

# **Dockless Shared Bicycle Flow Control by Using Kernel Density Estimation Based Clustering**

Shang-Yuan Chen<sup>\*</sup>, Tzu-Tien Chen

School of Architecture, Feng Chia University, Taichung, Taiwan

Received 31 October 2020; received in revised form 29 January 2021; accepted 03 February 2021

DOI: <https://doi.org/10.46604/aiti.2021.6666>

## **Abstract**

Since dockless sharing bicycles have become an indispensable means of everyday life for urban residents, how to effectively control the supply and demand balance of bikes has become an important issue. This study aims to apply Kernel Density Estimation based (KDE-based) clustering analysis and a threshold-based reverse flow incentive mechanism to encourage the users of bicycles to adjust the supply and demand actively. And it takes Shanghai Jing'an Temple and its surroundings as the research area. Its practical steps include: (1) compilation and processing of the needed data, (2) application of KDE-based clustering, partitioning, and grading, and (3) incentives calculation based on dockless shared bicycle flow control system. The study finds that the generalization function of KDE-based clustering can be used to estimate the density value at any point in the study area to support the calculation of the incentive mechanism for bicycle reverse flow.

**Keywords:** classification, clustering, unsupervised neural networks learning, R language, choropleth map

## **1. Introduction**

The promotion of shared bicycles seeks to resolve the problem of short-distance urban transportation, and shared bicycles are an important means of contemporary green transportation [1]. The docking shared bicycles found throughout Taiwan are chiefly aimed at tourists. However, unlike dockless shared bicycles, docking shared bicycles do not allow users to obtain and return the bikes at any location, thus docking shared bicycles cannot easily become important means of everyday transportation for urban residents. In addition, although dockless V-bikes made a short-lived appearance in Taiwan during 2017, little data concerning V-bikes is available. As a consequence, this study relies on observations and analysis of dockless shared bicycles in Shanghai, China to gain a better understanding of dockless shared bicycles and methods of improving their effectiveness, as well as specific methods of making them more acceptable to city residents and a more important means of transportation.

After the shared bicycles are introduced on a large scale in a city, the supply and demand problems of bicycles involving different areas often emerge. For instance, at certain important transportation nodes, the use of bicycles frequently increases during peak times, and may cause bicycle shortages. Therefore, it is necessary to dispatch bicycles in order to resolve the supply and demand problems. However, the current dispatching approach is often too inefficient to be of much use. Furthermore, it also imposes logistics costs on bicycle companies because manpower and trucks are required to move bicycles from oversupplied areas to shortage areas. Besides, since bicycles cannot be used while they are being transported, the capacity of urban transportation is reduced. As a consequence, how to dispatch shared bicycles in real-time, or in advance, is an especially important issue.

---

<sup>\*</sup> Corresponding author. E-mail address: shangyuanc@gmail.com

Tel.: +886-4-24517250

In this study, we apply cluster analysis with unsupervised learning to shared bicycle flow control and dispatching, and employ artificial intelligence to extract the real and full-scale transportation rules from open data. This study's analysis of the times and locations at which the oversupply or shortage of shared bicycles may occur in Shanghai will help bicycle companies to perform dispatching ahead of time. Furthermore, clustering results can be used in conjunction with incentive mechanisms employing threshold values to encourage bicycle users to actively perform bicycle dispersal and dispatching, which will boost the carrying capacity of urban transportation systems, and reduce the risk and cost of ineffective dispatching.

## 2. Literature Review

The scope of this study includes shared bicycles and cluster analysis; the following is a review of the literature:

### 2.1. Shared bicycles

Shared bicycle systems include docking and dockless models. The world's first intelligent public bicycle system was the Vélo'v docking public bicycle system, which was established in Lyon, France, in 2005 [2]. Nevertheless, docking shared bicycles have never caught on in China. Since the residents of China consider bicycles to be everyday means of transportation, many dockless shared bikes began appearing in China during the second half of 2016. They are not dependent on the fixed and managed parking areas, and can be left in any vacant space along roads. Consumers only need to use Wechat Pay, Alipay, or some other apps to pay a specified deposit before using the bikes. The following explains the clustering principles of the docking and dockless shared bicycles, and the differences between the two.

#### (1) Docking shared bicycles and their clustering principles:

It is much easier and more clear-cut to calculate the number of docking bikes that are taken or returned than dockless bikes. To predict the number of bikes at each group of docks in a city as a whole and after clustering, it is first necessary to perform the grouping of the docking points. In the case of docking shared bikes, because the locations at which bicycles can be obtained or returned are limited to places where docks are present, it is only necessary to perform clustering of the fixed docks, instead of the bikes themselves. If within a specific geographical scope, we divide  $n$  docks into  $m$  groups ( $n > m$ ), calculate the probability of bikes moving among the  $m$  groups, and analyze their historical data (including time, climate, notable events, and other influencing factors), we can then predict future movement trends [3].

