Study on Demand Forecasting and Scheduling Routes of Shared Bicycles

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

[1] Institute of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China
2206574108@qq.com, {zhouhaoyang,linshangjing}@bupt.edu.cn
[2] Institute of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China

Abstract. This paper endeavors to address the pressing issue of resource wastage in shared bicycles by proposing an innovative approach to optimize their utilization and cater to the demands of urban residents. The proposed solution involves devising an efficient vehicle dispatch roadmap based on predictive demand modeling. Leveraging the open-source Beijing shared bicycle dataset, the research analyzes the spatio-temporal correlations within order data. The Temporal Graph Convolutional Network (T-GCN) is selected as the predictive model to forecast shared bicycle demand. Subsequently, the Genetic Algorithm (GA) is employed to determine an optimal dispatch route, thereby significantly improving the overall utilization rate of shared bicycles.

Keywords: shared bicycle · spatio-temporal prediction · centralized dispatching

Foreword

In recent years, shared bicycles have become popular. However, there are several problems associated with them. The most significant problem that affects the user experience is the imbalance between supply and demand, leading to low turnover. While scheduling shared bicycles can alleviate this issue, there may still be instances where users have to wait for a long time due to scheduling delays. Therefore, the key to addressing this problem is to effectively use the scheduling method.

Many current algorithms rely on static scheduling methods, which schedule bicycles overnight based on the usage from the previous day in order to meet the needs of users for the following day. However, this method requires data from the previous day and since the flow of bicycles can vary from day to day, the scheduling outcome may not be optimal and there may still be a shortage of bicycles.

The flow of shared bicycles has a certain regularity in time and space. For instance, when it comes to time, the need for bicycles on weekdays remains consistent across most areas. In terms of location, areas with a higher demand for shared bicycles tend to

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. 414–424, 2025.
https://doi.org/10.1007/978-3-031-70507-6_31

have specific amenities such as homes, subway stations, and entertainment areas. While it is not possible to directly compare data from the first day to the second day, a deep learning algorithm can be utilized to make predictions. This algorithm takes into account the spatio-temporal characteristics of the data, resulting in more accurate predictions. These predictions are then fed into a scheduling algorithm, which produces a scheduling map. This map can be used to schedule cars in a way that meets the needs of users during various time periods.

At present, most of the solutions for traffic prediction are based on neural networks. Songjiang Li et al. [1] proposed an accurate multi-task learning recurrent neural network model (MTL-RNN) for short-term traffic prediction on highway networks, but it is limited to short-term changes. Shuyun Qian [2] integrated multi-graph convolution and attention mechanisms into the Seq2Seq framework for long-time traffic prediction, yet disregarded the spatial correlation of data. Jiansong Liu [3] introduced the Spatio-Temporal Graph-CoordAttention network (STGCA), which considers both temporal and spatial dimensions, but its prediction range is small, limiting its applicability on a large scale. Qiwen Yang [4] developed an Adaptive Parallel Structured Neural Network (APSNN) to address consistency problems in neural network training, but it struggles to adapt to the number of hidden nodes, affecting the neural network structure. Ming Wang [5] proposed a multi-timescale spatio-temporal graph network model, but it lacks stability.

Outside the realm of neural networks, Yusen Chen [6] designed a dynamic prediction system focused on highway networks, but its migration to shared bicycle prediction poses challenges.

In the realm of scheduling algorithms, the main focus has been on static scheduling. Hui Liu et al. [7] employed a modified hybrid taboo-particle swarm algorithm to address the problem. Zhiyong Zhang et al. [8] proposed an ant colony algorithm-based approach to optimize the model. Jiantong Zhang et al. [9] explored the use of Variable Neighborhood Search (VNS) algorithm to find local optimal solutions. Jun Zhou et al. [10] presented an equilibrium strategy for integrated bicycle scheduling around subway stations using multi-model technology, combining the home range method of animals in ecology and mean-shift clustering analysis. Ruochen Kong [11] opted for the genetic algorithm to satisfy the demand for bicycles on campus. Zichun Hu [12] performed scheduling area division through K-means clustering analysis and further improved the genetic algorithm to handle peak period scheduling.

