Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

A User-Based Relocation Strategy for Free-Floating Bike Sharing System: A case study of Beijing, China

S.K. Jason Chang[a], Zhuming Liu[a]

National Taiwan University. No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, R.O.C

Abstract

Over the past years, the free-floating bike sharing (FFBS) system has expanded rapidly all over the world. Such system allows users to rent or return a bike almost anywhere within the operating area. This characteristic makes the traditional operator-based relocation strategy become economically less attractive. In this paper, a user-based relocation system design to solve those imbalance problems through a monetary incentive to encourage users to return bikes at a nearby zone, where a shortage of bikes occurs. Firstly, based on GPS data, spatial clustering was used to construct transit Origin-Destination matrices. Additionally, in order to predict upcoming Origin-Destination traffic volume at certain place, depending on different temporal factors, a demand model was built using back-propagation neural network (BPNN). The prediction result was used to compute the redundant bike that could be dispatched and the insufficiency at each district. Then, combined with a user participation pattern that established based on an online survey, a dispatching scheme that considers minimizing both the incentives payout and the total bike shortfall developed and evaluated. In the end, a case study of FFBS system in Beijing was performed to show the effect of the proposed system within different types of regions.

Keywords: Free-Floating Bike Sharing System; User-based Relocation; Partition clustering; Demand prediction

Corresponding author. Tel.: +886-909-872-749; fax: +886-2363-9990.
E-mail address: r06521532@ntu.edu.tw

1. Introduction

Bike Sharing System (BSS) as a green and convenient transportation mode, can provide users as an alternative to those always crowded transit. BSS is growing rapidly in recent year around the world, see Shaheen, S. A. et al. (2010).

The station-based BSS, which have docks or stations, are widely used in various countries, but always criticized for its station. When the users do not know the location of the nearby station, or that station has already been empty or full, it become inconvenient for users to use that system. Additionally, Fishman, E. et al. (2014) found that convenience is the biggest motivation for users to choose the bike sharing system. In recent years, so-called free-floating bike sharing (FFBS) system expanded rapidly in many countries, especially China. Since there are dockless, the bike can be rented or returned almost anywhere within the operating area. To some extent, it solves the inconvenience problem in station-based system. However, due to the characteristics of "free-floating", bikes are always distributed imbalance. Therefore, relocation is necessary.

In general, the characteristics of FFBS cause unexpected high operating cost when the administrators choosing trucks for redistribution. In the existing research, there have some intelligent relocation methods applied for reducing the using of trucks, see Pfrommer, J. et al. (2014), Fricker, C. et al. (2014) and Singla, A. (2015). From the perspective of behavioral economics, it has been trying to solve the imbalance problem by offering monetary incentive to encourage users to redistribute bikes spontaneously. However, all of the above research were based on the station-based BSS. For the FFBS system, see Reiss, S. and Bogenberger, K. (2017), based on 10% to 80% fare discount for the user, a user-based relocation scheme was created. This scheme can combine with the operator-based scheme when more than 15% of the fleet need a relocation. In practical applications, Velib' system in Pairs introduced bonus V+, see Papanikolaou, D. (2011). If user could return the bike in the station that is 60 meters higher than the renting station, 15 minutes of free time will be offered as a bonus. Similarly, Mobike (2017), one of the most popular FFBS company in China, drew on the Pokémon Go design, monetary reward was used to attract users to find bikes which have low turnover rate then help company dispatch them to those popular places.

Based on previous research, the goal of this research is to develop a relocation system, by providing monetary incentive, "commission" users return the bike to the district whose inventory will have a shortfall in the upcoming period; thereby, the FFBS system could be rebalanced.

The structure of the proposed system shown in Fig. 1. Firstly, based on the historical trip data, K-means as a spatial clustering method used to construct Origin-Destination matrices. Additionally, in order to predict the volume of upcoming trips at certain district, depending on different temporal factors, a demand prediction model was built using back-propagation neural network (BPNN). Prediction result was used to compute the redundant bike that could be dispatched and the insufficiency at each district. Finally, by solving optimal problem with the objective of minimum total inventory shortfall and dispatching cost, the dispatching scheme and the incentives payout could be obtained.

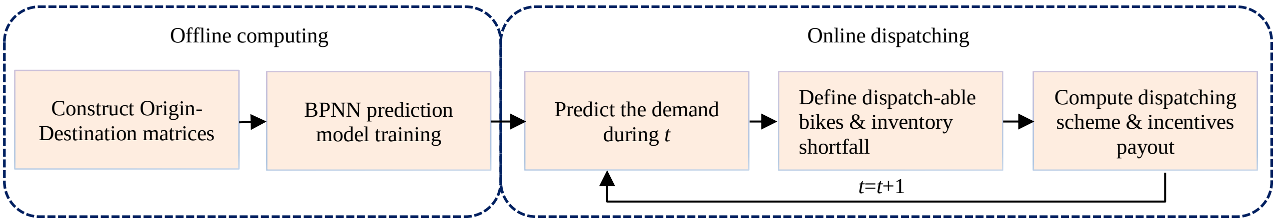

Fig. 1. System structure.

2. Data Analysis

2.1. *Historical data*

This research is based on the Mobike historical trip data, examples in Table 1, which includes 3,214,096 trips records of FFBS, during 2 weeks (May10th 2017 to May 24th 2017) in Beijing, see Mobike cup (2017).

