Dynamic Scheduling Strategy Based on Demand Prediction of Shared Bike

Xiangkai Qiu[1](✉), Wenbing Yang[1], Haoyang Zhou[2], He Wang[2], and Shangjing Lin[2]

[1] Institute of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
dnull123p@gmail.com

[2] Institute of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China
{zhouhaoyang,linshangjing}@bupt.edu.cn

Abstract. This paper aims to address the issue of imbalanced supply and demand in shared bikes by proposing the concept of 'red packet bike', utilizing the open-source Beijing Shared Bike Dataset. The Temporal Graph Convolutional Network (T-GCN) is selected to predict the demand for shared bikes based on the analysis of spatio-temporal correlations in the order data. Additionally, a user incentive scheduling algorithm is designed using the breadth-first algorithm (BFS) and presented in the form of distributing the red packet bikes, thereby delegating the scheduling problem to the users to solve.

Keywords: Shared bike · Spatio-temporal analysis · User incentive scheduling

Foreword

Shared bike has become a popular mode of transportation for many people. However, it faces several challenges in daily life, such as low usage rates, imbalances in supply and demand, and increased costs for companies. These issues have led to a decrease in user satisfaction. Merely investing in more bikes will not solve these problems, as it can lead to road congestion and damage to existing bikes. A better solution is to schedule the bikes effectively. However, this is a challenging task, as the flow of shared bikes depends heavily on time and location. Finding the right scheduling method that meets users' needs is a difficult balance to achieve.

Most current strategies for managing shared bike systems rely on static rebalancing algorithms that schedule bikes at night to accommodate demand the next day. However, this approach has a delay and does not address the underlying issue of long-term imbalances between supply and demand for bicycles.

The use of shared bikes follows a certain pattern. Analyzing the relationship between bicycle usage and time and location reveals a strong correlation. This data can be treated as spatio-temporal data and used in deep learning algorithms to predict bicycle traffic.

Fund project: This work is supported by Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications funding under grant 202302018.

X. Chen et al. (Eds.): IoTaaS 2023, LNICST 585, pp. 425–435, 2025.
https://doi.org/10.1007/978-3-031-70507-6_32

By adjusting for influencing factors, accurate predictions can be made. Platform can then use these predictions in real-time to deploy a user incentive system, encouraging users to schedule rides independently. By offering certain rewards, such as red envelopes, the platform can address the issue of supply and demand imbalance for shared bikes.

1 Related Work

1.1 Estimated Demand and Parking for Shared Bike

In the past, studies about predicting demand and parking for shared bikes relied on statistical models. Zou et al. [1] used a time series model to predict travel issues for users within an hour based on their usage information. Kamarianakis et al. [2] created a nonlinear traffic flow model using the time series model to analyze real-time traffic flow. However, statistical models struggle with processing large amounts of data and are better suited for making small-scale predictions.

Most studies on predicting shared bikes demand and parking patterns are based on deep learning techniques. Jia Xianguang et al.'s [3] Bi-LSTM model resulted in low accuracy when predicting e-hailing car trajectories. Lin et al. [4] used GNN to predict the hour-level demand of a single station in the station-based public bike network and employed the long and short term memory neural network to capture the temporal correlation of demand sequence. While they talked about the temporal correlation of shared bike data, their discussion on spatial correlation was limited. Qian Xingjian's [5] MGCN-TCN model, on the other hand, considered both the temporal correlation and spatial relationship between shared bike stations, resulting in more accurate predictions. Additionally, Miao Xiaofeng et al. [6] accounted for the impact of weather on the data and used the LSTM model to train separately through various feature inputs. Their findings showed that the addition of spatio-temporal features improved prediction accuracy.

It can be seen that considering only the time correlation does not produce ideal prediction results. Inputting shared bike data as spatio-temporal data can effectively improve prediction accuracy.