All the above algorithms can schedule shared bicycles, but their scheduling is not predicted in advance, and has a large lag, which can hardly meet the needs of peak users.

To address the aforementioned issues, this paper presents strategies to tackle shared bicycle demand prediction and scheduling challenges. The aim of this paper is to maximize bicycle utilization, enhance user experience, and minimize operational costs for enterprises.

1 Spatial and Temporal Characterization of Demand for Shared Bicycles

In this paper, we utilize the 2017 bike-sharing dataset of Haidian District, Beijing for analysis. The dataset comprises bike-sharing order data spanning 14 days, from 27th June to 10th July. The data information includes various attributes such as orderid, userid, bikeid, biketype, starttime, start_loc, and end_loc, as displayed in Table 1.

Table 1. Labels of the data set

label	data type	typology
orderid	int	1893973
userid	int	451147
bikeid	int	210617
biketype	int	1
starttime	datetime	2017/5/20 23:35
geohashed_start_loc	char	wx4snhx
geohashed_end_loc	char	wx4s nhj

The start and end positions are encrypted by Geohash and decoded to get the approximate latitude and longitude ranges.

The dataset was organized based on time, and the 14-day data was segmented into 355 hourly time periods. Subsequently, the location information from the data was extracted and subjected to clustering using the OPTICS algorithm to eliminate outliers. As a result, the primary bicycle usage area, representing the urban region, was identified. This area was further divided into grids of size 50 * 50 for subsequent data analysis and processing.

1.1 Time Dimension Correlation Analysis

Figure 1 illustrates the weekly change in shared bicycle orders within a grid. It shows that shared bicycle orders have a strong temporal correlation. The trend and number of daily order changes are nearly identical. Although there is a deviation in quantity on Saturdays and Sundays, the trend remains roughly the same.

1.2 Spatial Dimension Correlation Analysis

In order to analyze the spatial dimensions, the spatial distribution of all grid data at a specific time is shown in Fig. 2. Most orders are concentrated in the downtown area with few in the outskirts. At the same time, a small variation in order volume was noted in neighboring grids, indicating a strong correlation between bike flow and the surrounding environment.

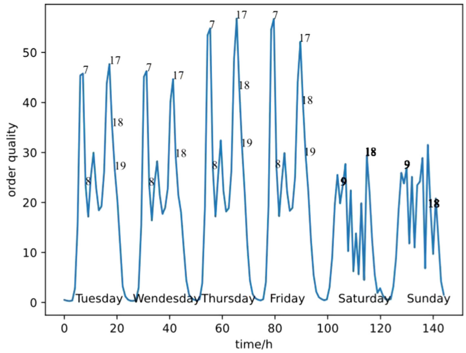

Fig. 1. The weekly change in shared bicycle orders within a grid

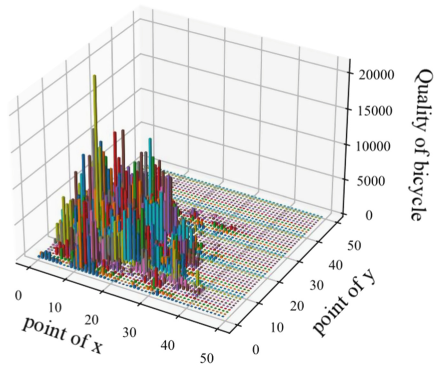

Fig. 2. Demand for bicycles in the whole grid at a specific time

2 Algorithm Design

In this paper, the following assumptions were made.

① The sum of the number of bicycles in all raster areas is constant without introducing additional bicycles.
② It is necessary to utilize the bicycles parked in the area as much as possible to meet the needs of users without introducing additional bicycles.
③ After scheduling, if the number of bicycles parked in a certain grid still cannot meet the demand of users, additional bicycles can be introduced in advance at this location, but the number of bicycles introduced must be minimal.