#### (2) Dockless shared bikes and their clustering principles:

Unlike the case of shared bicycles with fixed docks, the clustering of dockless shared bicycles changes with time, and it is challenging to calculate check-out and check-in probabilities. One feasible approach that can be employed to cluster dockless shared bicycles is to perform clustering of the coordinates of the bicycles' check-out and check-in locations. However, it is still hard to perform direct clustering due to the fact that check-out points and check-in points have an irregular and macro distribution. Moreover, the clustering of dockless shared bicycles must also reflect the precondition that users within each cluster can easily find out usable shared bicycles within a comfortable walking distance. Besides, the most suitable walking distance is affected by climatic conditions, vertical distance, and a city's state of economic development [4].

In the case of Shanghai community residents, taking a walking distance of 787m as a dividing point, residents would choose to walk when the distance is shorter than 787m, but would very seldom choose to walk when the distance is longer than 787m [5]. This study, therefore, adopts the round figure of 787m as the scope of the most suitable walking distance; 787m is established as the baseline scope of clusters in the remainder of this paper. Since 787m is roughly equivalent to the distance that can be walked in 10 minutes, it is a representative figure of daily walking distance. As shown in Fig. 1, in a zone, the distance for users to find a bicycle should not exceed the "optimum walking distance". The minimum diagonal of the equilateral rectangle is 787m, which corresponds approximately to the cell grid with a side length of 550m, reflecting the unit distance of the longitude and latitude grid in the Shanghai area.



Fig. 1 The hypothesis for the cell grid

## 2.2. Cluster analysis

As described above, the methods and algorithms used to cluster shared bikes affect the shared bicycle flow prediction and dispatching ability. In machine learning, the use of clustering methods and the derivation of clustering algorithms are referred to as “cluster analysis”. Cluster analysis is defined as the division of different objects into homogeneous subgroups [6].

Data clustering is ordinarily classified as unsupervised learning. While clustering and classification are easily confused, particularly in data mining, the difference between clustering and classification lies in the fact that clustering is a process of grouping to form samples, where the examples are adjacent in space. In contrast, classification involves attaching labels to samples, where the labels are derived from supervised external standards. In other words, clustering constitutes an unsupervised grouping process, where groups are formed after data has been input naturally [7], but the classification is supervised and involves other external information. Unsupervised neural networks can uncover essential characteristics, or that may be overlooked, from data on their own, without any preconceived expectation of the output values, and can, therefore, perform clustering of data [8-9].

Different clustering methods of unsupervised learning use mathematical formulas as their algorithms. Although cluster analysis has been used to develop various types of algorithms [10-16], those algorithms lack a uniform classification. In terms of classification, this study synthesizes the views of Gan et al. (2007) [11] and Saxena et al. (2017) [14]. Gan created the parsing tree based on the characteristics of clustering algorithms. From the top, this tree lists hard clustering and fuzzy clustering. Hard clustering includes partitional and hierarchical approaches. Hierarchical approaches may consist of either divisive or agglomerative methods [11]. Also, Saxena et al. suggested the following three additional categories for applying clustering techniques: distance-based methods, model-based methods, and density-based methods (Fig. 2) [14].

According to the nature of the dockless shared bicycle problem, they are often distributed in arbitrary shape in a city. Therefore, this study proposes to employ density-based clustering algorithms. Most partitioning methods are based on the distance between objects. Such methods can find only spherical-shaped clusters and encounter difficulty in discovering clusters of arbitrary shapes. However, the density-based method can be used to filter out the noise or outliers and identify the clusters of arbitrary shapes. There are two types of density-based approaches. One is density function, which is a total mathematical function, e.g. Kernel Density Estimation based (KDE-based) Clustering, and the other is density-based connectivity, e.g. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [17].

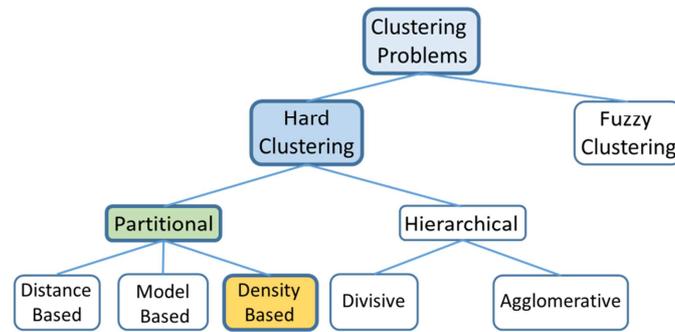


Fig. 2 The clustering method parsing tree

Comparing the KDE-based clustering with DBSCAN methods, the former offers the advantage of the setting of generalization function, which is convenient to partition and grading and is capable of being the basis for incentive mechanism. However, the latter can only be used to mark those clusters that meet a pre-set density requirement by the control radius  $\epsilon$  and MinPts quantity. KDE-based clustering has long been used for detecting traffic accident hot spots and has proven to be useful in traffic accident analysis [18]. In particular, the ordering capabilities of KDE-based clustering help compare the spatial structure to determine or assign the importance of mitigation measures [19-20]. As a consequence, the following research will discuss the introduction of KDE-based clustering applied to the dockless shared bicycle flow control system.