For simplify and reduce the computing time, we selected those trips which on weekday and their origin and destination location are both within the 3rd Ring Road (area 159 km^2) in Beijing. Additionally, based on trip start time, each weekday equally divided into 72 time slices.

Table 1. Historical trip data samples.

Order id	Bike id	Start time	Start location	End location
1893973	210617	2017/5/14 22:16	40.1035,116.2896	40.1007,116.2868
4657992	465394	2017/5/14 22:16	39.7904,116.3253	39.7972,116.3226
2965085	310572	2017/5/14 22:16	39.8824,116.5423	39.8755,116.5519

2.2. *Data analysis of bike spatial and temporal distribution*

Firstly, the temporal analysis reveals the daily trip pattern as shown in Fig. 2. It can be seen that the usage amount fluctuates greatly with time, and the usage peak period are consistent with the morning and evening commuting peak. It means that many bikes used as a commuter tool. According to the report from Mobike (2017), 44% bikes are active around subway station as a last mile connection tool. We calculated the discrepancy in the volume of bikes between rented and returned around each subway station within 1 week based on GPS data, and try to figure out the imbalance pattern in spatial distribution. As shown in Fig. 3, some stations defined as “destination-type station”, which mean more bikes were returned around those stations; some stations defined as “Origin-type station”, which mean more bikes were rented around those stations. The discrepancy can reflect the severity of imbalance in bike spatial distribution. If there have no relocation scheme, the reliability of the entire system will become lower.

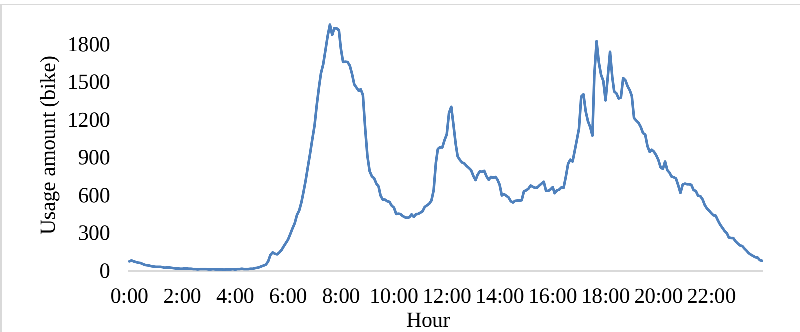

Fig. 2. Average usage rates on weekdays.

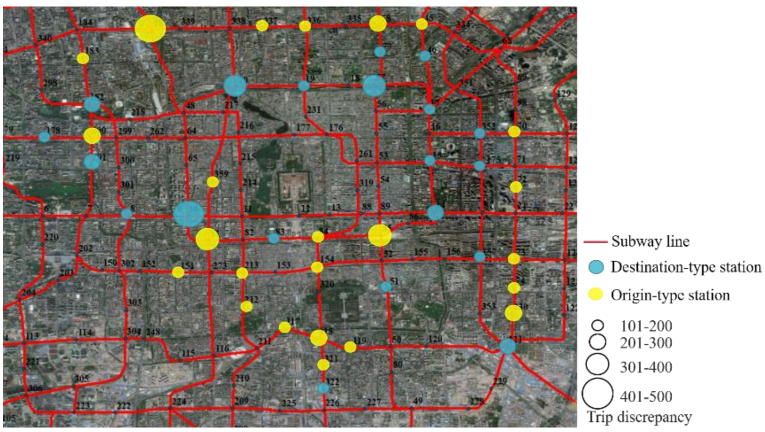

Fig. 3. Discrepancy between Origin trip and Destination trip around subway station.

3. Constructing transit O-D matrices with spatial clustering

3.1. *K-means-based clustering*

In this paper, clustering used to realize the partition management of FFBS. Firstly, with Hopkins Statistic, we can evaluate if the data is suitable for clustering. Hopkins Statistic, see Banerjee, A. and Dave, R. N. (2004), as an index to evaluate whether the data has a non-random structure, the value is located between [0, 1], 0.5 means that the data is evenly distributed, and a value close to 1 tends to indicate the data is highly clustered. The Hopkins Statistic of the data set used in this paper is 0.711, indicating that this data set is suitable for cluster analysis.

Vogel, P. et al. (2011) applied different clustering methods to the BSS to reveal the uneven distribution of bikes caused by user usage patterns. Luo, D. et al. (2017) used the K-means, objective of maximizing the ratio of average intra-cluster flow to average inter-cluster flow while maintaining the spatial compactness of all clusters; construct the Origin-Destination matrices base on data of the tram and bus of Haaglanden, Netherlands.

Due to the characteristics of free-floating, FFBS considered to have similarities with taxi in terms of clustering characteristics. Therefore, we also reference the clustering research of taxis. Lv, Y. Q. et al. (2010) used the GPS data from taxi to partition the research area by K-means. It found that K-means could reflect the homogeneity of trip.

According to previous research, follow the approach proposed by Luo, D. et al. (2017), K-means used in this research to construct transit O-D matrices.

