Ori station	Explanatory	Integer
Day of week	Explanatory	Integer
Elapsed time	Explanatory	Minutes
Latitude	Explanatory	Fractional
Longitude	Explanatory	Fractional
Movement direction	Explanatory	Fractional

4.3 Return Station Prediction Model

The return station prediction model was created by training XGBoost [2] on a classification task. For the return station prediction, the *Start slot*, *User ID*, *Ori station*, *Day of week*, *Elapsed time*, *Latitude*, *Longitude*, and *Movement direction* in the collected data were set as explanatory variables, and the *Dst station* was set as the objective variable.

4.4 Return Time Prediction Model

The return time prediction model was created by training XGBoost on a regression task. The return station prediction used the same explanatory variables as the return station prediction, but the objective variable was changed to *Total usage time*. It is assumed that the actual time of return is calculated using the total usage time and the start time of usage predicted by the model

5 Evaluation

The evaluation experiment was conducted using the one-way ECS service data[5] operated by the Nara Institute of Science and Technology (NAIST). We used the data between June 2020 and May 2021 to train two machine learning models of the proposed method and used the data from June 2021 to test the models. The return station prediction model, which is a classification task, was evaluated using the confusion matrix and the F value for each class. The regression task, return time prediction model, was evaluated using Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE), which was introduced to correctly evaluate the importance of the same error for long and short total usage.

[5] `https://naist-carshare.github.io/`

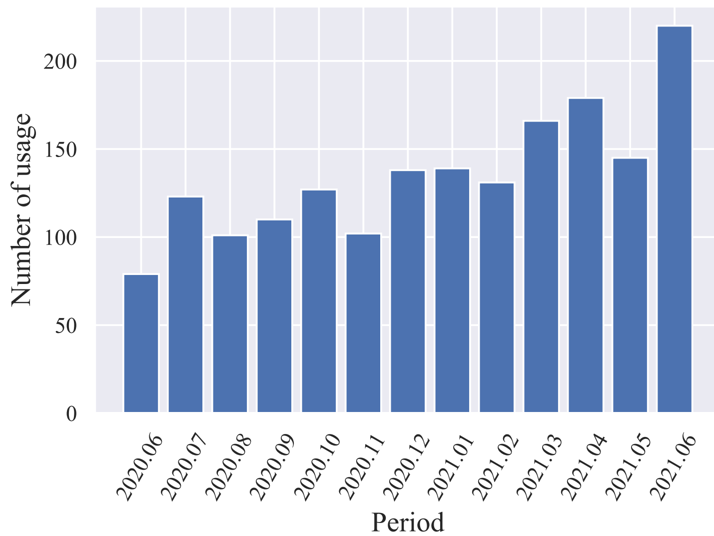

Fig. 2. Usage amount for each month.

Table 2. Number of times used at each original/destination station.

Ori \ Dst	School	Station	Facility
School	1208	202	28
Station	215	33	13
Facility	16	21	24

5.1 Data Collection

The data for this evaluation was collected from a one-way ECS that is in operation at NAIST. This is operated on a scale of three vehicles and three stations in total operation. The three stations are located in a school, near a railroad station, and near a research facility, respectively, and are referred to as "School", "Station", and "Facility". Each vehicle operated by the ECS is equipped with a GPS module, and GPS positioning data is stored every minute when the vehicle is in operation and available. The ECS is managed with a time slot similar to the one assumed in Section 4.1. Users bid for the time slots they wish to use using their own tokens. The user who has bid with the most tokens at the time of the start of the time slot wins the right to use the service. the number of usage for each month is shown in Fig. 2, and the number of times each station is used as an original and a destination is shown in Table 2.

5.2 Evaluation of Return Station Prediction

Prediction Result of Each Stations Table 3 shows the results of the return station prediction model for the test data, evaluated separately for each station.

Table 3. Prediction results of the return station prediction model.