### 3. Theory and Method

As mentioned above, this study proposes a “dockless shared bicycle flow control system” that is an incentive mechanism to encourage the reverse flow of bikes in which KDE-based clustering is introduced to partition and grading as the basis of threshold control. The activity of the dynamic bicycle control system is not evenly distributed but varied in intensity and scope following people’s everyday activities. Because the system’s most important function is to maintain a balance between supply and demand in clusters with different densities, the incentive mechanism is used to reverse supply and demand imbalances during earlier periods. This solution brings up the question of how to perform partitioning and grading of bicycle check-out points and check-in points. When check-outs and check-ins span different clusters, how should incentives be provided to reverse bicycle movement trends? The following explains KDE-based clustering partition and grading and dockless shared bicycle flow control system.

#### 3.1. KDE-based clustering partition and grading

KDE-based clustering analysis over a density surface is based on the technologies of spatial smoothing and spatial interpolation [21-22]. The kernel function analyzes and converts the point distribution of the plane in the area, and infers the spatial density distribution in the area. The spatial density is the property that describes the spatial distribution in an intensive or sparse pattern in quantity. Therefore, the point data is mapped from the inseparable 2-dimensional space to the separable 3-dimensional space. Since a density value is calculated as an attribute of each grid cell in the sub-region, it is, therefore, possible to represent the spatial distribution employing homogenous and easily comparable areas with choropleth map visualization. This estimated density is used to cover the top of the research domain with a grid cell, and the estimation of density is performed based on each grid cell’s center point (Fig. 3). Based on defined kernel function and bandwidth, KDE provides an estimate of the intensity in each position of the grid cell by moving three-dimensional functions. The defined kernel function weights events within its sphere of influence according to their distance from the point at which the intensity is estimated (Fig. 4) [23]. In other words, the nature of the optimization and generalization function is a significant characteristic of this density estimation method, which can perform fitting and generalization of large bodies of complex data following the approximation function and bandwidth, and can possess estimation ability.

Commonly used density functions include Uniform, Epanechnikov curve, Quartic curve, and Gaussian curve, etc. (Fig. 5). Among them, Quartic and Gaussian are more suitable radial basis function kernels but the Quadratic kernel is less computationally intensive than the Gaussian kernel and can be used as an alternative when using the Gaussian becomes too time-consuming and tedious [20]. After optimization, the system can estimate the density value of any position in the study area.

Since a density value is calculated as an attribute of each quadrat sub-region, it is, therefore, possible to represent the spatial distribution employing homogenous and easily comparable areas with choropleth map visualization [20]. The general form of a kernel density estimator is expressed as:

$$f(s) = \sum_{i=1}^n \frac{1}{h^2} k \left[ \frac{d(s, c_i)}{h} \right] \tag{1}$$

where  $f(s)$  is the density measured at location  $s$ ,  $h$  is the search radius (bandwidth) of the KDE [only events within  $h$  are used to estimate  $f(s)$ ],  $c_i$  is the observed event point,  $k[ ]$  is the weight of event  $c_i$  at distance  $d(s, c_i)$  to location  $s$ . The KDE usually models the so-called kernel function,  $k$ , as a function of the ratio between  $d(s, c_i)$  and  $h$  so that the “distance decay effect” can be taken into account in density estimation. The evaluation of KDE requires two parameters: the bandwidth  $h$  and the kernel function  $k$ , which determine the weighting of the points. In this study, we use the Quartic kernel function, which is one of the most commonly used features:

$$k \left[ \frac{d(s, c)}{h} \right] = \frac{3}{4} \left\{ 1 - \frac{[d(s, c)]^2}{h^2} \right\} \tag{2}$$