In predicting bicycle mobility, this paper utilizes the T-GCN algorithm which treats input data as spatio-temporal data resulting in more precise prediction outcomes. The T-GCN algorithm takes into account both spatial and temporal aspects of the data, making it very effective in boosting the accuracy of predictions.

To process the predicted data, the following variables are defined.

$\mathbf{Store_{t,x,y}}$: At the beginning of the time period t, the number of bicycles parked at raster coordinates (x, y).

$Out_{t,x,y}$: During the time periodt, the number of bicycles ridden out from the raster coordinates(x, y).

$In_{t,x,y}$: During the time periodt, the number of bicycles ridden in to the taster coordinates(x, y).

Initially $Store_{0,x,y}$ is known data, and with the flow of time, the $Store_{t,x,y}$ is updated according to the Eq. (1).

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

Therefore, the goal of this paper is to calculate the amount of scheduling needed $Scheduled_{t,x,y}$ at each moment by scheduling algorithm to make sure that $Store_{t+1,x,y}$ satisfies the Eq. (2).

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

Therefore, the $Store_{t+1,x,y}$ can be stabilized after scheduling.

2.1 Objective Function

Based on the above assumptions, it is necessary to make the bicycles used in each time period come from the storage capacity of the region as much as possible, so the following indicator is proposed in Eq. (3).

$$DemandRate_t = \frac{\sum_{x,y} Min\left(Store_{t,x,y}, Out_{t,x,y}\right)}{\sum_{x,y} Out_{t,x,y}} \tag{3}$$

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

2.2 Algorithmic Ideas

Genetic Algorithm (GA) originated from the computer simulation research on biological systems, is a stochastic global search optimization method, which simulates the phenomena of replication, crossover and mutation occurring in natural selection and heredity, starting from any initial population, through random selection, crossover and mutation operations, to produce a group of individuals better suited to the environment, making the group evolve to better and better regions in the search space, so that generation after generation, it keeps reproducing and evolving, and finally converges to a group of individuals that are best adapted to the environment, so as to find a high-quality solution to the problem.

Specify the transportation path of the transporter as a series of sequential rasters, for example:

$$\text{Routine} = \left(x_1, y_1\right)\left(x_2, y_2\right)\ldots\ldots\left(x_n, y_n\right) \tag{4}$$

It can be assumed that the number of bicycles in reserve on the grid will match the demand for the next moment once the transportation vehicles have completed their scheduling and passed through each point. All bicycles on the route will be distributed based on the proportion of demand for each grid to the total demand. The number of bicycles on each grid will be balanced according to the Eq. (5). The total number of bicycles on the grids will be added up and then allocated accordingly.

$$\mathrm{Store}_{t+1,x,y} = \frac{\sum_{i=1}^{\mathrm{Routine}} \mathrm{Store}_{t,x,y}}{n} \tag{5}$$

Now it is necessary to consider how to choose a suitable route so that at the end of the time period t, the reserves of bicycles can meet the demand of the next time period (so that the DemandRate of bicycles in the area of the next time period is as large as possible), which needs to have predicted the demand distribution of the next time period. In order to select the optimal scheduling points and routes, this paper needs to use genetic algorithm optimization twice.

Each raster is coded and expressed as an integer, and this integer is used as a gene on the chromosome as shown in the equation.

$$\mathrm{Gene}_i = y_i \times 50 + x_i \tag{6}$$

The selected path is then represented as a chromosome as:

$$\mathrm{Routine}_i = \mathrm{Chorm}_i = (\mathrm{Gene}_1, \mathrm{Gene}_2 \ldots\ldots \mathrm{Gene}_n) \tag{7}$$

Next set up the crossover and mutation operations in the genetic algorithm.