1.2 Recommended Routes for Dispatch Vehicles

Currently, academic research on the scheduling of shared bikes primarily concentrates on addressing the Vehicle Routing Problem (VRP). Shi Bing et al. [7] have attempted to delegate scheduling tasks to users through incentive strategies and multiple algorithms to reduce costs. Liu Xinyu et al. [8] have utilized a genetic algorithm and several rounds of algorithm iteration to establish static scheduling for bikes. HaiTao Xu et al. [9] have implemented the K-means and random forest algorithm to forecast bike flow and further used the enhanced genetic algorithm (E-GA) for scheduling to obtain a more precise scheduling strategy. Zeng Qiongyan [10] has developed a scheduling model for shared bikes that aims to maximize total profit by using the simulated annealing algorithm. Kadri [11] transformed the problem into an optimization issue of site balance and scheduling, proposed algorithms like the greedy algorithm, and gave examples to prove its practicality. Caggiani et al. [12] have focused on free-floating systems that allow bikes to be scheduled anywhere and utilized dynamic redistribution to enhance user satisfaction.

Many algorithms have been developed to optimize shared bike systems, but most rely on static scheduling algorithms which cannot keep up with the needs of today's society. Therefore, a new dynamic algorithm is necessary to solve the problem in real-time.

2 Definition of Concepts

In predicting the bike flow data, this paper adopts the T-GCN algorithm to make accurate predictions. To streamline the process, the paper simplifies the data representation by rasterizing the latitude and longitude coordinate information from the bike dataset, see 4.1.
The following assumptions are made:

① Without introducing additional bikes,the total number of bikes in all grid areas is constant.
② Without introducing additional bikes,it is necessary to maximise the use of bikes parking in the area to meet the needs of the users.
③ After scheduling, if the number of bikes parked in a grid still does not satisfy the user's demand, additional bikes can be brought forward to that location, but the number of bikes brought forward must be minimal.

For the purpose of processing the predicted data, some variables are declared in Table 1.

Table 1. Used variable declaration

Variable	Meaning
$Out_{t,x,y}$	the number of bikes ridden out at the time period t with raster coordinates (x, y)
$In_{t,x,y}$	the number of bikes ridden in at the time period t with raster coordinates (x, y)
$Scheduled_{t,x,y}$	The total number of bikes artificially introduced or dispatched to a grid with grid coordinates (x, y) in the t time period
$ScheduledIn_{t,x,y}$	The number of bikes introduced to a grid with grid coordinates (x, y) in the t time period
$ScheduledOut_{t,x,y}$	The number of bikes dispatched to a grid with grid coordinates (x, y) in the t time period
$Store_{t,x,y}$	the number of bikes parked at the time period t with raster coordinates (x, y)

$Store_{0,x,y}$ is known, and as time flows, $Store_{t,x,y}$ is updated according to the Eq. (1):

$$Store_{t+1,x,y} = Store_{t,x,y} - Out_{t,x,y} + In_{t,x,y} \quad (1)$$

The goal of this paper is to compute, by means of a scheduling algorithm, the amount of $Scheduled_{t,x,y}$ required at each moment such that $Store_{t,x,y}$ satisfies the Eq. (2).

$$Store_{t+1,x,y} = Store_{t,x,y} - Out_{t,x,y} + In_{t,x,y} + Scheduled_{t,x,y} \tag{2}$$

The number of schedules in each raster at each moment in time is:

$$Scheduled_{t,x,y} = ScheduledIn_{t,x,y} - ScheduledOut_{t,x,y} \tag{3}$$

Besides, define the following variables:

Definition 1. ***DemandRate$_t$*** In order to improve the bike utilization rate, based on the assumption that the bikes used per unit of time will be derived as much as possible from the storage capacity in the area, the Eq. (4) are proposed.

$$DemandRate_t = \frac{\sum_{x,y} Min\left(Store_{t,x,y}, Out_{t,x,y}\right)}{\sum_{x,y} Out_{t,x,y}} \tag{4}$$

It indicates the proportion of bikes used that come from regional storage.