Since K-means needs to input the number of clusters *k* so that the algorithm could begin the iterations. In this paper, the average radius of each cluster was employed to help select the initial range of *k* because it means the average dispatching distance of the relocation system. Referred to the station distance of station-based bike sharing system in real world, see ITDP (2018), 300meters as a walkable distance selected as the median value of the average radius, which result in 562 clusters. Subsequently, each *k*, ranging from 510 to 624 (denote the average radius from 285 meters to 315 meters) tested with following analysis to find the optimal number of clusters.

3.2. *Distance-based metric*

The construction of the distance-based metric adopted the Davies–Bouldin index (DBI) proposed by Davies, D. L., and Bouldin, D. W. (1979). As a function of the ratio of the intra-cluster scatter, to the inter-cluster separation, a lower value of DBI will mean that the clustering is better. The intra-cluster scatter, denoted by E_i:

$$E_i = \sqrt[2]{\frac{1}{T_i}\sum_{j=1}^{T_i}\left|X_j - A_i\right|^2} \tag{1}$$

where A_i is the centroid of cluster i and T_i is the size of the cluster i. X_j is a location (rent or return bike location) assigned to cluster i. E_i means the average Euclidean distance between the centroid of the cluster A_i, and each individual feature vectors X_j.

The inter-cluster measure, $M_{i,j}$, denotes the separation between cluster i and cluster j, it defined as follow:

$$M_{i,j} = \sqrt[2]{\sum_{z=1}^{2}\left|a_{k,i} - a_{k,j}\right|^2} \tag{2}$$

here $a_{k,i}$ is the z^th^ element of A_i. $M_{i,j}$ means the average Euclidean distance between centroid A_i and A_j.Then define $R_{i,j}$ as the similarity between cluster i and cluster j, it is used to define D_i, the similarity between cluster i with its most similar cluster.

$$R_{i,j} = \frac{E_i + E_j}{M_{i,j}} \tag{3}$$

$$D_i = \arg\max R_{i,j} \tag{4}$$

The Davies–Bouldin index, τ, defined as follows:

$$\tau = \frac{1}{k}\sum_{i=1}^{k} D_i \tag{5}$$

To obtain the optimal value of k, the idealized situation is that no cluster be similar to another, and hence τ should be minimized.

3.3. *Trip-based metric*

The Origin-Destination matrices should reflect the divergence of different area, and hence the trip-based metric can be used to determine the optimal value of *k*. The proportion of intra-cluster trips to total trips, δ is adopted and defined shown in function (6).

$$\delta = \frac{\sum_{i=1}^{k}\sum_{i=j} f_{i,j}}{\sum_{i=1}^{k}\sum_{j=1}^{k} f_{i,j}} \tag{6}$$

here $f_{i,j}$ denotes the number of trips from cluster *i* to cluster *j*. δ should be minimized so that more trip assigned as inter-cluster in the O-D matrix.

3.4. *Optimal number of clusters*

In order to obtain the optimal number of clusters, we consider both distance-based and trip-based metrics. Firstly, a standardizing procedure as follows applied to both metrics so that their magnitudes were comparable.

$$\tau_k^{'} = \frac{\tau_k - \tau_{\min}}{\tau_{\max} - \tau_{\min}} \tag{7a}$$

$$\delta_k^{'} = \frac{\delta_k - \delta_{\min}}{\delta_{\max} - \delta_{\min}} \tag{7b}$$

A simple method that takes the sum of δ_k' and τ_k' (denote the standardized trip-based and distance-based metrics for *k* as the clusters number, respectively) applied as follows so that the optimal number of clusters k^*could be determined.

$$k^* = \underset{k\in[k_{\min},k_{\max}]}{\arg\min}\ (\tau_k^{'} + \delta_k^{'}) \tag{8}$$

The variation of clustering result for different *k* is shown in Fig. 4. Fig. 4a reveals two standardized metrics for comparing. As expected, as the value of *k* increases, although there has fluctuating, the trend of trip-based metric is getting decrease. For distance-based metric, there is no clear specific patterns. Aggregated two metrics, the integrated result shown in Fig. 4b, the optimal number of clusters come to be 598. According to this result, the study area is divided into 598 clusters.

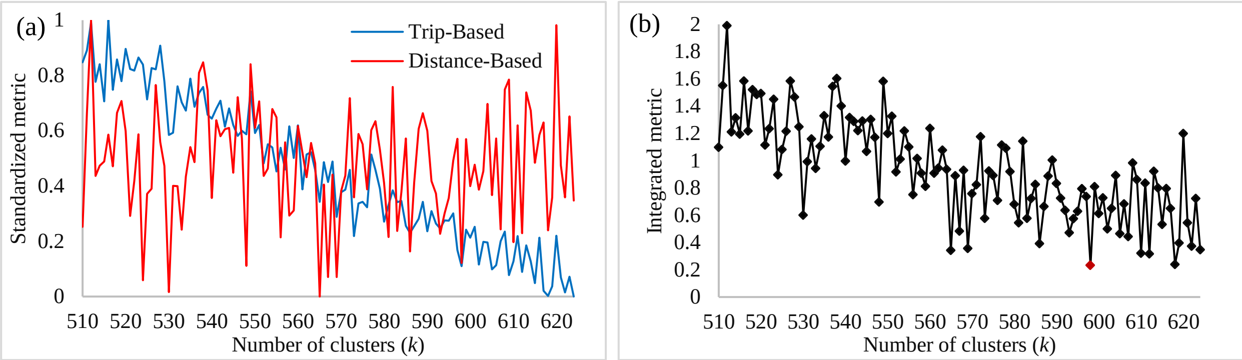

Fig. 4. (a) Standardized trip-based metric and distance-based metric for each *k*; (b) Integrated metric for each *k*.