	Precision	Recall	F-value
School	0.94	0.98	0.96
Station	0.77	0.58	0.66
Facility	0.96	0.84	0.90

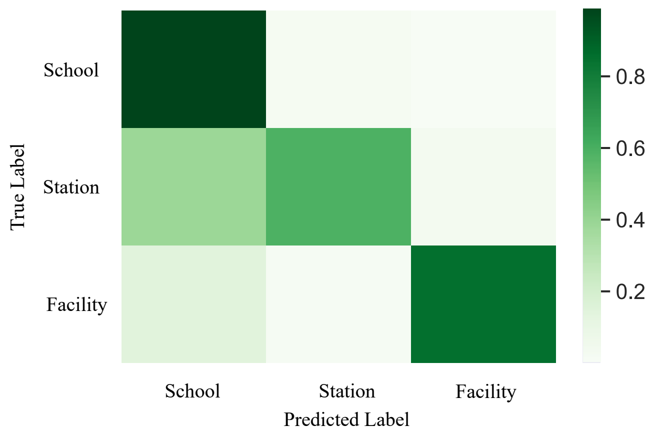

Fig. 3. Confusion matrix of return station prediction model's precision.

This table shows that the station prediction model is able to predict stations with a high degree of accuracy. In fact, the percentage of correct predictions for the entire test data was as high as 0.93. The reason for this high accuracy can be attributed to the large number of usage for School in the test data, and the high accuracy of the prediction for the use where the destination station is a School, as can be seen from the table.

Also, Fig. 3 shows a confusion matrix. Figure shows that the predictions were highly accurate when the target station was a Facility. The reason for this is that only a small number of users used the Facility as their destination station in the usage data, and the high accuracy is thought to be due to the fact that the *User ID* features were used as input. However, when the destination station is the Station, the prediction accuracy is lower than in the other cases. The reason for this is yet to be clarified, and future efforts should be made to improve the accuracy by adding distance information to each station as a feature value.

Time-series Change of Station Prediction Results The proposed method updates time-series information, such as the *Movement direction* from the start of use to the end of use, and makes predictions. Intuitively, the closer the end of use is, the closer the predicted return station is to the target station, so the prediction accuracy should improve. In fact, about half of the incorrect

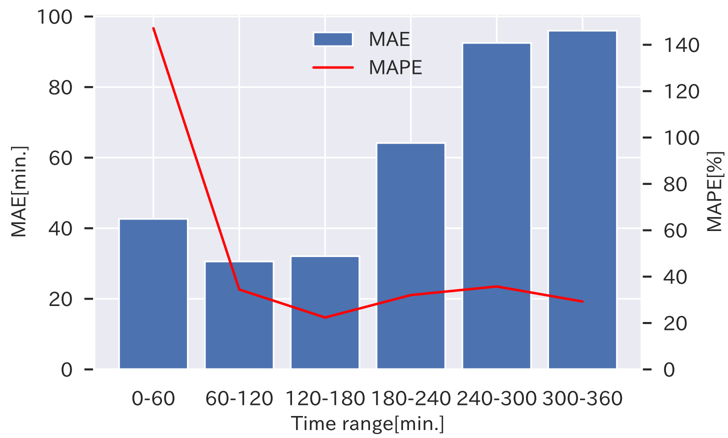

Fig. 4. Results of the return time prediction model for each time-range.

predictions, that were made at beginning of the use, actually became correct near the end of the use. Therefore, it is possible to improve prediction accuracy by continuing to update features and make predictions in a time-series manner. In the future, it is necessary to confirm what characteristics (travel route, user ID, etc.) are observed in these cases in which the predicted results did not change to correct ones.

5.3 Evaluation of Return Time Prediction

Accuracy of Return Time Prediction The prediction results of the return time prediction model for all test data were 42.3 min. for MAE and 47% for MAPE. Fig. 4 shows the results of comparing MAE and MAPE for the usage divided by the *Total usage time*. The bar graph in the figure shows MAE and the line graph shows MAPE. From the results of Fig. 4, MAPE was high at 146% for use within 60 minutes. This is a reasonable value considering that the MAE is 42.6 min, but considering the application to a real system, the value needs further improvement.

On the other hand, for the use between 60 and 180 minutes, the MAE was about 31 min. in both cases, and the MAPE was 34% and 22% lower than in the previous time period. In June 2021, the median of the total time of use [6] was 1.6 hours, and in nearly half of the cases, the *Total usage time* was between 60 and 180 minutes. This indicates that a practical model for predicting the return time has been obtained.

