

## Research Article

# Shared Bicycle Scheduling Model Based on Price Incentive Mechanism

Xiaoming Han, Yingshan Chong , and Qianwen Huangfu

College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Correspondence should be addressed to Yingshan Chong; 644170934@qq.com

Received 19 September 2018; Accepted 15 November 2018; Published 29 November 2018

Academic Editor: Elena Zaitseva

Copyright © 2018 Xiaoming Han 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.

In order to realize the rational delivery and effective dispatch of urban public self-vehicles, according to the randomness and time-varying of public bicycle demand, a demand forecast based on random forest and spatiotemporal clustering is proposed, and based on this, the user-based site rebalancing price incentive mechanism is implemented. Combining the demand of public bicycles with time factors, meteorological factors, associated sites, and other variables, using logarithmic optimization to reduce the impact of outliers, establish a random forest regression model. Secondly, based on this, a dynamic price incentive model is constructed to realize the rebalancing of user-based rental vehicles. The validity and feasibility of the dynamic price incentive model are verified by taking the historical data of public bicycle operation in the Bay Area as an example.

## 1. Introduction

With the advancement of rapid urbanization and motorization, the development of public bicycle systems at home and abroad has been promoted [1]. The emergence of urban public bicycles not only has greatly alleviated the traffic pressure, but has been gradually improved by the advantages of convenience, low carbon and environmental protection. The public bicycle system is an important part of the public transportation system, effectively solving the problem of “last 1 km”. In its rapid development, it provides convenient transportation for urban residents [2]. At the same time, due to the tidal traffic and the untimely dispatching of vehicles, it also faces the problem of “difficult to rent a car and return the car”, which hinders the development of public bicycles [3]. Therefore, it is essential to improve the scheduling mechanism and achieve the “balance of the car with rent and availability”.

At present, a lot of researches have been done on the scheduling problem of public bicycle systems at home and abroad. The scheduling of scholars' research focuses on the optimization of transportation vehicle routes. Yajun Zhou et al. [4] use the Markov chain model to predict the demand to reduce the scheduling cost. Schuijbroek J et al. [5] carried out an algorithm improvement for vehicle path optimization.

Kloimullner et al. [6] proposed an optimization model that minimizes the number of bicycles to be dispatched. The existing research is to assume the static demand of each site demand, which is inconsistent with the randomness of customer demand in the public bicycle system, which is not conducive to the long-term sustainable development of the enterprise [7]. In order to effectively alleviate the problem of “returning a car to rent a car” in public bicycle systems and improve the utilization rate of public bicycles [8], a new user-based rebalancing mechanism is proposed.

Based on the demand forecasting of random forests [9], this paper proposes a price incentive mechanism based on user self-balancing and considers the randomness of user demand to achieve the balance of supply and demand of public bicycle system. While improving service levels, the need for dispatchers to relocate vehicles is reduced.

## 2. Public Bicycle Demand Distribution Characteristics

*2.1. Time Distribution Characteristics.* The use of public bicycles is affected by time [10]. This paper uses the bicycle demand data of No. 70 site in Bay Area in 2015 to analyze the timing variation. The results are shown in Figure 1. Traffic

TABLE I: Correlation coefficient between demand and meteorological factors.