Definition 2 ***DemandDegree*** To issue red packets to users, the user's eligibility must first be determined based on scheduling conditions. The *DemandDegree* is then used to evaluate the user. It can be expressed as:

$$\begin{aligned} DemandDegree = & \max_{1 \le i \le 6} Out_{endtime+i, end_x, end_y} \times (7 - i) \\ & - \max_{1 \le i \le 4} Out_{starttime+i, start_x, start_y} \times (5 - i) \end{aligned} \tag{5}$$

DemandDegree is determined by two factors: the forecast demand for bicycles at the starting and ending locations, and the length of the forecast time from now. It represents how beneficial a user's bike route is to the scheduling strategy. The higher the *DemandDegree*, the more the user's cycling direction helps with bike scheduling and the more likely the user is to receive a red packet reward.

Definition 3 ***redPacket*** *redPacket* in this paper refers to the sum of red packets that users issue. To ensure normal scheduling and sufficient user participation, we inform users of nearby scheduling points when they scan the car. This encourages users to actively move to high-demand areas, helping to solve scheduling problems. The red packet amount can be adjusted based on the *DemandDegree*. Therefore, the red packet amounts are as follows:

$$redPacket = max\left(0, Riding\mathrm{Cost} - 1 + \frac{\ln(DemandDegree)}{6}\right) \tag{6}$$

3 Scheduling Frameworks

Red packet bike is by sending red packets to the user, inducing them to change the destination, so as to achieve the purpose of scheduling. On the definition of scheduling, that is, make $In_{t,x_1,y_1} - 1$, and make $In_{t,x_2,y_2} + 1$, and produce a certain cost of spending, which (x_1, y_1) is the user was intended to arrive at the grid, (x_2, y_2) is the scheduling of the user's actual arrival after the grid. The scheduling scheme can be expressed as a series of pairs of numbers such as $[(x_1, y_1), (x_2, y_2)]$. The scheduling algorithm needs to produce a scheme that minimizes the scheduling cost while making the DemandRate curve smooth at 1.

In this paper, red packet bike scheduling is divided into two solutions:

① Inducing the user to change the borrowing point: while recommending to the user the endpoints with a large red packet value.
② Inducing the user to change the point of return: when the user is ready to park the bike, he or she is guided to park the bike in a parking spot with a red packet.

The optimization objective is DemandRate. Let Δ DemandRate be the amount of DemandRate change before and after scheduling. Since redPacket ≥ 0 although Δ DemandRate can be positive or negative, it is necessary to ensure that each scheduling in the scheduling scheme can make Δ DemandRate > 0. It is easy to know that Δ DemandRate > 0 can only be made if a bike parked in a region with an overflow of bikes is moved to a region with an insufficient number of bikes. When a user intends to park the bike in an area with an overflow of bikes, the system initiates a search in the nearby grid areas to identify potential regions that require bike dispatching. The system then calculates the red packet value based on the nearest zone and sends this value as an incentive to the user. By doing so, the user is encouraged to change their bike's end position, thus actively participating in the bike redistribution process. This approach effectively induces users to contribute to the optimization of bike scheduling by moving bikes to areas where demand is higher, ultimately improving the overall bike-sharing system's efficiency.

4 Example Analyses

4.1 Dataset Introduction

This paper has chosen the Beijing shared bike data set from Kaggle platform, which has a vast amount of data (over 3.2 million) covering the period between May 10, 2017 and May 24, 2017. This data set records the order of shared bikes in Beijing, the capital of China, which is a super first-tier city. The selected data set of Beijing can effectively reflect some characteristics of shared bicycle orders, and it is also easy to analyze and process due to its large size. The data set includes fields such as orderid, userid, bikeid, biketype, starttime, start_loc, and end_loc, as shown in Table 2.

The starting and ending positions within this data are encoded using Geohash. Decoding these Geohashes yields the approximate latitude and longitude data.

To facilitate analysis and observation, the dataset is segmented according to time, resulting in 355 distinct hourly intervals over the course of the 14-day period. Location

Table 2. Labels of the data set

label	type	example
orderid	int	3454200
userid	int	81375
bikeid	int	356681
biketype	int	1
starttime	datetime	2017/5/14 22:17
geohashed_start_loc	String	wx4g15x
geohashed_end_loc	String	wx4g1ej

information is then extracted from the data. The OPTICS algorithm is applied to cluster the points and remove outliers, thereby revealing the primary usage areas for the bicycles.

