Unbalance Adjustment of Shared Bicycle Station Based on Data Analysis

Meng Gao and Yanyan Tan
School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China
Email: gaomeng_sdnu@outlook.com

Abstract. Because the number of shared bicycles riding out of the shared bicycle station and the number of shared bicycles returned by users on this station are not balanced, the problem that users have no shared bicycle riding and no redundant positions within the station for the return of the shared bike with the user. In order to solve the problem, this paper puts forward the imbalance adjustment of the unbalanced stations according to the distance, the imbalance ratio and the unbalanced sharing of the quantity of bicycles. The quad-tree algorithm is used to find out the highest frequency of riding a shared bicycle. According to the formula to determine whether the station is imbalance, if the station is imbalance, then classify it, and calculate the imbalance ratio. The unbalanced stations are adjusted according to the distance between stations, the imbalance ratio, and the quantity of unbalanced shared bicycles. By using this method, the number of shared bicycles in the station is too small or the problem of vehicle accumulation can be solved. This paper demonstrates the feasibility of this method by conducting experiments on shared bicycle data in the City of Chicago.

1. Introduction
At present, shared bicycles are popular with people because of their moderate riding prices, convenient return, low-carbon environmental protection, and fitness advantages [1]. This convenient transportation has been deployed in most cities and plays an important role in people’s daily lives. By analysing the data of shared bicycles, we can get a population's mobility in a certain city. The more frequent use of shared bicycles, the greater the flow of population. In contrast, when the number of shared bicycles is small, the population flow is relatively small. Regarding the situation reflected in the data, we can take corresponding measures to solve the existing problems in many cities to some extent [3]. Taking the Chicago of the United States as an example, the large amount of private car trips will inevitably cause a lot of social problems such as exhaust pollution and traffic jams. It is in this social context that the U.S. government proposes using the “Internet+” model. Shared bicycle service to solve the “last mile” problem.

With the increase in the frequency of users using shared bicycles, the problems that arise are increasingly emerging, seriously affecting the user experience. For example, the imbalance in the number of bicycles between the site and the site is the most obvious issue. This paper only focuses on this issue to make inquiries and solutions. For example, shared bicycles in some stations only ride out but rarely ride in; on the other hand, some stations have too many bicycles to ride in, causing tension in the parking spaces. Users cannot find free parking spaces at nearby stations to return vehicles, etc. This is the situation where the number of bicycles and the number of free parking spaces at different sites are uneven. In the long run, there will be shared bicycles that are not available at some sites, and some sites may have problems due to the excessive accumulation of bicycles and the resulting shortage of returned areas. In addition, the different lifestyle habits of the users and different working
hours also result in different frequency of use of each shared bicycle site during the day and night [4-7]. What is more obvious is that the prosperous areas between different regions will also cause imbalances in the use of bicycles between stations. For example, in busy commercial areas or around schools, shared bicycles are in short supply; in some remote locations, shared bicycles are in excess supply. Moreover, there are also a large number of unbalanced number of shared bicycles between sites and sites in certain special locations (the number of shared bicycles returned within the site and the number of shared bicycles riding out are not balanced). For example, on weekends, many students will ride out shared bicycles in schools and park them in busy areas such as tourist attractions or business districts. After the game is over, very few students will ride shared bicycles back to school and the number of shared bicycles in schools will decrease. It is difficult for users to find shared bicycles, and it is difficult to use shared bicycles. On the other hand, in the tourist attractions and commercial areas, there will be a large accumulation of shared bicycles.

In order to solve the above problems, in order to improve user satisfaction and improve the utilization rate of shared bicycles, this paper proposes the use of quad-trees to find the most frequently used cycling areas, to determine whether there are site imbalances in this area, and to use regulatory functions. Unbalanced site imbalance adjustment [4].

2. Overview

2.1. Related Concept Definitions

2.1.1 Whether imbalance. This paper uses the ratio of users returning shared bicycles at a shared bicycle site to the number of users riding a shared bicycle from this site as an indicator of whether or not there is an imbalance in the site. If the ratio is within the normal range, mark the site as a normal site. The ratio is above and below the normal range, marking the site as returning unbalanced and riding out of balance.

\[
\frac{I}{O} \in \begin{cases} 
[1-\delta, 1+\delta] & \text{Not out of balance} \\
(1+\delta, \infty) & \text{Return imbalance} \\
(0, 1-\delta) & \text{Riding out of balance}
\end{cases}
\]  

(1)

\(I\) represents the number of shared bicycles returned by the user at this site, \(O\) represents the number of users riding a shared bicycle from this site, and \(\delta\) is a setting to determine whether the site is out of balance.

2.1.2. Imbalance ratio.

\[
ratio_i = \frac{\max[I_i, O_i] - \min[I_i, O_i]}{\min[I_i, O_i]}
\]

(2)

The \(ratio_i\) represents the imbalance ratio of the \(i\) site. \(I\) and \(O\) represent the number of shared bicycles returned by the user at the \(i\) site and the number of shared bicycles riding. \(ratio \in [0, \infty)\), the greater the \(ratio\) value, the more severe the site imbalance.