	temp	humid	Wind	Cloud	weather
Correlation coefficient	0.19	-0.12	-0.26	0.06	0.36

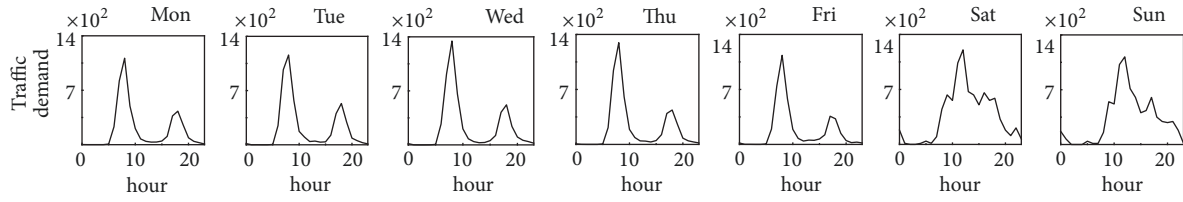
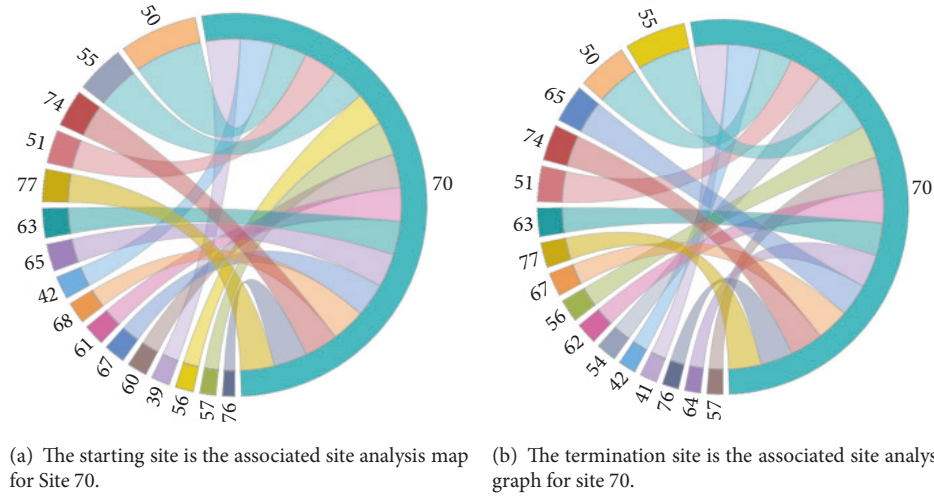


FIGURE 1: The seasonal variation of demand in the Bay Area.



(a) The starting site is the associated site analysis map for Site 70.

(b) The termination site is the associated site analysis graph for site 70.

FIGURE 2: Analysis of the starting site of the 70th station riding.

conditions on weekdays are similar including morning peak hours, daytime hours, night peak hours, and night time periods, while weekends and holiday periods are similar including night time, travel time, and late time. The total traffic volume on weekdays is much larger than that on weekends and holidays, and the traffic volume during peak hours and travel time is much larger than that of other time periods. Although bicycle usage varies from site to site, they are all closely related to time.

**2.2. Meteorological Distribution Characteristics.** Bicycles are a type of vehicle that is significantly affected by weather [11, 12]. Table 1 shows the correlation analysis between the demand for Site No. 70 of the Bay Area Public Bicycle System. It can be seen from the table that there is a correlation between the demand for bicycle and five meteorological factors, and there is a significant positive correlation with temperature, cloud amount, and weather conditions, while the same humidity and wind speed have a negative correlation. Bicycle demand is most correlated with wind speed and weather conditions, with  $-0.26$  and  $0.36$ , respectively.

**2.3. Related Site Distribution Characteristics.** The user rents a car from a certain rental point, and after a certain time, to

another rental point nearby [13], this kind of flow behavior makes a certain relationship between the lease point and the lease point: one of the lease points at the current moment and the other lease. There is a certain correlation between the number of rented cars before a certain time. Specifically, the associated site of a site is determined by the number of trips between the two sites. Therefore, when predicting the demand for a site for a certain period of time, the input variables should also consider the demand for other sites associated with the site, as shown in Figure 2, the relationship between Site 70 in the US Gulf Area and other sites. It can be seen from Figure 2 that most users who start to ride from station 70 will end their riding at stations 50 and 55. When the user ends the riding at station 70, the starting stations are generally stations 50 and 55.

**2.4. Site Inventory Forecasting Model.** The bicycle site inventory refers to the number of bicycles owned by the site at the current moment, which is the difference between the number of borrowed vehicles and the number of returned vehicles at the current time. Accurate and efficient forecasting of bicycle site inventory is a prerequisite for price incentives. Because public bicycle travel is affected by time, weather, and associated site conditions, it has strong randomness, time

variation, and nonlinearity. Aiming at the above problems, a regression prediction algorithm for public bicycle inventory based on random forest is proposed, and logarithmic optimization is used to reduce the impact of outliers on prediction performance.