For data with a *Total usage time* of 180 minutes or more, the MAE is considered to be larger than in the above case because there was less training data. On the other hand, the MAPE was about 40%, which is thought to be due to

[6] `https://naist-carshare.github.io/logs/2021-07-02-log-202106/`

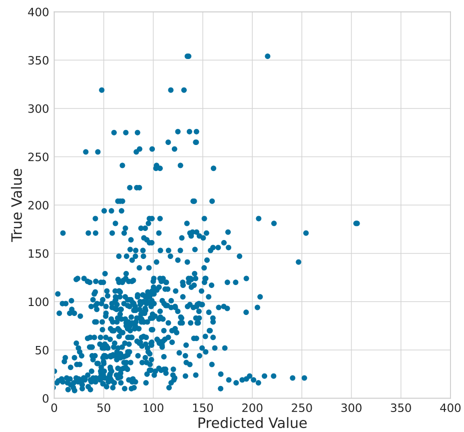

Fig. 5. Result for usage within 15 minutes from the start of use. This result shows the model's *Total usage time* prediction results for the input data within 15 minutes of the start.

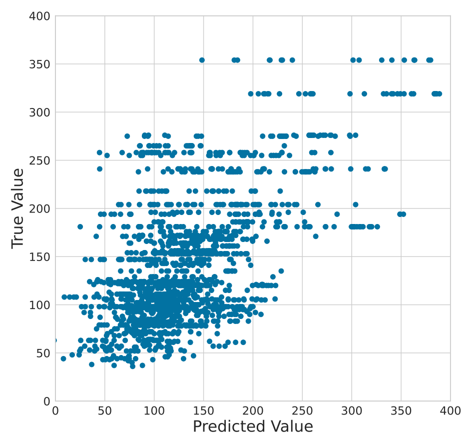

Fig. 6. Result for 15 minutes after the start - 15 minutes before the end. This result shows the model's *Total usage time* prediction results for the input data from 15 minutes after the start to 15 minutes before the end.

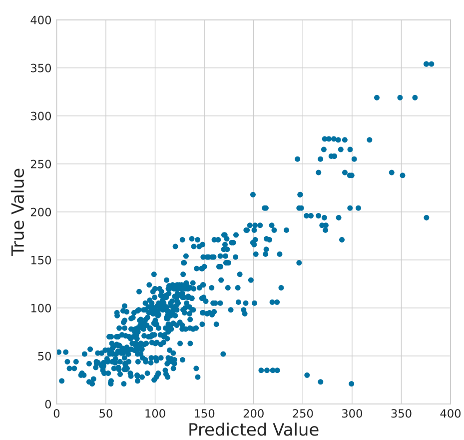

Fig. 7. Result for within 15 minutes from the end of use. This result shows the model's *Total usage time* prediction results for the input data within 15 minutes of the end.

the fact that the increase in *Total usage time* was larger than the increase in prediction error.

Time-series Changes in Time Prediction Results Figs. 5, 6, 7 shows the prediction results of the return time prediction model within 15 minutes of the user's start of use, from 15 minutes after the start to 15 minutes before the end, and within 15 minutes of the user's end of use. The closer to the diagonal line from the lower left to the upper right of the graph, the more accurate the prediction results are. In the Fig. 5, the model predict shorter use times than the true value in many cases. Fig. 6 show an increase in density around the diagonal compared to Fig. 5. Furthermore, Fig. 7 were clearly more concentrated around the diagonal than those for the earlier time periods. This indicates that the prediction accuracy of the actual return time improves with the *Elapsed time* of use. Therefore, it is shown that the proposed method improves prediction accuracy over time. On the other hand, there were cases where large prediction errors occurred even within the last 15 minutes of use, so it is necessary to improve the return time prediction model by analyzing the characteristics common to these uses.

6 Conclusion

In this study, we propose a vehicle return prediction model that is necessary for vehicle relocation to solve the vehicle uneven distribution problem in a one-way ECS. The proposed method performs vehicle return prediction by adding real-time vehicle location and direction information in addition to the user's past usage history, and keeps updating the prediction results while the vehicle is in use. The evaluation results showed that the proposed method achieved a high accuracy rate of 0.93 for the entire test data in the return station prediction. MAE and MAPE were smaller for the 60 to 180 minute time period than for the earlier time period. Furthermore, we confirmed that the prediction results became more accurate as the time of use increased.

Future plans include adding features such as the relative distance to each station to solve the problem that the return station prediction model is inaccurate for specific station. In addition, to solve the problem that the prediction accuracy of the return time prediction model does not improve even at the end of use for some users, it is necessary to analyze the characteristics of these usage. Furthermore, since the data used in this evaluation is for a small ECS, it is necessary to investigate whether similar results can be obtained for a larger ECS with a different station configuration. In addition, we would like to study vehicle relocation methods that use the proposed method, focusing on user-based vehicle relocation with incentives as described in the literature [12].