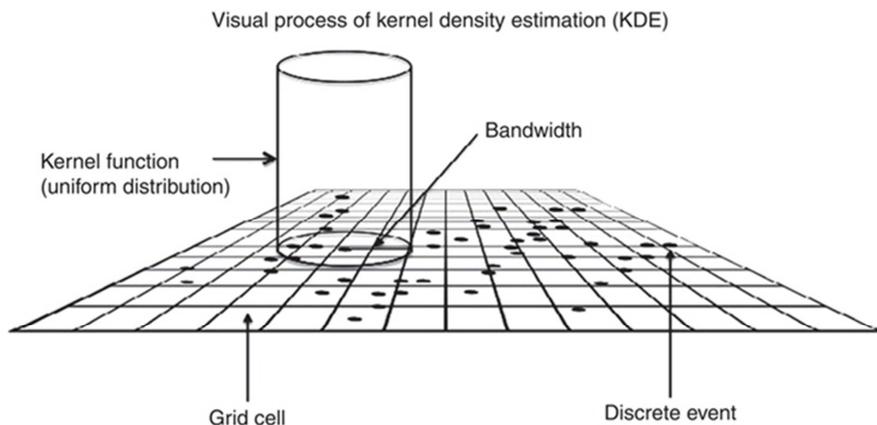


Fig. 3 Performing density estimation based on the center point of each grid cell [24]

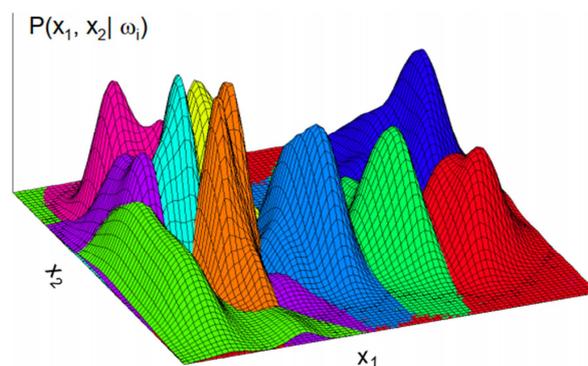


Fig. 4 Weighting ( $\omega_i$ ) based on a specific kernel function and bandwidth [25]

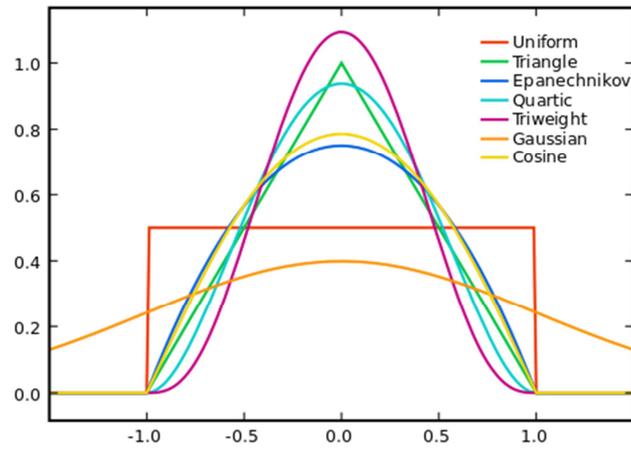


Fig. 5 Kernel functions in a standard coordinate system [26]

3.2. Dockless shared bicycle flow control system

It is crucial that shared bicycle companies possess real-time or ahead of time dispatching and management ability. This study proposes a dockless shared bicycle flow control system model and an incentive mechanism for the reverse flow of bikes based on threshold values. The clustering results can be used in conjunction with incentive mechanisms employing threshold values to encourage bicycle users to perform bicycle dispersal and dispatching actively. It boosts the carrying capacity of urban transportation systems, and reduces the risk and cost of ineffective dispatching.

As shown in Fig. 6, this system considers the number of check-outs within a unit time and a unit area to be the bicycle demand of that area, and the number of check-ins within a unit time and a unit area to be the bicycle supply of that area. It is necessary to use incentives or compulsory rules to control the rate of check-outs and check-ins to ensure that each region has a proper amount of bicycle stock. When the stock is excessively high—higher than the upper threshold—the system triggers the check-out incentive mechanism. When the stock is too low—below the lower limit—the system activates the check-in incentive mechanism. These incentive mechanisms controlling reverse flow based on threshold values can be expected to ensure that the stock can be in a better state than when not using the incentive mechanisms. Apart from this, when the system detects that the bicycle density reaches the level of oversupply or shortage, the shared bicycle company have to dispatch trucks to perform transport and force the system back within the normal range [27].