The random forest algorithm is a decision tree classifier fusion algorithm proposed by Breiman [14]. It belongs to the bagging algorithm in Ensemble Learning and is a classification and regression algorithm [15]. The basic idea of the random forest regression algorithm is to use the self-sampling method (Bootstrap) to randomly sample the original samples and construct several decision trees. When each decision tree is split, the optimal attributes are selected from some attributes for splitting. Finally, combining these decision trees constitutes a random forest, and the average of all decision trees in the random forest is used as the final prediction result. Random forest regression can be seen as a strong predictor (random forest) integrated by multiple weak predictors (decision trees).

Random forest selects  $n$  samples by self-sampling method by borrowing the bagging idea of integrated learning. Different from bagging selects all features for decision tree generation, random forest selects  $k$  features from all feature attributes. Use these samples and features to construct multiple decision trees to form a "forest." Its "random" characteristics are manifested in the random selection of samples, and the selection of feature attributes is also random. Through the random introduction of these two, the random forest has been overfitting and has good antinoise ability. A random forest is a single classifier consisting of a series of training samples  $X$  and random vectors  $\beta_k$  with a sample size of  $K$ . A set of  $\{h(X, \beta_k), K=1, 2, \dots, K\}$ . Each decision tree model  $h(X, \beta_k)$  has one vote to select the final classification result. The classification decision is as shown in

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y) \quad (1)$$

$H(x)$  represents random forest classification results,  $h_i(x)$  represents a single classification result,  $Y$  represents a classification target, and  $I$  is an indicative function. This formula is a classification problem of random forests, that is, the majority of the results of each decision tree are the final results. For the regression problem of random forests, the expectation of the results of each decision tree can be selected as the final result.

The general training process for random forests is as follows:

- (1) Resampling using the bootstrap method from the sample training set, randomly generating  $k$  sample training sets  $\theta_1, \theta_2, \dots, \theta_k$ . Each sample training set corresponds to a corresponding decision tree  $T(\theta_1), T(\theta_2), \dots, T(\theta_k)$ . Assuming that the number of samples in the sample training set is  $n$ , and each sample is subjected to random sampling with a return, the probability of extraction is  $1/n$ , and the extraction is repeated  $n$  times. In the training sample set, each sample is not extracted. The probability is  $(1-1/n)^n$ . When

the number of training samples is large enough,  $(1 - 1/n)^n \approx 0.368$ , the probability of not being extracted is about 37%, which guarantees the difference of the training set.

- (2) Assuming that the sample training set features are  $M$ -dimensional,  $m$  features are randomly selected from the  $M$ -dimensional features as the split feature set of the current node, and the features with the smallest variance among the  $m$  features are selected for splitting, and no pruning is required. This tree is fully grown. (The requirement here is that  $m$  is much smaller than  $M$ , i.e.) The formula for calculating the variance is as follows:

$$V = \frac{1}{n} \sum_{i=1}^N (y_i - \mu)^2 \quad (2)$$

where  $n$  represents the total number of samples,  $y_i$  represents the label of a sample, and  $\mu$  represents the mean of all samples.

- (3) Repeat the above steps to establish  $K$  decision trees to form a random forest.
- (4) Output the result. For the new data that is not predicted, the prediction is the random forest's result, and the prediction result is the average of the calculation results of each decision tree.

Random forest regression can be regarded as a strong predictor integrated by weak predictor generated by randomly sampling samples and randomly selecting features. It almost inherits all the advantages of decision trees and avoids overfitting problems and reducing generalization errors. There is no need to normalize the input variables. In the public bicycle raw data, weather characteristic data such as temperature, wind speed, humidity, etc. are numerical variables, and data such as holidays and working days are categorical variables. In the random forest model, we can use these features directly as input variables without the need for additional conversions. However, the regression accuracy of random forests depends heavily on the accuracy of a single decision tree, ensuring that the predictive power of a single decision tree can improve the classification effect of random forests. Therefore, logarithmic optimization is used to reduce the influence of outliers on decision tree performance and clustering algorithm is used to cluster training samples. Different categories use different prediction models, and the results of each model are summarized to obtain the final prediction.