4. Demand predictions

4.1. *Prediction method*

Due to the complexity of the FFBS and the non-linearity of usage pattern, there are fewer researches in the demand prediction of FFBS. Therefore, some researches about demand prediction in taxi and car sharing token as a reference in this paper and found that three common methods for short-term time series prediction include: Autoregressive integrated moving average (ARIMA), Kalman filter, and Neural Network.

ARIMA is a prediction method based on linear system theory. In Vlahogianni, E. I. et al. (2013), the bias errors of linear modes and nonlinear modes removed, a consistent evaluation framework established to compare the prediction accuracy of Neural Network and classical time series modes. It found that the optimized neural network has higher accuracy no matter univariate or multivariate.

The Kalman filter is also a linear prediction method. Li, Y. et al. (2017) pointed out that the Kalman filter is too simple to operate and the system complexity is not enough, for complex transportation systems, it cannot meet the requirements of prediction accuracy or offer a dynamic feedback.

In Guo, R. X. research (2017), back propagation neural network (BPNN) used to predict the shortfall of car sharing during different periods. It found that the BPNN can better reflect the autocorrelation and nonlinear pattern in the demand of car sharing, in the case of massive training data, the mean absolute percentage error of the prediction model is only 6.9%.

Based on the review above, this paper chose nonlinear mode, BPNN. The neural network model trained for each district with its historical Origin-Destination trip data so that the model could be used to predict the upcoming demand.

4.2. *Neural network training*

The back propagation neural network (BPNN) belongs to "supervised learning" and multi-layer network that uses error bike propagation for network training. BPNN consists of an input layer, hidden layers, and an output layer. There is no connection between the neurons in the same layer, but the neurons in different layers are fully connected.

In order to higher the prediction accuracy and make the training duration controllable, in this paper, each district trained separately for its prediction model.

For model training, firstly, it is necessary to determine the parameters of input layer. Based on the data processed above, the average correlation coefficient of bike usage amount between time t and its near time slice t-h in all district computed. The results shown in Fig. 5. It can be seen that as time goes by, the value does not decrease linearly as h increases, a sharp drop can be observed when the span is less than 160 minutes. Moreover, the average correlation value between different weekdays at a same time of day (*t*-72, *t*-144...*t*-360) analyzed, the results shown in Table 2. The high correlation between time *t* and the same time of previous week *t*-360 shows the periodic characteristics in weekly bike usage.

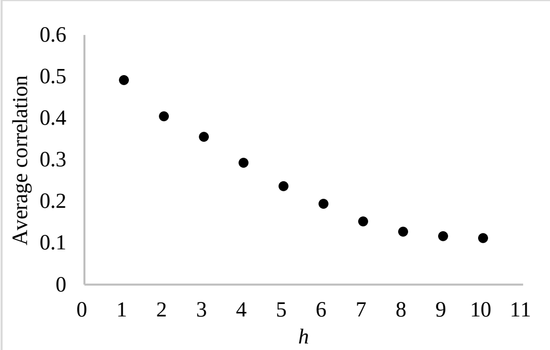

Fig. 5. Average correlation of bike usage amount between time slice *t* and near time slice *t-h*.

Table 2. Average correlation of bike usage amount at the same time of day.

Time slice	*t*-72	*t*-144	*t*-216	*t*-288	*t*-360
Average correlation	0.480	0.401	0.385	0.383	0.401

Take into account the model training time and the correlation above, , there are 12 parameters: *t*-1, *t*-2, *t*-3, *t*-4, *t*-5, *t*-6, *t*-7, *t*-72, *t*-144, *t*-360, the day of a week *td* ($td \in [1, 5]$), and the time of a day *ts* ($ts \in [1, 72]$) are selected as input layers.

In the determination of the neurons number of hidden layers, single hidden layer neural network selected for reducing training time. Follow the empirical equation (9) applied in Qiu, L. et al. (2018), the range for subsequent attempts is obtained.

$$M = \sqrt[2]{n+m} + a \tag{9}$$

Here, M denotes the range of the neurons number of hidden layer, n means the neurons number of input layer; m means the neurons number of output layer, a is a constant between [1, 10].

The historical data divided into training data set and testing data set, respectively. Then, the training data set used to train the model with different hidden layer neurons (within M). BPNN will follow the gradient descent rule. The error back propagation principle will feedback layer by layer and re-adjust the connection weight between different layers. This will be done repeatedly so that the connection weights are continuously adjusted until the predetermined convergence condition of the minimum error is satisfied. Then the testing data set brought into the obtained BPNN prediction model. Comparing the prediction result with testing data, the model with the lowest root mean square error (RMSE) selected as the optimal prediction model of that district. It will be used in future demand prediction.