The area under consideration was classified as urban and partitioned into a 50*50 raster grid to facilitate subsequent data analysis and processing.

4.2 Spatio-Temporal Characterisation of Demand for Shared Bike

Time Dimension Correlation Analysis

In order to analyze the data, one grid is randomly selected out of 2500 for visualization. This process obtains the weekly order situation of shared bikes, as shown in Fig. 1.

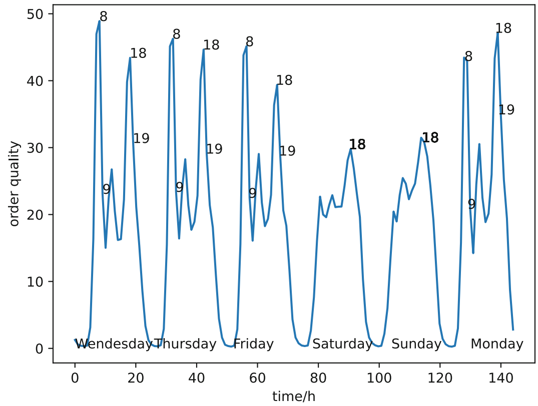

Fig. 1. One-week change in shared bike orders

Based on the time distribution, it is evident that the collected data exhibits a notable periodicity with a double hump trend, particularly at 8 am and 6 pm. However, it should be noted that the trend of the selected data is affected by rain interference on Saturdays and Sundays, resulting in a different pattern.

Spatial Dimension Correlation Analysis

During a certain period of time, the order data of all grids was randomly selected to create a thermal distribution diagram and perform spatial feature analysis, as seen in Fig. 2. The blue color represents bikes heading out while the red color represents bikes coming in. It is evident that the majority of orders were placed in the downtown area, while there were few orders in the outskirts. Additionally, it can be observed that the increase and decrease of bikes are adjacent and staggered, this means that bike flow is closely tied to the surrounding area.

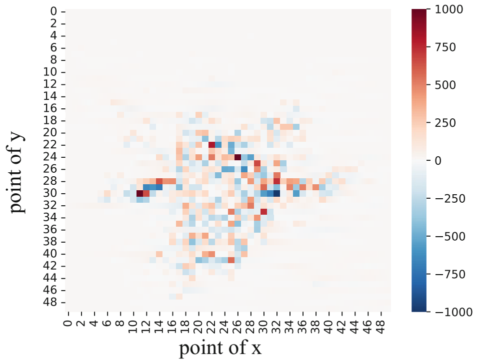

Fig. 2. Heat distribution of shared bike orders

T-GCN Model Predictions

We utilize the time series data in conjunction with the spatial adjacency matrix as our input data. Through the application of the Temporal Graph Convolutional Network (T-GCN) algorithm, we can derive more accurate predictions. After the appropriate.

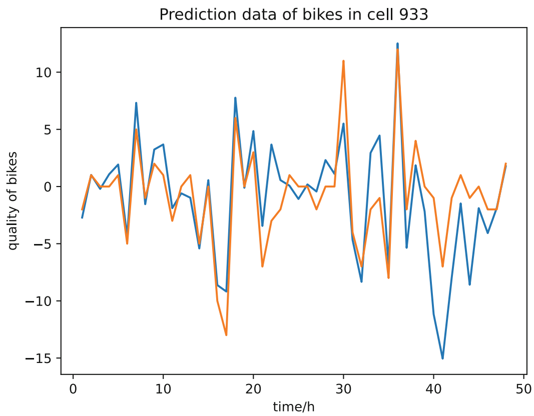

Fig. 3. Predicted results of changes in the number of bicycles in a given grid

preprocessing of the data, the resulting adjacency matrix and feature matrix are separately input into the model, producing the following predictive outcomes:

In Fig. 3, the orange curve represents the actual data while the blue curve depicts the predicted data. Observing the graph, we can deduce that the predicted data will also exhibit morning and evening demand peaks for one of the rasters in the selected area. The prediction results align reasonably well with the original data, suggesting a degree of reliability for both the chosen area and the predictions. The data predicted by this algorithm can be utilized in subsequent scheduling programs.