### 3. Price Incentive Model Based on User Rebalancing

**3.1. Model Parameter.** If appropriate incentives are provided, the user may make a contribution to the rebalancing of the vehicle [16]. Currently, it is possible to communicate with users and pay incentives through existing infrastructure (self-service terminals or mobile applications) [17]. Therefore, a user-based rebalancing price incentive mechanism is proposed to dynamically select the amount of money provided

TABLE 2: Symbol Definition.

symbol	description
$N_s$	N neighbors around the S site
$Q(t)$	The demand for the site at t time
$M(t)$	The number of users returning at time t
$N(t)$	Change the amount of car due to the acceptance of the incentive price
$\lambda$	Incentive price coefficient
$\mu$	Distance perception coefficient

TABLE 3: Site space relative distance table (KM).

Station	69	70	71	72	73	74	75	76
69	0	0.233	1.612	1.829	1.722	1.052	0.962	1.197
70	0.233	0	1.632	1.852	1.738	1.048	0.971	1.219
71	1.612	1.632	0	0.571	0.631	1.601	1.064	0.411
72	1.829	1.852	0.571	0	1.194	2.078	1.539	0.819
73	1.722	1.738	0.631	1.194	0	1.194	0.861	0.692
74	1.052	1.048	1.601	2.078	1.194	0	0.541	1.260
75	0.962	0.971	1.064	1.539	0.861	0.541	0	0.723
76	1.197	1.219	0.411	0.819	0.692	1.260	0.723	0

to the user to change the user's return location in a manner that improves the overall service level.

In the public bicycle rental system, some of the technical terms are symbolized, as shown in Table 2.

**3.2. Incentive Mechanism Model.** It is set  $P_s^n(t)$  that, during the time period from  $t \in T$  to  $t+1$ , the rental system provides the customer with the incentive price for changing the vehicle from the original planned returning station  $s$  to the returning station  $n$ , and  $f(x)$  is the site inventory vehicles at time  $t$ . The number of stock vehicles,  $f(x)$ , is related to the site demand  $Q(t)$ , the number of users returning the vehicle  $M(t)$ , and the number of returned vehicles  $N(t)$  that change the returning location by accepting the incentive price. The riding incentive price of each riding user obeys the uniform distribution  $P \sim U[0, P_{max}]$ , and the number of returned vehicles that accept the incentive price and change the returning location  $N(t)$  satisfies formulas (3)-(6):

$$f(x) = Q(t) - M(t) - N(t) \quad (3)$$

$$\pi_{s,n} = \lambda P_{s,n} - \mu d_{s,n} \quad (4)$$

$$\delta(t) = Q(t) - M(t) \quad (5)$$

$$\sum_{n \in N_s} \pi_{n,s} \delta_n = 0.8N(t) \quad (6)$$

Equation (3) indicates that the change in the site vehicle inventory level  $f(x)$  is related to the site demand amount  $Q(t)$ , the number of users returning the vehicle  $M(t)$ , and the number of returned vehicles  $N(t)$  that accepts the incentive price and changes the returning location. Equation (4) represents the probability of accepting the price incentive to change the returning station from  $s$  to  $n$ , and the price incentive coefficient  $\lambda$ , the user changes the incentive price of the

returning station  $P_{s,n}$ , the distance perception coefficient  $\mu$ , and the distance between stations  $d_{s,n}$  related. Equation (5) represents the number of vehicles that need to be stimulated by the neighboring site through price incentives. Equation (6) indicates that the neighboring site region  $N_s$  of the site  $s$  can meet the needs of 80% of users. From the relational formulas of (3) to (6), combined with the objective function and constraints of the price incentive model, the price incentive mechanism model can be described as

$$\min \left[ \sum_{n=1}^N Q(t) - M(t) - N(t) \right] \quad (7)$$

s.t

$$\sum_{n \in N_s} \pi_{n,s} \leq 1, \quad \forall s \in S, \quad \forall t \in T \quad (8)$$

$$0 \leq P_{s,n}(t) \leq P_{max}, \quad \forall s \in S, \quad \forall t \in T \quad (9)$$