5. User decision model

5.1. *Survey study*

A questionnaire designed to obtain respondents response to the incentive scheme. The questionnaire is divided into two parts: peak period (07:00-09:00 &17:00-20:00) questionnaire and off-peak period (time other than peak period) questionnaire. The problems in both two part are the same except for the time when the relocation occurs. The respondents were asked about their general riding behavior on frequency and purpose at first. Then an investigation on the user's willingness to participate in the dispatching conducted. Such: If you ride a FFBS from point A to point B on a weekday, when you returning the bike at your destination B, the bike application prompts: "If you are willing to ride the bike that you just used from point B to point C, you will get a bonus immediately." What are the longest distance and the lowest bonus that you will be willing to ride from B to C? Note that for the maximum incentives payout we set in the questionnaire, $price_{max}$, it set based on the average riding speed in Beijing (8.6 km/h), see Xinhua

(2017), we assumed participants will arrive at the dispatching destination within 5 minutes, then they will walk to their final destination within 10 minutes at half riding speed. With these assumptions and considering the minimum hourly wage in Beijing, see Beijing HR (2017), it is reasonably set that the $price_{max}$= 5CNY.

5.2. *User decision model*

There are 795 valid answers, the cumulative percentage of users who are willing to participate in the dispatching calculated under different dispatching distances and incentive prices shown in Fig. 6. Identical with expectation, people are more willing to participate in relocation during off-peak period when the incentives price and dispatching distance are the same. It is also worth noting that nearly 10% of users are unwilling to participate in the dispatching at any distance and incentive.

Along the lines of Pfrommer, J. et al. (2014), in order to simplify the solving of the subsequent optimal dispatching scheme, the linear regression is used to model the relationship between the user participate probability and the independent variables, such as the incentive price and the dispatching distance; therefore, equation (10) obtained.

$$Pr_{\text{peak}} = 0.164\,price - 0.321\,dist \qquad (R^2=0.755) \tag{10a}$$

$$Pr_{\text{off\_peak}} = 0.174\,price - 0.311\,dist \qquad (R^2=0.788) \tag{10b}$$

Where, Pr_{peak} and $Pr_{\text{off\_peak}}$ denotes the probability of users who will participate in the relocation during peak and off-peak period respectively; *price* means the incentives payout (CNY); *dist* denotes the dispatching distance (km).

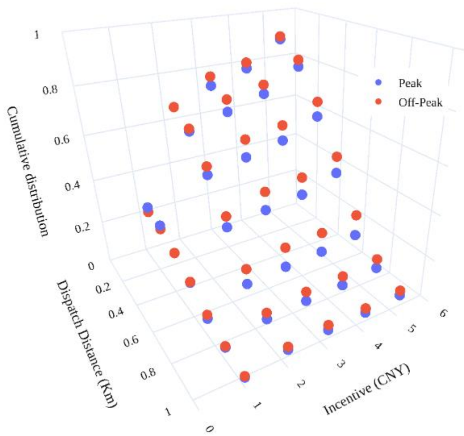

Fig. 6. User participation willingness under different dispatching distance and incentives price.

6. Optimized dispatch scheme

6.1. *Define the dispatchable bike and the inventory shortfall*

To dispatch bikes, first, we need to know when and where have dispatchable bikes and where they are needed.

In order to guarantee the reliability of FFBS system, $[\underline{In}_s, \overline{In}_s]$ defined as the range of inventory that district *s* should hold. The value of $\underline{In}_s$ and $\overline{In}_s$ are related to the population of district *s* and defined as follows.

$$\underline{In}_s = pop_s / \pi_{low} \tag{11a}$$

$$\overline{In}_s = pop_s / \pi_{high} \tag{11b}$$

Where, pop_s means the population of district s. π_{low} and π_{high} scale the lower and upper bound of bikes that per1,000 residents should hold, in this paper, 20% and 80% of average density used to determine the lower and upper bound respectively. Combined with the demand prediction model described above, R_s^{t+1} defined as the number of dispatch-able bikes in $t+1$ at district s. Q_s^{t+1} is denoted as the scale of inventory shortfall in $t+1$ at district s. The computing of these two parameters shown in Algorithm 1.

The algorithm first compares the $Inve_s^{t+1}$ which denotes the inventory of district s at upcoming t+1. If $Inve_s^{t+1}$ smaller than $\underline{In}_s$, the district s considered to be in "insufficient" state within $t+1$. If district s will become “insufficient”, we define the inventory shortfall Q_s^{t+1} equal to $\underline{In}_s$-$Inve_s^{t+1}$. When, $Inve_s^{t+1}$ lager than $\overline{In}_s$, those bikes arriving within t+1defined as available for dispatch.

Algorithm 1 Computing the R_s^{t+1}、Q_s^{t+1}	
Require: $Inve_s^t$	▷At the end of time t , the bike inventory of district s
O_s^{t+1}、D_s^{t+1}	▷Predicted volume of Origin or Destination trip for s in time t+1
[$\underline{In}_s, \overline{In}_s$]	▷The range of inventory that district s should hold
ts	▷The time slice within a day $ts \in [1, 72]$
Begin	
for (s in S)	
$Inve_s^{t+1} = Inve_s^t - O_s^{t+1} + D_s^{t+1}$	▷The inventory for s in the end of t+1 without relocation
if(ts >18 & ts <67)	▷System enable time 06:00-22:00
if($Inve_s^{t+1} > \underline{In}_s$)	▷If district s will become “insufficient” state at the end of t+1
if($Inve_s^{t+1} > \overline{In}_s$)	
R_s^{t+1}=$D_s^{t+1} - \max(0, \underline{In}_s - Inve_s^t)$	▷Redundant bikes for district s in t+1
Q_s^{t+1}= 0	▷The scale of inventory shortfall for district s in t+1
if($Inve_s^{t+1} \leq \overline{In}_s$)	
R_s^{t+1}=0	
Q_s^{t+1}=0	
if($Inve_s^{t+1} \leq \underline{In}_s$)	
R_s^{t+1}=0	
Q_s^{t+1}=$\underline{In}_s - Inve_s^{t+1}$	
end	
Output: R_s^{t+1}、Q_s^{t+1}	