Crossover: Select a series of position genes from the chromosome, and exchange these genes in the two chromosomes, which is equivalent to exchanging the selection points on the path, so as to obtain two new scheduling paths. The specific operation is as follows: First, a random sequence of 01 with a length equal to the number of selected chromosome genes, such as 00100, is generated, and then the two selected crossover position genes at the corresponding position (that is, the gene with the number of 1 in the corresponding random sequence) are numerically exchanged, as shown in Fig. 3.

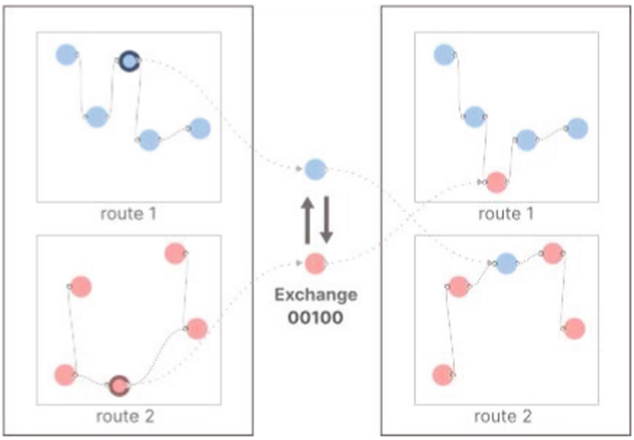

Fig. 3. Schematic diagram of cross operation

Mutation: To introduce mutation in the chromosome, a fixed probability p_i is assigned to each gene selected as a target. For a chromosome of length n, approximately $\sum_{i=0}^{n} p_i$ genes will be mutated. These selected genes are randomly displaced within the threshold value to obtain the path after mutation, as shown in Fig. 4.

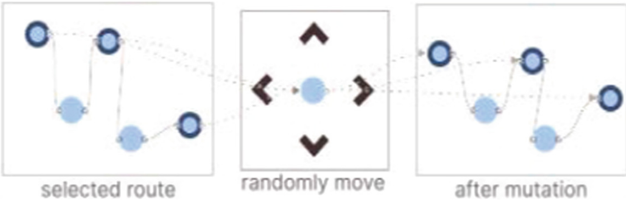

Fig. 4. Schematic diagram of mutation operation

Finally, the objective function for optimization is set, which calculates DemandRate after scheduling. This data employs a function to compute DemandRate after scheduling each grid based on the scheme.

3 Experiment

Based on the existing data, the prediction of the change in the number of bicycles is obtained by forecasting their demand and further using genetic algorithms, scheduling routes are given to improve the utilization of bicycles.

3.1 Demand Projections

Considering the number of shared bicycle orders as a set of time series data, after the corresponding preprocessing of the data, the obtained adjacency matrix and feature matrix are regarded as spatial data input into the T-GCN model, and the following prediction results can be obtained:

In the Fig. 5, the orange curve is the actual data and the blue is the forecast data. The forecast curve basically coincides with the actual curve, and the trend of the forecast curve still matches the actual in the time period where there is a large difference. Therefore, the prediction results have strong accuracy. In the subsequent scheduling program, the above data prediction results are utilized for scheduling.

3.2 Experimental Steps

No intervention Experiments

The changes in DemandRate without introducing bicycles are shown in the Fig. 6. As can be seen from the graph, DemandRate intermittently drops to around 0.8, indicating the continuous need to introduce new bicycles during this time period to meet the users' demand for bicycles.

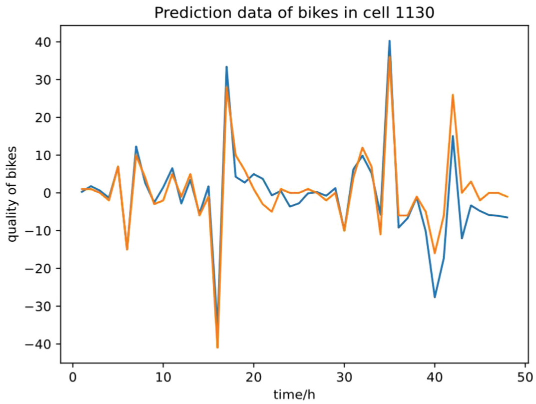

Fig. 5. Comparison of predicted and actual results of changes in the number of bicycles in a grid