Acknowledgment

This work was supported in part by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research number JP21H03431 and JP21K11879.

References

1. Baptista, P., Melo, S., Rolim, C.: Energy, environmental and mobility impacts of car-sharing systems. empirical results from lisbon, portugal. Procedia-Social and Behavioral Sciences **111**, 28–37 (2014)
2. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
3. Curtale, R., Liao, F., van der Waerden, P.: Understanding travel preferences for user-based relocation strategies of one-way electric car-sharing services. Transportation Research Part C: Emerging Technologies **127**, 103135 (2021)
4. Huang, K., An, K., Rich, J., Ma, W.: Vehicle relocation in one-way station-based electric carsharing systems: A comparative study of operator-based and user-based methods. Transportation Research Part E: Logistics and Transportation Review **142**, 102081 (2020)
5. Huo, X., Wu, X., Li, M., Zheng, N., Yu, G.: The allocation problem of electric carsharing system: A data-driven approach. Transportation Research Part D: Transport and Environment **78**, 102192 (2020)
6. Jorge, D., Correia, G.H., Barnhart, C.: Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems. IEEE Transactions on Intelligent Transportation Systems **15**(4), 1667–1675 (2014)
7. Liu, D., Lu, J., Ma, W.: Real-time return demand prediction based on multisource data of one-way carsharing systems. Journal of Advanced Transportation **2021** (2021)
8. Luo, M., Du, B., Klemmer, K., Zhu, H., Ferhatosmanoglu, H., Wen, H.: D3p: Data-driven demand prediction for fast expanding electric vehicle sharing systems. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **4**(1), 1–21 (2020)
9. Luo, M., Zhang, W., Song, T., Li, K., Zhu, H., Du, B., Wen, H.: Rebalancing expanding ev sharing systems with deep reinforcement learning. In: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. pp. 1338–1344 (2021)
10. Park, S., Yu, W.: Analysis of system parameters for one-way carsharing systems. Transportation Letters pp. 1–11 (2021)
11. Senju, K., Mizumoto, T., Suwa, H., Arakawa, Y., Yasumoto, K.: Designing strategy for resolving maldistribution of vehicles in one-way car-sharing through active trip request to potential users. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 83–88. IEEE (2018)
12. Stokkink, P., Geroliminis, N.: Predictive user-based relocation through incentives in one-way car-sharing systems. Transportation Research Part B: Methodological **149**, 230–249 (2021)

13. Wang, G., Qin, Z., Wang, S., Sun, H., Dong, Z., Zhang, D.: Record: Joint real-time repositioning and charging for electric carsharing with dynamic deadlines. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 3660–3669 (2021)
14. Wang, G., Vaish, H.R., Sun, H., Wu, J., Wang, S., Zhang, D.: Understanding user behavior in car sharing services through the lens of mobility: Mixing qualitative and quantitative studies. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **4**(4), 1–30 (2020)
15. Wang, L., Zhong, Y., Ma, W.: Gps-data-driven dynamic destination prediction for on-demand one-way carsharing system. IET Intelligent Transport Systems **12**(10), 1291–1299 (2018)
16. Wang, L., Zhong, H., Ma, W., Zhong, Y., Wang, L.: Multi-source data-driven prediction for the dynamic pickup demand of one-way carsharing systems. Transportmetrica B: Transport Dynamics **8**(1), 90–107 (2020)
17. Wang, N., Guo, J., Liu, X., Liang, Y.: Electric vehicle car-sharing optimization relocation model combining user relocation and staff relocation. Transportation Letters **13**(4), 315–326 (2021)
18. Wang, N., Liu, Q., Guo, J., Fang, T.: A user-based adaptive joint relocation model combining electric car-sharing and bicycle-sharing. Transportmetrica B: transport dynamics **10**(1), 1046–1069 (2022)
19. Yang, S., Wu, J., Sun, H., Qu, Y., Li, T.: Double-balanced relocation optimization of one-way car-sharing system with real-time requests. Transportation Research Part C: Emerging Technologies **125**, 103071 (2021)
20. Yu, D., Li, Z., Zhong, Q., Ai, Y., Chen, W.: Demand management of station-based car sharing system based on deep learning forecasting. Journal of Advanced Transportation **2020** (2020)