6.2. *Computing the optimized relocation scheme*

In this subsection, we construct the relocation scheme. Based on the linearized user decision model, the number of bikes that dispatched from s to its neighbor n during t denoted as $x_{s\to n}^t$. Here we assume that all bikes dispatched at t will arrive at the relocation destination within t. $x_{s\to n}^t$ could be stated as follows.

$$x_{s\to n}^t = Pr_{s\to n} R_s^t \tag{12}$$

After relocation, the inventory shortfall, $\tilde{Q}_s^t$ obtained as follows:

$$\tilde{Q}_s^t = Q_s^t - \sum_n^{N_s} x_{n\to s}^t \tag{13}$$

Where N_s denotes the set of districts that having s as one of its neighbors.

Considered minimize the total inventory shortfall and dispatching cost, the programming problem can be stated as follows:

$$MinZ = \sum_{s}^{S}\sum_{n}^{N_s} price_{s\to n}^{t} \tag{14}$$

Such that

$$0 \le price_{s\to n}^{t} \le price_{\max} \tag{15a}$$

$$\sum_{n}^{N_s} Pr_{s\to n}^{t} \le 1 \tag{15b}$$

$$\sum_{n}^{N_s} x_{n\to s}^{t} \le Q_s^t \tag{15c}$$

$$\tilde{Q}_s^t \ge \lambda_t Q_s^t \tag{15d}$$

Equation (15a) ensures that the incentives price within [0, $price_{max}$]; (15b) limits no more than 100% users participate in the relocation; (15c) means the bike dispatched into district s no more than its inventory shortfall. Equation (15d) try to reduce the shortfall as possible, $\lambda_t \in [0,1]$ will decrement 0.01 in each iteration when there is no solution until a feasible solution is found or there have no bike are available for relocation.

7. Case study and simulation

7.1. *Case setting*

Due to the training time of BPNN, three different types regions are selected in this paper for case study. As shown in Fig. 7, each region includes 60 districts.

We randomly selected 80% of the historical O-D data from May 17th to May 23th to train the neural network models for each district in case region and utilize the remaining 20% data for accuracy testing. Then choose the model with the lowest RMSE to predict the O-D volume in May 24th. The comparison between the predicted result and the real state value as shown in Fig. 8. It can be seen that the disparity between prediction and real are comparatively small. The RMSE of origin and destination volume of each district are 0.755 and 0.738 respectively, which shows a high prediction accuracy. This prediction model could reflect the variation of the usage with time.

7.2. *Simulation results*

With considering the users schedule as well as morning and evening commuting peak period in Beijing, this relocation system designed to be initiated from 06:00 to 22:00 on weekday. Applying the proposed relocation scheme, this study simulated the system for 24 hours and compared the total inventory shortfall after relocation and before dispatching. The following indicator E used to evaluate the effect of proposed relocation system.

$$E = 1 - \frac{\sum_{t}^{T}\sum_{s}^{S}\tilde{Q}_s^t}{\sum_{t}^{T}\sum_{s}^{S}Q_s^t} \tag{16}$$

The relocation result, see Table. 3, shows that the above system is more effective when the shortfall is relatively low. The total payout is also economical than hire operator to do the relocation. For the different FFBS usage characteristics of each region, three different relocation patterns as shown in Fig. 9. The temporal distribution of dispatching and shortage are consistent, comprehensive area has relatively few shortfalls; for commercial area, the scarcity and dispatching peak are mainly concentrated in the morning commuting peak period; for the tourist attraction that peak comes later, most in noon and evening commuting peak. It is worth noting that when the shortfall is too large in a time slice (such as region B, a bustling commercial area), the system will have a relatively lower optimization effect due to the limited quantity of dispatch-able bikes that around the insufficient district.