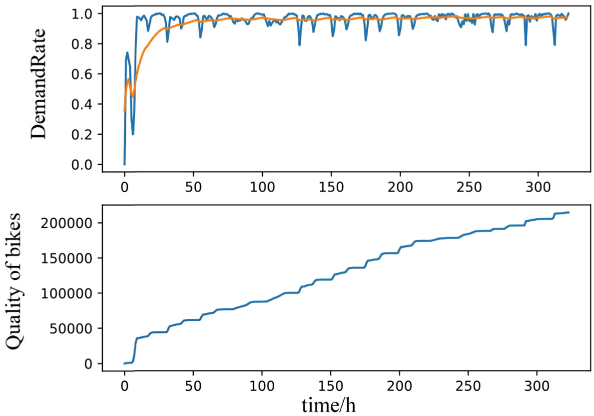

Fig. 6. Changes in the rate of demand for bicycles and changes in the number of bicycles under unmanned intervention

Demand Scheduling of Shared Bicycles Based on Genetic Algorithm

In order to accelerate the convergence of the genetic algorithm, the greedy algorithm was used once first, and the rasters with more remaining bicycles and more lack of bicycles were selected as the genes of the initial population, which enabled the algorithmic process to enter into a better population state in advance.

After the optimization is completed, the genetic algorithm is used to calculate the model of the TSP problem and the shortest path to the optimal scheduling point. This process also uses the greedy algorithm to calculate the better initial population, thus accelerating the convergence of the genetic algorithm.

After two optimization choices of genetic algorithm, we can get the scheduling route map of the dispatching vehicle in each time period (Figs. 7 and 8).

After receiving the scheduling plan, the dispatching vehicle can real-time schedule the available bicycles to the grids with a shortage of bicycles, thus accomplishing the scheduling task efficiently.

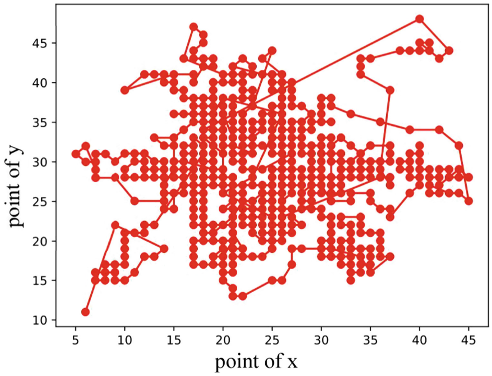

Fig. 7. Peak Dispatch Road map

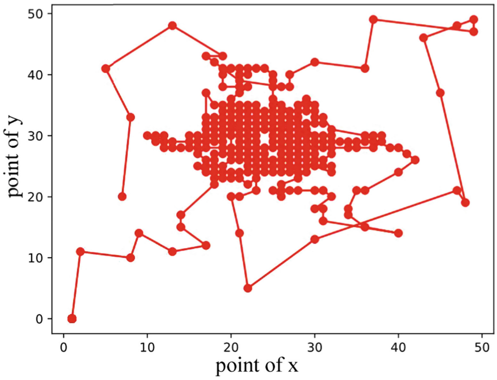

Fig. 8. Low-peak dispatch road map

The following results are obtained after implementing the scheduling scheme described above (Figs. 9 and 10):

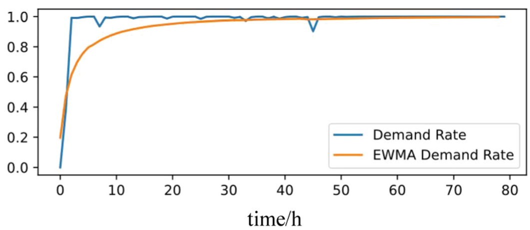

Fig. 9. Demand curve after genetic algorithm scheduling

Over time, the DemandRate gradually converges to 1 and does not suddenly decrease significantly most of the time, stabilizing the bicycle utilization rate at a high level. At the same time, the number of bicycles required drops to about 150,000 and approaches