The objective function (7) minimizes the sum of the unbalanced vehicles of the site of the leased site neighbor network; the constraint (8) indicates that the sum of the probability that the user chooses to accept the price incentive of the neighboring site is not more than 1; the constraint (9) indicates that the rental system provides. The price incentive to the user is not greater than the maximum price incentive value  $P_{max}$ .

## 4. Experiment Analysis

**4.1. Experimental Data Analysis.** The data used in this paper is from August 2013 to August 2016 in the Gulf of California area. There are 8 sites in a common area, and the spatial distance between the sites is shown in Table 3. Now the price incentive period  $T$  is one day, which is divided into 13 time

TABLE 4: Initial site demand information data table.

Station	Number of station lock	Optimal stock	Initial stock	Dispatching vehicles
69	23	19	4	15
70	19	11	5	6
71	19	5	12	-7
72	23	5	3	2
73	15	7	12	-5
74	23	2	9	-7
75	19	2	15	-13
76	19	5	7	-2

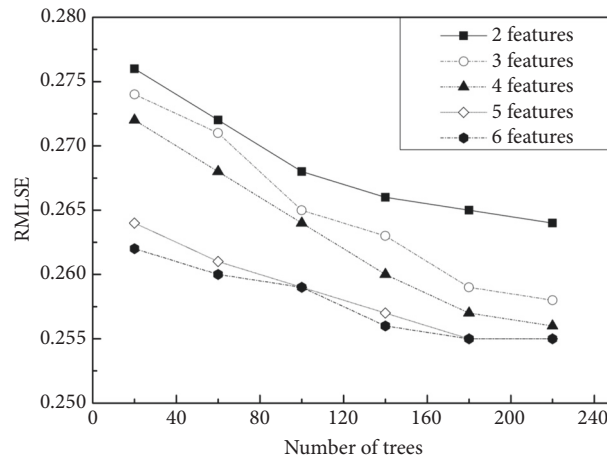


FIGURE 3: Random forest parameter selection and error.

segments, and the price incentive opening period is from 9:00 to 21:00 every day, and the price period is opening from 9:00 to 21:00 every day by dividing the T discretization into 13 time segments.

4.2. *Experimental Parameter Selection.* There are two main factors affecting the performance of random forest algorithms, the number of random forest algorithm trees and features used in building trees. By analyzing the affecting factors of shared bicycle inventory, including time, spatial, meteorological, and associated site screening for the factors that have a greater impact on it, Figure 3 shows the results of constructing a random forest prediction model using different parameters. Considering the time consumption and the error of the model, the number of trees is 180 and the number of features is 5 to construct the prediction model.

4.3. *Analysis of Results.* Through the logarithm-optimized random forest model to solve the inventory quantity, the

optimal inventory level of each station at different times is obtained. The optimal inventory level, initial inventory, and number of vehicles scheduled for each station at the initial time  $t=0$  are shown in Table 4.

Analyze the influence of time factors, meteorological factors, and associated sites on the demand of public bicycle systems, use the log-optimized random forest which is used to find the optimal inventory level of each site, in the price incentive mechanism, and use the optimized ion swarm algorithm to solve. The site price incentive matrix obtains the system price incentive matrix at the initial moment and loops the solution at the next moment to provide the customer with the real-time incentive price of the site, as shown in Table 5.