4.3 Red Packet Bike Scheduling Strategy Application

No Intervention Experiments

Calculate the image changes of the DemandRate without scheduling as follows (Fig. 4):

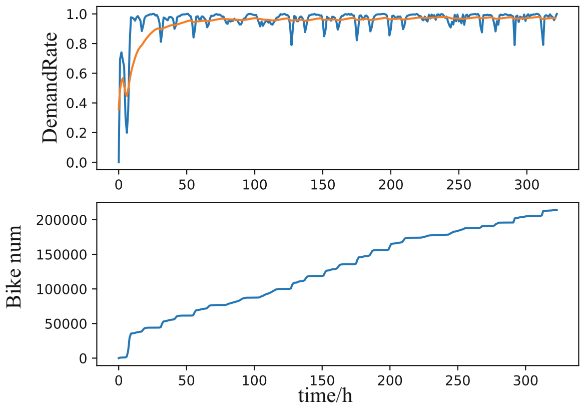

Fig. 4. Changes in DemandRate and number of bikes during natural migration

DemandRate intermittently plummeted to around 0.8, making it necessary to continuously introduce new bikes to meet user demand over this time period.

Simulation of Red Packet Bike Allocation Based on BFS

Programme Implemention

Considering the user's starting point as the center, and using demand and distance as measures, we recommend to the user both the starting and ending points where the value of the red packet can be maximized, as illustrated in Fig. 5.

In the Fig. 5, the green dot is the user's starting point, the darker the green means the more recommended this place is as the starting point, and the darker the red means the more recommended this place is as the end point. In this way, the user is provided with a variety of options to dispatch the bike and get the red packet, the user can view the red packet map when they want to use the shared bike and accordingly drive the car from the area with too many bikes to the area with too few bikes as much as possible. This allows the user to dispatch their own bikes.

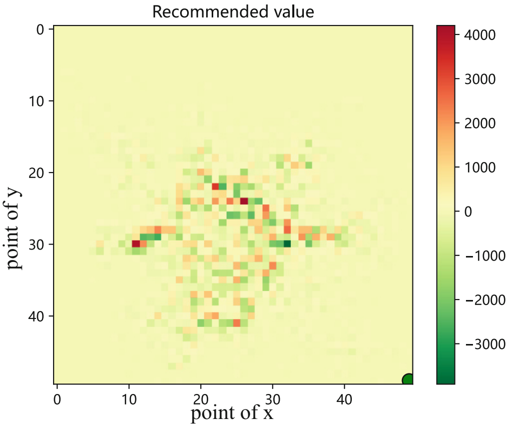

Fig. 5. Recommendation map of bike using by users of a certain grid

Results-Based Assessment

By employing the Manhattan distance as the distance metric for calculating the redPacket, , we obtain the following results after executing the above-described scheduling scheme (Figs. 6 and 7).

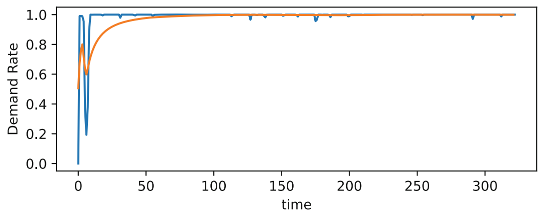

Fig. 6. *DemandRate* change curve after scheduling

After a certain period of time, the total demand for bikes decreases to approximately 40,000 and then stabilizes, all the while incurring continuous dispatch costs. Concurrently, the curve of DemandRate at 1 becomes smoother, resulting in a relatively high level of bike usage.

In practical applications, users can utilize the red packet car by receiving a discount on cycling or a direct red packet reward after successfully scheduling the bike. However, there are potential issues with this approach. For instance, some users may not want to use the red packet car method of scheduling and may choose unsuitable routes instead. Additionally, if the red packet is too large, some people may use it as a way to make money by riding the red packet car, which can be more expensive for the company than group scheduling of dispatch cars. If the red packet is too small, users may not find it worthwhile to use the red packet car. These practical concerns have yet to be addressed.