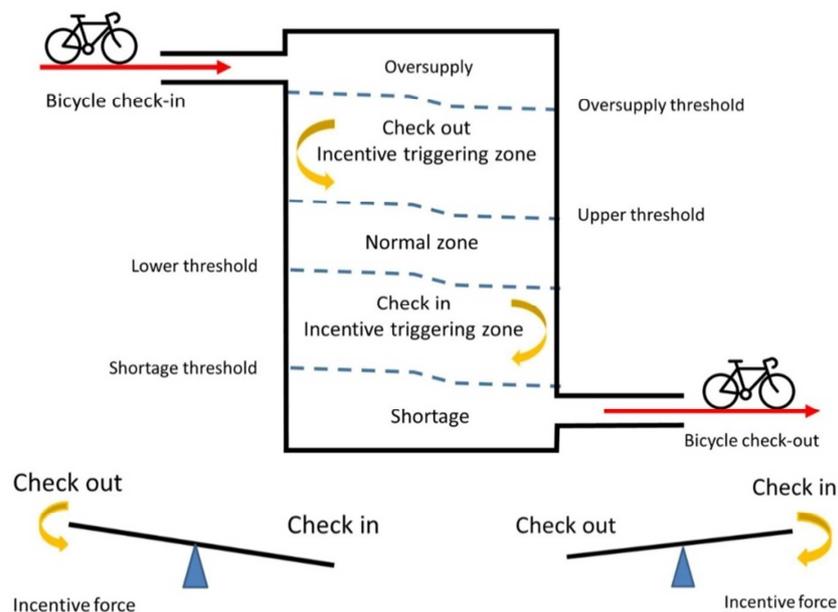


Fig. 6 Dockless shared bicycle control system model

For every partition, the most important aspect is maintaining a balance between supply and demand within each cluster in this dynamic system, and incentives must seek to reverse imbalances in supply and demand during the previous period. In this study, we consequently design a check-out incentive formula (referring to Eq. (3)) and a check-in incentive formula (referring to Eq. (4)), in which  $X$  is the check-out rate,  $Y$  is the check-in rate,  $t_n$  is the time, and  $n$  is an integer.

If  $X_{t_{n-1}} \leq Y_{t_{n-1}}$  at the time  $t_n$ , then the check-out incentive amount is expressed as:

$$\frac{|Y_{t_{n-1}} - X_{t_{n-1}}|}{2^i} \quad (3)$$

where  $i$  is a whole number and is the check-out sequence at a time  $t_n$ . By Eq. (3), at the time  $t_n$ , the incentive for check-out of the first bicycle is maximum, and the impetus for the check-out of each subsequent bike decreases progressively with the increasing root of  $1/2$  with the increase in the number of bikes checked out:  $1/2^1$ ,  $1/2^2$ ,  $1/2^3$ , and so on. The incentive for the final bicycle checked-out approaches 0 without limit, but never reaches 0.

If  $X_{t_{m-1}} \leq Y_{t_{m-1}}$  at the time  $t_m$ , then the check-in incentive amount is expressed as:

$$\frac{|X_{t_{m-1}} - Y_{t_{m-1}}|}{2^j} \quad (4)$$

where  $j$  is a whole number and is the check-in sequence at a time  $t_m$ . By Eq. (4), at time  $t_m$ , the incentive for check-in of the first bicycle is maximum, and the impetus for the check-in of each subsequent bike decreases progressively with the increasing root of  $1/2$ :  $1/2^1$ ,  $1/2^2$ ,  $1/2^3$ , and so on. The incentive for the final bicycle checked-in approaches 0 without limit, but never reaches 0.

As described above, if a bicycle is checked out from area  $p$  at time  $t_n$  and checked in area  $q$  at  $t_m$ , the incentive amount is the sum of the check-in incentive in area  $p$  during at time tout and the check-out incentive in area  $q$  during at time  $t_{in}$ . The total incentive amount can be represented as:

$$Sum = \frac{|Y_{t_{n-1}} - X_{t_{n-1}}|}{2^i} + \frac{|X_{t_{m-1}} - Y_{t_{m-1}}|}{2^j} \quad (5)$$

#### 4. Realization and Verification

This section realizes and tests a dockless shared bicycle flow control system following the preceding theory and method. We employ 101,843 data items issued in August 2016 for the Shanghai Open Data Innovative Application Competition to perform modeling and verification [28]. These data items include order numbers, bicycle serial numbers, user serial numbers, check-out times, check-out longitudes, check-out latitudes, check-out times, check-in longitudes, check-in latitudes, and routes. We use this data to find the optimal clustering methods for shared bikes in Shanghai by examining relatively small areas (the area around Shanghai's Jing'an Temple) and relatively large areas (the part of Shanghai within the fourth ring roads—121.75°E-121.125°E, 30.95°N-31.45°N) and also establish the models corresponding to the state of clustering. This control system applies the KDE-based clustering analysis to propose an incentive mechanism for the reverse flow of bikes based on partition and threshold to encouraging bicycle users to evacuate and dispatch actively. We employ the following three steps in this process: (1) compilation and processing of the needed data, (2) application of KDE-based clustering, partitioning, and grading, and (3) incentives calculation based on dockless shared bicycle flow control system. In the development of the application, the RStudio software is used as an integrated development environment for writing R language programs [29].