2.2. Using Quad tree Algorithm to Find Frequent Areas

Because the quad-tree algorithm has the advantage of fast query, fast deletion, etc., this paper uses the quad-tree algorithm to divide the frequency of use of all shared bicycle stations, and it can get the area where users use shared bicycles most frequently.
2.3. Judging Whether the Site is out of Balance
Determine if there are any imbalances in the sharing bike sites in the most frequently-ridden areas. If there are imbalance problems, identify which kind of imbalances exist (return imbalances, riding out of balance).

2.4. Adjust Imbalance Stations
Adjust unbalanced sites based on the distance between sites, the imbalance ratio and the number of shared bicycle imbalances.

3. Site Imbalance Adjustment and Implementation

3.1. Finding Frequent Areas
The data used in this article is the July 2015 public data of the Chicago shared bicycle, with a total of 533,709 data. The data includes the latitude and longitude coordinates of the shared bicycle site that the user rides, the shared bicycle site number, the name of the shared bicycle site, the time the user started riding, the return time, the latitude and longitude coordinates of the shared cycling site, the number, the name of the return site, and the like.

This paper uses quad-tree algorithm to divide the city into different regions for calculation. We set the total access threshold in the city to 1%. If the number of visits in a region is greater than 1% of the total count, then four equally large small regions are further divided. This process continues until all areas have equal or less than 1% of total count (see Algorithm 1).

Algorithm 1. Quad-tree algorithm
Step1 Create a root node $n$, with $P1$ and $P2$ as polygon limits;
Step2 Calculate the $count(v)$ of bike rides shared inside polygon $n$;
Step3 if $count(v) > visits$ threshold $Threshold_{visits}$ then
  Divide node $n$ into four equal regions;
  Create four nodes $n1, n2, n3, n4$ for each of divided regions;
  Assign the four nodes $n1, n2, n3, n4$ as children to node $n$;
  Now repeat the step 3 for all the four child nodes of $n$;
Step4 end

Figure 1. Frequency of sharing bike cycling
Figure 2. Whether there is an imbalance in the station

According to the quad-tree algorithm, we divide the data according to the frequency of each bicycle sharing site. We need to input the maximum and minimum latitude and longitude $P1 (lat1, lon1)$ and $P2 (lat2, lon2)$ threshold $t$ of allowed total count $v$ in a region.

According to the use of the quad-tree algorithm to divide the frequency of sharing bicycle cycling, it is divided into 1985 areas. From Figure 1, it can be seen from Figure 1 that the area where the user rides a shared bicycle most frequently is area 142.

3.2. Judging Whether the Station is Imbalance

According to Formula 1, imbalanced judgments are made for all shared bicycle stations. If there are imbalances in the shared bicycle stations and they are marked with tags, the results are shown in Figure 2 according to the imbalance of each shared bicycle station.

Among them, the only ride out tag means that the user only rides a shared bicycle from the shared bicycle site, the only return tag indicates that the user is sharing the bicycle site only to return the shared bicycle, the balance tag indicates that the shared bicycle site is not imbalanced, and the unbalance of return tag represents this sharing. There is an unbalanced return problem at the cyclist's site. The unbalance of ride out tag indicates that there is a problem of riding out of balance at the shared cyclist's site.

3.3. Imbalanced Site Adjustment

This paper randomly selects the unbalanced site $i$ in the area of 142 to adjust. $i$ site data has a latitude of 41.879255, a longitude of -87.639904, and the number of shared bicycles on site $i$ is 1318, and the number of returning shared bicycles is 450. According to Formula 1, judging that $i$ site has a ride imbalance, and the imbalance ratio is 1.93.

Due to the particularity of shared bicycles (it is not convenient to move a large number of shared bicycles), we need to change the location of the shared bicycles during site adjustment, and it may be necessary to adjust multiple sites to adjust one site. Therefore, it is necessary to use this area. There are multiple shared bicycle stations, and the regulatory sites cannot be too far apart from each other. Due to various reasons, we have determined the range to be 1 km (radius $R$). We look for any unbalanced sites within 1 km of the $i$ site. The results are shown in Figure 3.
According to Figure 3, there are many unbalance of return stations (represented by the red dots) within the distance from site i to R. In order to adjust the site imbalance problem, according to the return of the distance between the unbalanced site and the i site, the return of the site’s imbalance ratio and the return of the site's shared bicycle imbalance. The number of i sites together to adjust and generate adjustment function, see formula 3

$$\max \left\{ \frac{O_j - I_j}{I_i - O_i} \cdot \frac{\text{ratio}_j}{d(i, j)} \right\}$$

$I_j$ and $O_j$ indicates the count of shared bicycles returned by users in the unbalanced site and the count of users riding shared bicycles. $d(i, j)$ indicates the distance between two sites. The specific results are shown in Figure 4. $I$ indicate that there is a need to adjust imbalanced sharing bike sites, the size of the return site label is determined by 1% of the number of unbalance shared bicycles and numbered.