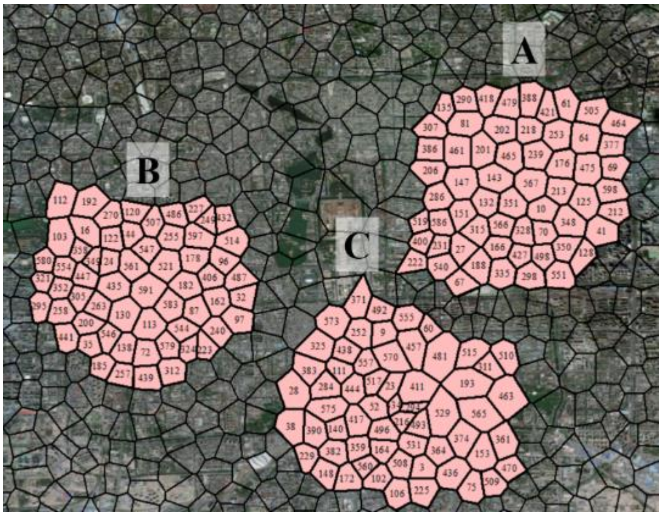

Fig. 7. Case region

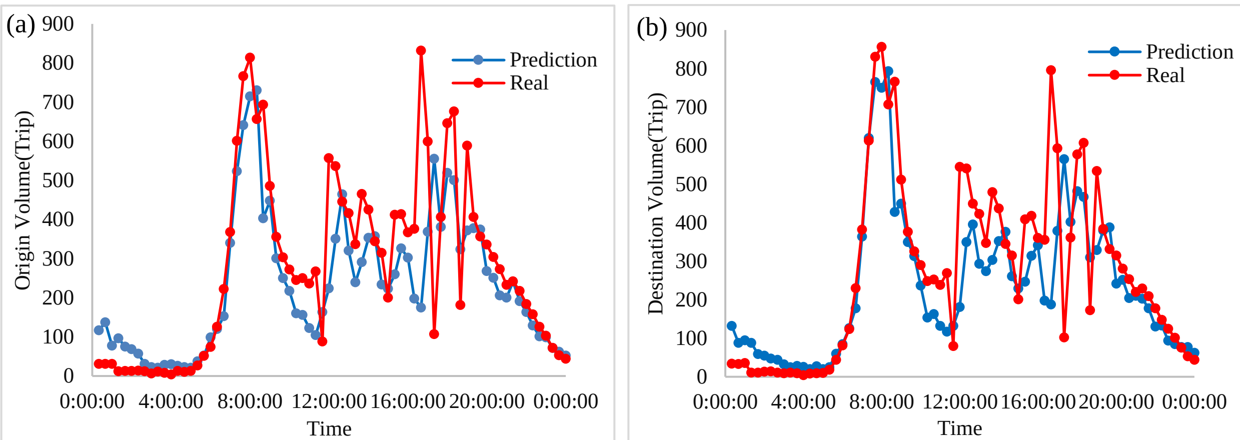

Fig. 8 The predicted and actual trip volume: (a) Origin (b) Destination.

Table 3. Relocation result for different study regions.

Region	Type	Total dispatching volume	Total shortfall	Effect	Total incentives payout
A	Comprehensive area	34	59	57.63%	54.96
B	Commercial area	261	638	40.91%	302.38
C	Tourist attraction	167	233	71.67%	242.18

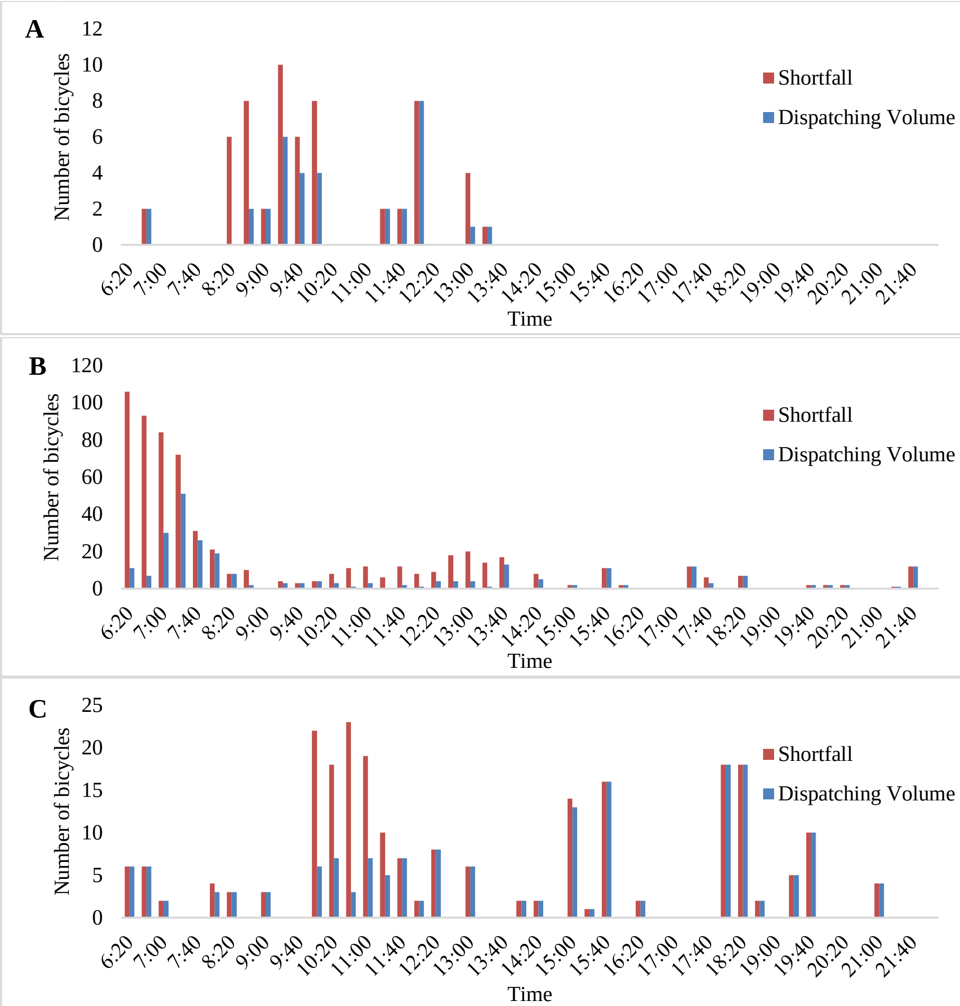

Fig. 9. The inventory shortfall and dispatching volume under different time in case study regions.