4.1. Stage 1: Compilation and processing of data

As of August 2016, 223,000 Mobike bikes are in use in Shanghai [28]. Since each Mobike bike in China is ridden by 5.4 persons per day [30], we can roughly estimate that there are approximately 36,000,000 use records each month for Mobike bikes in Shanghai during August 2016. Based on the Shanghai area’s latitude and longitude, as well as a maximum suitable walking distance of 787m, which corresponds approximately to the cell grid with a side length of 550m, we can estimate that there is an average of approximately 2,857 bikes in each 550m×550m grid square at that latitude and longitude (discarded, damaged, and under-repair bikes were not subtracted). Assuming that 50% of bikes in the urban area of Shanghai can still be found, we can estimate that there are approximately 1,428 bikes in each grid square. The following modeling, therefore, employs a common stock of 1,428 bikes in each 550m×550m grid square as the reference bicycle stock value for Shanghai during August 2016. 550m×550m squares are designated as grid cells for KDE-based clustering calculations.

4.2. Stage 2: Application of KDE-based clustering, partitioning, and grading

As described above, the activity of the dynamic bicycle control system is not evenly distributed but varied in intensity and scope following people’s everyday activities. Because the system’s most important function is to maintain a balance between supply and demand in clusters with different densities, the incentive mechanism is used to reverse supply and demand imbalances during earlier periods. The following is an explanation of how this is done:

Following the analysis in section 3.2, the kernel density spatial clustering method is used to perform partitioning and grading of check-out and check-in density in the study area around Shanghai’s Jing’an Temple on 8/1/2016. In this study, we set 787m as bandwidth and choose the Quartic kernel function to execute KDE-based clustering.

(1) Check-out density partitioning and grading:

As shown in Table 1, according to the results of calculations, the level 1 area around Jing’an Temple has a check-out density of approximately 3,600, while the adjacent peripheral area has a check-in density of roughly 900. This number shows that the density difference between the central region and the adjacent outer space is around  $3,600 - 900 = 2,700$ . Following the depth of the purple color in Fig. 7, the system recommends that there be 12 clusters; when 2,700 points are divided among levels 1 to 12, we obtain the clustering thresholds for each level shown in Table 1.

Table 1 Check-out clustering thresholds for each level

Areas with different check-out clustering levels	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10	Level 11	Level 12
Upper threshold	$\geq 3600$	3375	3150	2925	2700	2475	2250	2025	1800	1575	1350	1125
Lower threshold	3375	3150	2925	2700	2475	2250	2025	1800	1575	1350	1125	$\leq 900$

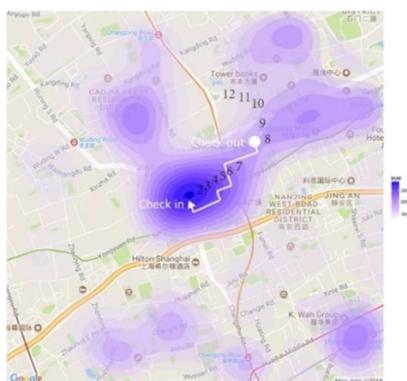


Fig. 7 Partitioning of check-out density in the area around Shanghai’s Jing’an Temple on 8/1/2016

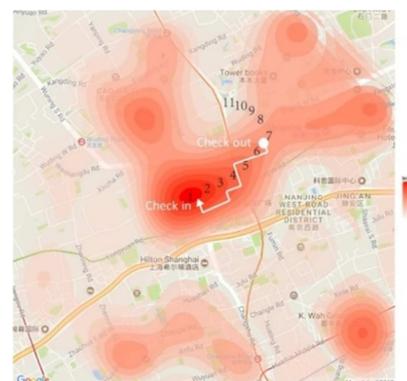


Fig. 8 Partitioning of check-in density in the area around Shanghai’s Jing’an Temple on 8/1/2016

## (2) Check-in density partitioning and grading:

As shown in Table 2, according to the results of calculations, the level 1 area around Jing'an Temple has a check-in density of approximately 3,100. In contrast, the adjacent peripheral area has a check-in density of roughly 800. This number shows that the density difference between the central region and the adjacent outer space is around  $3,100 - 800 = 2,300$ . Following the depth of the orange color in Fig. 8, the system recommends that there be 11 clusters; when 2,300 points are divided among levels 1 to 11, we obtain the clustering thresholds for each level shown in Table 2.