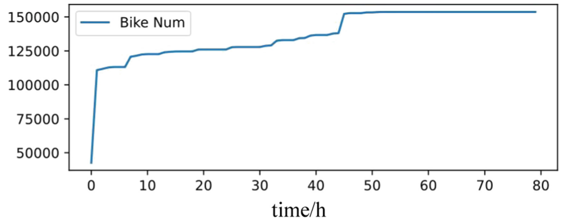

Fig. 10. Change in the number of bicycles after genetic algorithm scheduling

equilibrium in a short period of time, thus eliminating the need to continuously introduce new bicycles and achieving efficient bicycle utilization.

4 Conclusion

This paper examines the correlation between shared bicycle orders and time and location through spatio-temporal analysis to better cater to the needs of urban shared bicycles. The T-GCN algorithm is utilized to predict bicycle flow and generate prediction results. To improve the utilization and environment of shared bicycles, dispatching vehicles are introduced using a genetic algorithm to dispatch bicycles according to a designated route.

References

1. Songjiang, L., Shaojing, Z., Huamin, Y., et al.: Traffic prediction based on spatio-temporal correlation multitask neural network. Comput. Appl. Softw. **38**(09), 286–292 (2021)
2. Qian, S.: Research on long-time traffic prediction based on multi-graph convolution and attention. Mod. Comput. **29**(03), 34–38 (2023)
3. Liu, J.S., Kang, Y., Li, H., et al.: A spatio-temporal Graph-CoordAttention network for traffic flow prediction. Comput. Sci. **50**(S1), 568–574 (2023)
4. Qi-Wen, Y.A.N.G., Yue, L.I., Jun-Na, W.U., et al.: Traffic flow prediction based on adaptive parallel structure neural network. Comput. Meas. Control **31**(04), 42–48 (2023). https://doi.org/10.16526/j.cnki.11-4762/tp.2023.04.007
5. Wang, M., Peng, S., Huang, F.: A traffic flow prediction model based on multi-timescale spatio-temporal graph networks. Comput. Sci. **49**(08), 40–48 (2022)
6. Yusen, C.: Dynamic traffic prediction of highway networks. Highw. Transp. Sci. Technol. **04**, 1–8 (1995)
7. Liu, H., Zhong, J.: A particle swarm algorithm-based optimization method for interstation scheduling of shared bikes. J. Xichang Coll. (Nat. Sci. Ed.) **33**(02), 67–69 (2019). https://doi.org/10.16104/j.issn.1673-1891.2019.02.015
8. Zhang, Z.-Y., Zhang, X.: Shared bikes scheduling under users' travel uncertainty. IEEE Access **8**, 3123–3143 (2020). https://doi.org/10.1109/ACCESS.2019.2961628
9. Zhang, J., Dai, Q., Ding, Y.: Optimization study of shared bicycle scheduling considering demand splittable. Shanghai Manag. Sci. **45**(01), 119–125 (2023)
10. Zhou, J., Zhou, Q., Lin, Y.: The active home domain of shared bicycles around subway stations and its scheduling strategy. In: Academic Committee on Urban Transportation Planning, China Society of Urban Planning. Green-Wisdom-Integration-Proceedings of the 2021/2022 Annual Conference on Urban Transportation Planning in China. [Publisher unknown], pp. 19 (2022). https://doi.org/10.26914/c.cnkihy.2022.059197

11. Kong, R., He, S., Sheng, Y., et al.: A genetic algorithm-based scheduling method for shared bicycles on university campuses. Transp. Sci. Technol. **311**(02), 132–135 (2022)
12. Huizhi, C.: Optimization research on peak period shared bicycle scheduling based on improved genetic algorithm. Nanjing University of Finance and Economics (2021). https://doi.org/10.27705/d.cnki.gnjcj.2021.000298