8. Conclusion

Ensuring a high reliability is essential for the free-floating bike sharing system. In this paper, a user-based relocation strategy for such system developed and evaluated. By spatial clustering, we constructed Origin-Destination matrices based on the historical trip data and thereby achieving partition management of the FFBS. The back propagation neural network used to predict the upcoming demand, simulation result shown high prediction accuracy. The case study in different type of region in Beijing shown that monetary incentives are efficacious for FFBS relocation under an affordable payout. However, due to the usage pattern variance in different region and the limitations of the quantity of dispatch-able bike, only part of the shortfall could be improved with exclusively user-based strategy.

Therefore, some further studies are worth noting here. Firstly, through machine learning or other artificial intelligence methods, the willingness of different users to participate in dispatching under different circumstances could be modeled more accurately. Secondly, dispatching based on short-term prediction has the defect of “short-sighted”, so it is recommended to utilize rolling optimization to get a more effective relocation scheme. Thirdly, a robust relocation system should combine both operator-based and user-based strategy for better optimization effect and relocation economy.

References

Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia: past, present, and future. Transportation Research Record, 2143(1), 159-167.

Fishman, E., Washington, S., Haworth, N., & Mazzei, A. (2014). Barriers to bikesharing: an analysis from Melbourne and Brisbane. Journal of Transport Geography, 41, 325-337.

Pfrommer, J., Warrington, J., Schildbach, G., & Morari, M. (2014). Dynamic vehicle redistribution and online price incentives in shared mobility systems. IEEE Transactions on Intelligent Transportation Systems, 15(4), 1567-1578.

Fricker, C., & Gast, N. (2016). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. Euro journal on transportation and logistics, 5(3), 261-291.

Singla, A., Santoni, M., Bartók, G., Mukerji, P., Meenen, M., & Krause, A. (2015). Incentivizing users for balancing bike sharing systems. In Twenty-Ninth AAAI Conference on Artificial Intelligence.

Reiss, S., & Bogenberger, K. (2017). A Relocation Strategy for Munich's Bike Sharing System: Combining an operator-based and a user-based Scheme. Transportation Research Procedia, 22, 105-114.

Papanikolaou, D. (2011). The market economy of trips (Doctoral dissertation, Massachusetts Institute of Technology).

Mobike cup. Online https://biendata.com/competition/mobike/data/. Accessed June 25, 2017.

Mobike. Riding Mobike can make money, "Mobike Bonus Car" help refined operation. https://mobike.com/cn/news/75. Accessed Mar. 23, 2017.

Banerjee, A., & Dave, R. N. (2004). Validating clusters using the Hopkins statistic. In 2004 IEEE International Conference on Fuzzy Systems, 1, 149-153.

Vogel, P., Greiser, T., & Mattfeld, D. C. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. Procedia-Social and Behavioral Sciences, 20, 514-523.

Luo, D., Cats, O., & van Lint, H. (2017). Constructing transit origin–destination matrices with spatial clustering. Transportation Research Record, 2652(1), 39-49.

Lv, Y. Q., Qin, Y., Jia, L. M., Dong, H. H., Jia, X. B., & Sun, Z. Y. (2010). Dynamic traffic zone partition based on cluster analysis of taxi gps data. Logistics technology, 29(9), 86-88.

ITDP. The Bikeshare Planning Guide. Institute for Transportation and Development Policy, 2018.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, (2), 224-227.

Petrovic, S. (2006). A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. In Proceedings of the 11th Nordic Workshop of Secure IT Systems, 53-64.

Gong, M., Fei, X., Wang, Z. H., & Qiu, Y. J. (2014). Sequential framework for short-term passenger flow prediction at bus stop. Transportation Research Record, 2417(1), 58-66.

Zhang, L., Ma, J., & Zhu, C. (2012). Theory modeling and application of an adaptive kalman filter for short-term traffic flow prediction. Journal of Information &Computational Science, 9(16), 5101-5109.

Guo, R. X. (2017) Short-Term Prediction on Demand of E-Hailing Trips Based on BP Neural Network. Beijing Jiaotong University, Beijing.

Vlahogianni, E. I., & Karlaftis, M. G. (2013). Testing and comparing neural network and statistical approaches for predicting transportation time series. Transportation Research Record, 2399(1), 9-22.

Li, Y., Lu, J., Zhang, L., & Zhao, Y. (2017). Taxi booking mobile app order demand prediction based on short-term traffic forecasting. Transportation Research Record, 2634(1), 57-68.

Qiu, L., Zhang, D., Huang, H., Xiong, Q., & Zhang, G. (2018). BP Neural Network Based Prediction of Potential Mikania micrantha Distribution in Guangzhou City. Forest Res, 7(216), 2.

Xinhua net. Online http://www.xinhuanet.com/itown/2017-04/26/c_136237422.htm. Accessed Apr 26, 2017.

Beijing HR . Online http://www.bjrbj.gov.cn/mtgz/mtgz_1714/201709/t20170920_67974.html. Accessed Sep 10, 2017.