4.4. *Evaluation Index.* For the evaluation indicators, the shared bicycle site service level is used to calculate, and the formula is as follows:

$$\text{System balance} = \frac{\text{Number of vehicles required} - \text{Number of unavailable vehicles}}{\text{Number of vehicles required}} \tag{10}$$

When the maximum incentive price is increased, the service level of the public bicycle system site has been

significantly improved; the effectiveness of dynamic price incentive model is proved to improve the imbalance of the



TABLE 5: Incentive price for each site at the initial moment (in US dollars).

Station	69	70	71	72	73	74	75	76
69	0	0	0	1.9	1.7	0	0	0
70	0	0	0	5	0	3.2	0.5	0
71	0	4.9	0	0.3	3.5	0.3	1.3	0
72	3.5	5	0	2.4	0.8	3.1	2.3	5
73	4.8	2.7	0.3	4.5	0	5	4	5
74	3.6	0	0	4.1	2.9	0	3.8	4.0
75	0.66	0	0	0.4	0	2	0	4.8
76	0	3.5	0	1.4	5	0	0	0

The abscissa is the original planned returning station, and the ordinate is the returning site after the price incentive.

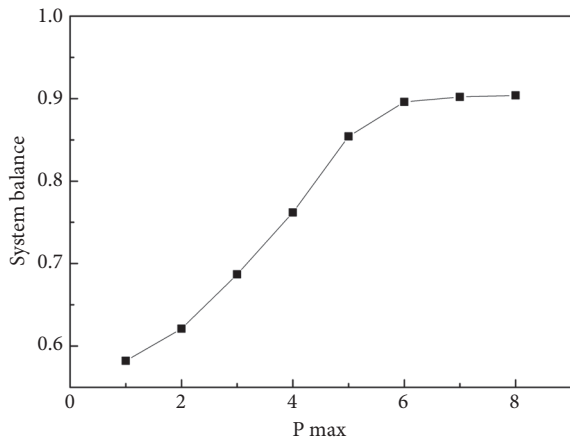


FIGURE 4: Balance between the maximum incentive price and the rental system.

public bicycle rental system. In order to further analyze the simulation results, by analyzing the impact of different price incentives on the system to find the best maximum price incentive value, the experimental results are shown in Figure 4.

Figure 4 shows that the maximum incentive price is raised from 2 to 5, and the balance of the rental system is increased from 62% to 90%, which greatly reduces the operating costs of the leasing companies. On this basis, three sets of tests are continued, and the test results show that appropriately increasing the maximum incentive price  $P_{max}$  can greatly improve the balance of the rental system, but excessively increasing the value of  $P_{max}$  does not significantly improve the system balance; the total system cost is significantly related to the size of  $P_{max}$ . When the  $P_{max}$  value is too small, the price incentive mechanism does not. To achieve the incentive effect, the customer does not choose to change the returning site, the vehicle can not be effectively configured, and the system loss cost is high; when the  $P_{max}$  value is too large, the balance of the rental system does not increase greatly with the increase of the incentive, and the larger. The  $P_{max}$  value brings too much incentive payment to the enterprise, and the total system cost is too high. Appropriately increasing the maximum incentive price  $P_{max}$  can greatly improve the balance of the rental system and

improve customer service level and vehicle utilization. When  $P_{max}$  reaches 5 and continue to increase  $P_{max}$  value, the system balance does not improve significantly, the growth rate slows down, and the maximum incentive price  $P_{max} = 5$  system balance of 90% can meet the needs of leasing business operations management.

## 5. Conclusion and Future Work

Aiming at the problem of leased self-vehicle configuration in urban public bicycle system, this paper proposes a vehicle scheduling model based on dynamic price incentive mechanism, which realizes user-based self-balancing, can realize long-term configuration of rental vehicles, and analyzes by random forest algorithm. The demand for rental vehicles has the following conclusions:

- (1) The factors of time and meteorological and associated sites are important to influence the behavior of public bicycles. Due to the randomness in use of public bicycles, the quantity demanded has abnormal values that are reduced by logarithmic optimization. This algorithm makes the random forest-based prediction model more accurate and provides a basis for the next step of self-balancing based on user price incentive.
- (2) The dynamic price incentive matrix is obtained through the price incentive model. The price incentive matrix can help public bicycle rental companies provide real-time dynamic price incentive decisions to customers.
- (3) The test results show that an appropriate increase in the maximum incentive price  $P_{max}$  can significantly improve the balance of the rental system. The maximum incentive price was raised from 2 to 5, and the balance of the rental system was increased from 62% to 90%, which greatly improved the balance of the rental system.