Table 2 Check-in clustering thresholds for each level

Areas with different check-in clustering levels	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9	Level 10	Level 11
Upper threshold	$\geq 3100$	2891	2681	2472	2263	2055	1845	1636	1427	1218	1009
Lower threshold	2891	2681	2472	2263	2055	1845	1636	1427	1218	1009	$\leq 800$

## 4.3. Stage 3: Incentives calculation based on dockless shared bicycle flow control system

As shown in Figs. 7-8, during the period from 12:00 to 18:00 on 8/1/2016 (i.e., the period following 06:00 to 12:00 on 8/1/2016), the first bicycle is checked-out from the check-out point and returned at the check-in point half an hour later, which is during the same period. The calculations of the check-outs and check-ins are presented as follows.

- (1) As for the check-outs, the estimated check-out rate  $X$  at the check-out point is 1600, and this area is in a level 7 check-in cluster (threshold: 1,845-1,636); since  $1600 < 1636$  (implying that the check-out rate  $X$  is slower than the minimum estimated check-in rate), consequently when  $X_{\min} \leq X_{t_{n-1}} \leq X_{\max}$ , and  $X_{t_{n-1}} < Y_{\min}$ . The check-out incentive system should be activated. According to Eq. (3),  $Y_{t_{n-1}}$  is entered as  $Y_{\min}$ ,

$$\frac{|1636 - 1600|}{2^1} = 18 \quad (6)$$

- (2) As for the check-ins, the estimated check-in density  $Y$  of the check-in point is 2,800; and this area is in a level 2 check-out cluster (threshold: 3,375-3,150); since  $2750 < 3150$  (implying that the check-in rate  $Y$  is slower than the minimum estimated check-out rate), when  $Y_{\min} \leq Y_{t_{m-1}} \leq Y_{\max}$ ,  $X_{\min} \geq Y_{t_{m-1}}$ . The check-in incentive system should be activated. According to Eq. (4), when  $X_{t_{m-1}}$  is entered as  $X_{\min}$ ,

$$\frac{|3150 - 2750|}{2^1} = 200 \quad (7)$$

- (3) The total incentive is the sum of the two amounts. According to Eq. (5), a total incentive amount of 218 units is obtained.

$$Sum = 18 + 200 = 218 \quad (8)$$

The preceding is a discussion of partitioning and grading and the use of an incentive mechanism relying on threshold values to reverse the flow of bikes. The data for the period from 12:00 to 18:00 on 8/1/2016 obtained from on-site observations is very consistent with the results of our model. Because 8/1/2016, Monday, is the first working day of the week, it is easy to find shared bikes in the outer marginal area of Jing'an Temple. But in contrast, the Jing'an Temple center is located in a dense commercial and office area, so it is often difficult to find bikes that can be borrowed.

## 5. Conclusions

With KDE-based clustering, this study can use the choropleth map visualization to represent spatial distribution to achieve partitioning and grading. It can support the calculation of the threshold-based incentive mechanism for bicycle reverse flow. The analysis of the implementation results is as follows:

- (1) If an area is in a peak usage period, the flow rate is fast (i.e., the pick-up rate is fast, and the return speed is also fast). Alternatively vice versa, if the relative time is in the off-peak period. When the absolute value of the difference between the check-out rate and the check-in rate is in a range, the number of bikes in the area is considered to be in a state in which regular operation can be maintained.
- (2) However, if a domain is in a fast check-out speed for a long time, and the check-in speed is very slow or almost stationary, it means that the number of bikes in the area has been slower than the shortage threshold. Conversely, if the check-in rate is rapid for a prolonged period within a specific area, but the check-out rate is almost 0, this indicates that the number of bikes in that area quickly rises above the oversupply threshold. When states of oversupply or shortage occur, trucks should be dispatched to move and redistribute the bikes in order to ensure that the system returns to a healthy state.

KDE-based clustering has its constraints to establish dispersal and dispatching strategies. As analyzed above, although the incentive mechanism can alleviate the situation of imbalance between supply and demand, the oversupply or shortage of bicycles still cannot be completely ruled out. The DBSCAN clustering method, is better than KDE-based clustering obviously in this situation. Therefore, how to use The DBSCAN cluster analysis to find out the oversupply or shortage of distribution of bikes in the study area to establish dispersal and dispatching strategies will be the core issue of the next research.