As shown in Figure 4, when adjusting the I site, according to formula 3, it is preferable to choose site 3 or station 9 to adjust the I site. After adjusting, we use formula 1 to judge whether the I site is unbalanced or not, and the adjusted imbalance ratio is 0.07. The adjustment result is very satisfactory.
4. Multi-Station Adjustment Test

Table 1. Multiple imbalance station adjustment tests

| No. | Latitude | Longitude | Returns | Rents | Imbalance label | Imbalance ratio | Imbalance label after adjustment |
|-----|----------|-----------|---------|-------|----------------|----------------|---------------------------------|
| 1   | 41.88918 | -87.6277  | 440     | 167   | +              | 1.634          | -                              |
| 2   | 41.8810317 | -87.62408432 | 471     | 111   | +              | 3.243          | -                              |
| 3   | 41.884451 | -87.629892 | 608     | 149   | +              | 3.081          | -                              |
| 4   | 41.8807   | -87.63947 | 448     | 706   | +              | 0.576          | -                              |
| 5   | 41.872373 | -87.633523 | 11      | 240   | +              | 20.818         | -                              |
| 6   | 41.879255 | -87.639904 | 450     | 1318  | +              | 1.929          | -                              |
| 7   | 41.870257 | -87.639474 | 2       | 22    | +              | 10.000         | -                              |
| 8   | 41.882091 | -87.639833 | 392     | 894   | +              | 1.281          | -                              |
| 9   | 41.886875 | -87.62603  | 280     | 49    | +              | 4.714          | -                              |
| 10  | 41.864059 | -87.623727 | 21      | 149   | +              | 6.095          | -                              |
| 11  | 41.880317 | -87.635185 | 1278    | 177   | +              | 6.220          | -                              |
| 12  | 41.882664 | -87.63253  | 358     | 63    | +              | 4.683          | -                              |
| 13  | 41.872077633 | -87.629543773 | 47      | 225   | +              | 3.787          | -                              |
| 14  | 41.88338  | -87.64117  | 701     | 1537  | +              | 1.193          | -                              |
| 15  | 41.867888 | -87.623041 | 75      | 437   | +              | 4.6827         | -                              |
| 16  | 41.879472352 | -87.625688606 | 375     | 52    | +              | 6.212          | -                              |
| 17  | 41.884576228 | -87.63188991 | 539     | 68    | +              | 6.926          | -                              |
| 18  | 41.881319815 | -87.629520919 | 1088    | 111   | +              | 8.801          | -                              |
| 19  | 41.860384 | -87.625813 | 17      | 198   | +              | 10.647         | -                              |
| 20  | 41.888243 | -87.63639  | 565     | 136   | +              | 3.154          | -                              |

This paper uses the adjustment method to adjust and test several sites and shows the results in Table 1, in which + indicates the imbalance of the site, - indicates that there is no imbalance in the site. The final adjustment results show that the use of this method can adjust the imbalance of the shared bicycle site and the accuracy rate is 95%.

5. Conclusion

Based on the analysis of shared bicycle data in Chicago, this paper first uses the quad-tree algorithm to obtain the most frequent use of shared bicycles. Second, according to formula 1, judging whether there is an imbalance problem in all shared bicycle stations, if there is an imbalance problem in this site, then determines which kind of imbalance state belongs to, and then query all unbalanced sites in the frequent area, and adjust the imbalanced site. After adjusting the unbalanced site, formula 1 is used to determine whether the unbalanced shared-vehicle is imbalanced. If the shared-cycling site after adjustment is still unbalanced, adjustment will continue until the site reaches equilibrium. If the shared bicycle still has imbalance problems after a certain number of consecutive adjustments, it will no longer be adjusted. This paper demonstrates the feasibility of this regulation method based on the test results of multiple unbalanced sites.

6. Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant number: 61401260) and the Natural Science Foundation of Shandong, China (Grant number: BS2014DX006)

7. References

[1] Larsen J (2013) Bike-sharing programs hit the streets in over 500 cities worldwide vol 1( Plan B Updates)
[2] González M C and Hidalgo C A(2008) Understanding individual human mobility patterns (Nature) pp 779-782
[3] Y Zheng, F Liu and H Hsieh (2013) U-air: when urban air quality inference meets big data (ACM SIGKDD ) pp 1436–1444
[4] R Beecham, J Wood and Bowerman A (2014) Studying commuting behaviours using collaborative visual analytics (Computer Environment and Urban System) pp 5-15
[5] N Lathia, C Smith, J Froelich and L Capra (2013) Individuals among commuters: Building personalised transport information services from fare collection systems vol 9 (Pervasive and Mobile Computing) pp 643- 664
[6] N Lathia, S Ahmed and L Capra (2012) Measuring the impact of opening the London shared bicycle scheme to casual users vol 21 (Transportation Research Part C) pp. 88-102
[7] M Vogel, R Hamon, G Lozenguez, L Merchez, P Abry, J Barnier, P Borgnat, P Flandrin, I Mallon, and Robardet (2014) From bicycle sharing system movements to users: a typology of Velo’v cyclists in Lyon based on large-scale behavioural dataset (Journal of Transport Geography) pp 280-291