The above conclusions show that the model and method can be better to complete the vehicle configuration in the lease system, realize the balance of supply and demand of the lease network, and provide theoretical basis and practical reference for the study of vehicle allocation problem in the public bicycle rental system.

## Data Availability

The data used to support the findings of this study are from (<https://www.fordgobike.com/>). This is a public data set.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Funded by Shanxi Natural Science Foundation (2015011052).

## References

- [1] E. Fishman, "Bikeshare: a Review of Recent Literature," *Transport Reviews*, vol. 36, no. 1, pp. 92–113, 2016.
- [2] A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs, "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 455–466, 2010.
- [3] C. Contardo, C. Morency, and L. Rousseau, "Balancing A Dynamic Public Bike-Sharing System," Technical Report, CIR-RELT, 2012.
- [4] Y. Zhou, L. Wang, R. Zhong, and Y. Tan, "A Markov Chain Based Demand Prediction Model for Stations in Bike Sharing Systems," *Mathematical Problems in Engineering*, vol. 2018, no. 1, Article ID 8028714, pp. 1–8, 2018.
- [5] J. Schuijbroek, R. C. Hampshire, and W.-J. van Hoes, "Inventory rebalancing and vehicle routing in bike sharing systems," *European Journal of Operational Research*, vol. 257, no. 3, pp. 992–1004, 2017.
- [6] C. Kloimüller, P. Papazek, B. Hu, and G. R. Raidl, "Balancing bicycle sharing systems: an approach for the dynamic case," in *Evolutionary computation in combinatorial optimization*, vol. 8600 of *Lecture Notes in Comput. Sci.*, pp. 73–84, Springer, Heidelberg, 2014.
- [7] L. Cagliero, T. Cerquitelli, S. Chiusano, P. Garza, and X. Xiao, "Predicting critical conditions in bicycle sharing systems," *Computing*, vol. 99, no. 1, pp. 39–57, 2017.
- [8] C. Fricker and N. Gast, "Incentives and regulations in bike-sharing systems with stations of finite capacity," *Adaptation and Self-Organizing Systems*, 2012.
- [9] W. Y. Loh, "Classification and regression trees," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011.
- [10] W. El-Assi, M. Salah Mahmoud, and K. Nurul Habib, "Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto," *Transportation*, vol. 44, no. 3, pp. 589–613, 2017.
- [11] K. Gebhart and R. B. Noland, "The impact of weather conditions on bikeshare trips in Washington, DC," *Transportation*, vol. 41, no. 6, pp. 1205–1225, 2014.
- [12] J. Corcoran, T. Li, D. Rohde, E. Charles-Edwards, and D. Mateo-Babiano, "Spatio-temporal patterns of a Public Bicycle Sharing Program: The effect of weather and calendar events," *Journal of Transport Geography*, vol. 41, pp. 292–305, 2014.
- [13] T. Raviv and O. Kolka, "Optimal inventory management of a bike-sharing station," *Institute of Industrial Engineers (IIE). IIE Transactions*, vol. 45, no. 10, pp. 1077–1093, 2013.
- [14] A. Liaw and M. Wiener, "Classification and regression by random Forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [16] M. Benchimol, "Balancing the stations of a self service 'bike hire' system," *RAIRO - Operations Research*, vol. 45, no. 1, pp. 37–61, 2011.
- [17] T. Raviv, M. Tzur, and I. A. Forma, "Static repositioning in a bike-sharing system: models and solution approaches," *EURO Journal on Transportation and Logistics*, vol. 2, no. 3, pp. 187–229, 2013.