## Acknowledgments

This work was supported by project: Cluster Analysis of Unsupervised Learning Applied to Dockless Shared Bicycle Traffic Prediction and Scheduling, MOST 108-2221-E-035-005.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- [1] S. K. Chang, H. W. Chang, and Y. W. Chen, Green Transportation, Slow, Friendly, and Sustainable: People-Oriented Transportation Environment Makes The City Smoother and Life Better, Taipei: Neonaturalism, 2013. (In Chinese)
- [2] W. Zhu, J. Y. He, and D. Wang, "Methods and Empirical Research on the Distribution of Public Bicycle Systems in France: Case Study on Paris and Lyon," *Urban Planning International*, vol. 30, no. S1, pp. 64-70, 2015.
- [3] Y. Li, Y. Zheng, H. Zhang, and L. Chen, "Traffic Prediction in a Bike-Sharing System," *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, no. 33, November 2015, pp. 1-10.
- [4] J. Speck, *Walkable City: How Downtown Can Save America, One Step at a Time*, New York: North Point, November 2012.
- [5] N. Wang and Y. C. Du, "Resident Walking Distance Threshold of Community," *Transport Research*, vol. 1, no. 2, pp. 20-24, June 2015. (In Chinese)
- [6] M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, London: Pearson Education Limited, 2011.
- [7] J. C. Príncipe, N. R. Euliano, and W. C. Lefebvre, *Neural and Adaptive Systems: Fundamentals through Simulations*, New York: Wiley, 2000.
- [8] M. C. Su and X. D. Zhang, *Machine Learning: Neural Networks, Fuzzy Systems, and Gene Algorithms*, 2nd ed. Taipei: Chuan Hwa, 2004. (In Chinese)
- [9] A. D. Kulkarni, *Computer Vision and Fuzzy-Neural System*, New Jersey: Prentice Hall PTR, 2001.

- [10] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*, London: Chapman and Hall/CRC, 2013.
- [11] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, USA: Society for Industrial and Applied Mathematics, 2007.
- [12] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Massachusetts: Morgan Kaufmann, 2012.
- [13] S. K. Popat and M. Emmanuel, "Review and Comparative Study of Clustering Techniques," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 805-812, 2014.
- [14] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, et al., "A Review of Clustering Techniques and Developments," *Neurocomputing*, vol. 267, pp. 664-681, December 2017.
- [15] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 1st ed. London: Pearson, pp. 491-493, 2005.
- [16] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. USA: Elsevier, December 2009.
- [17] H. Shah, K. Napanda, and L. D'mello, "Density-Based Clustering Algorithms," *International Journal of Computer Sciences and Engineering*, vol. 3, pp. 54-57, 2015.
- [18] Z. Xie and J. Yan, "Detecting Traffic Accident Clusters with Network Kernel Density Estimation and Local Spatial Statistics: an Integrated Approach," *Journal of Transport Geography*, vol. 31, pp. 64-71, July 2013.
- [19] T. K. Anderson, "Kernel Density Estimation and K-Means Clustering to Profile Road Accident Hotspots," *Accident Analysis & Prevention*, vol. 41, no. 3, pp. 359-364, May 2009.
- [20] W. Yu and T. Ai, "The Visualization and Analysis of Urban Facility Pois using Network Kernel Density Estimation Constrained by Multi-Factors," *Boletim de Ciências Geodésicas*, vol. 20, no. 4, pp. 902-926, October 2014.
- [21] M. C. Jones, "Variable Kernel Density Estimates, and Variable Kernel Density Estimates," *Australian Journal of Statistics*, vol. 32, no. 3, pp. 361-371, 1990.
- [22] T. C. Bailey and A. C. Gatrell, *Interactive Spatial Data Analysis*, Harlow: Longman, 1995.
- [23] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson, "Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology," *Transactions of the Institute of British Geographers*, vol. 21, no.1, pp. 256-274, 1996.
- [24] T. Hart and P. Zandbergen, "Kernel Density Estimation and Hotspot Mapping: Examining the Influence of Interpolation Method, Grid Cell Size, and Bandwidth on Crime Forecasting," *Policing: An International Journal of Police Strategies & Management*, vol. 37, no. 2, pp. 305-323, May 2014.
- [25] G. O. Ricardo, "Kernel Density Estimation," [http://csyue.nccu.edu.tw/ch/Kernel%20Estimation\(Ref\).pdf](http://csyue.nccu.edu.tw/ch/Kernel%20Estimation(Ref).pdf), August 2019.
- [26] "Kernel (Statistics)," [https://en.wikipedia.org/wiki/Kernel\\_\(statistics\)](https://en.wikipedia.org/wiki/Kernel_(statistics)), August 2019.
- [27] S. Y. Chen and T. T. Chen, "Application of Cluster Analysis with Unsupervised Learning to Dockless Shared Bicycle Flow Control and Dispatching," *Computer-Aided Design and Application*, vol. 17, no.5, pp. 1067-1083, 2020.
- [28] "Soda Open Data," <http://shanghai.sodachallenges.com/data.html>, August 2019. (In Chinese)
- [29] J. P. Lander, *R for Everyone: Advanced Analytics and Graphics*, Boston: Addison-Wesley Professional, December 2013.
- [30] Beijing Planning Design Research Institute, "2017 Sharing Bike, and Urban Development White Paper", <http://www.199it.com/archives/581592.html>, 2017. (In Chinese)



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).