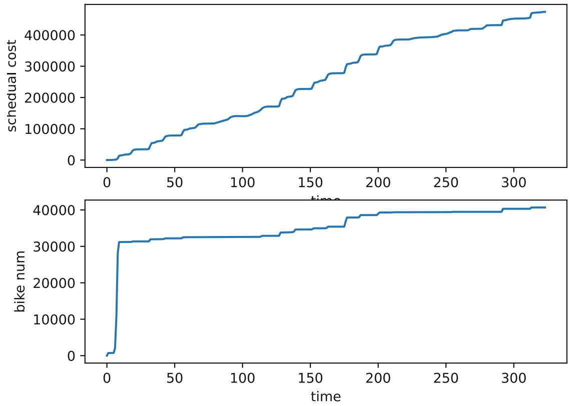

Fig. 7. Dispatch generation costs and changes in number of bikes

5 Conclusion

This paper focuses on the forecasting and scheduling of shared bicycles. The study uses a dataset from the Haidian District in Beijing, which is input as spatio-temporal data and analyzed using the T-GCN algorithm. The results are relatively accurate and effective. Additionally, the BFS algorithm is introduced to improve the utilization rate of bikes through the use of red packet cars for active scheduling of users. This helps to create a better environment for the use of shared bikes. However, there may be potential problems with this approach that require further investigation and analysis.

References

1. Zou, Y., Zhu, X., Zhang, Y., et al.: A space–time diurnal method for short-term freeway travel time prediction. Transp. Res. Part C Emerg. Technol. **43**, 33–49 (2014)
2. Kamarianakis, Y., Shen, W., Wynter, L.: Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. Appl. Stoch. Model. Bus. Ind. **28**(4), 297–315 (2012)
3. Jia, X., Feng, C., Su, Z., et al.: Traffic prediction by Bi-LSTM for urban traffic grid clustering. J. Chongqing Univ., 1–14 (2023). http://kns.cnki.net/kcms/detail/50.1044.N.20230706.0952.002.html
4. Lin, L., He, Z., Peeta, S.: Predicting station-level hourly demand in a large-scale bikesharing network: a graph convolutional neural network approach. Transp. Res. Part C Emerg. Technol. **97**, 258–276 (2018)
5. Xingjian, Q.: Research on demand prediction of shared bicycle based on spatio-temporal dependence. East China Normal Univ. (2021). https://doi.org/10.27149/d.cnki.ghdsu.2021.000710
6. MIAO Xiaofeng,FAN Shurui,CAO Dandan et al. Demand prediction of shared bicycle stations based on LSTM and spatio-temporal combination[J]. Journal of Inner Mongolia University of Technology(Natural Science Edition), 2020,39(03):24–31.https://doi.org/10.13785/j.cnki.nmggydxxbzrkxb.2020.03.004
7. Shi, B., Huang, X., Song, Z., Xu, J.: User incentive based bike-sharing dispatching strategy. J. Comput. Appl. **42**(11), 3395–3403 (2022)

8. Xin-yu, L.I.U., Qun, C.H.E.N.: An optimization model for bike repositioning in bike-sharing systems considering both demands for borrowing or returning bikes and costs of repositioning operations. China J. Highw. Transp. **32**(7), 146–157 (2019)
9. Xu, H., Duan, F., Pu, P.: Dynamic bicycle scheduling problem based on short-term demand prediction. Appl. Intell. **49**, 1968–1981 (2019). https://doi.org/10.1007/s10489-018-1360-6
10. Zen, Q., Yang, S.: Research on dynamic scheduling of shared bicycles based on simulated annealing algorithm. J. Compr. Transp. **45**(02), 75–79 (2023)
11. Kadri, A.A.: Simulation and optimization models for scheduling and balancing the public bicycle-sharing systems. Other [cs.OH]. Université de Lorraine (2015). English. ⟨NNT : 2015LORR0268⟩. ⟨tel-01754554⟩
12. Caggiani, L., Camporeale, R., Ottomanelli, M., Szeto, W.Y.: A modeling framework for the dynamic management of free-floating bike-sharing systems. Transp. Res. Part C Emerg. Technol. **87**, 159–182 (2018). https://doi.org/10.1016/j.trc.2018.01.001