Location Analysis of Urban Electric Vehicle Charging Metro-Stations Based on Clustering and Queuing Theory Model

Chongyu Chen[1], Teng Li[1], Shugui Wang[1], Zhenyao Hua[1,2,3], Zecheng Kang[1,2,3], Dongyang Li[2,3], and Weian Guo[1,2,3](✉)

[1] Sino-German College of Applied Sciences, Tongji University, Shanghai 201804, China
guoweian@tongji.edu.cn

[2] School of Information and Electronics Engineering, Tongji University, Shanghai, China

[3] Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China

Abstract. The holding capacity of electric vehicles becomes increasingly huge. Limited by the battery capacity and charging speed, the market of the electric charging stations is very large. Moreover, due to the restrict of the urban planning resources, it is important that electric charging stations are reasonably located. In this paper, through the analysis of electric vehicles' trajectory in Chengdu and the application of K-Means clustering algorithm and Queuing Theory, the reasonable charging station locations and the numbers of charging piles are proposed. The model is conducive to reduce the users' queue length for charging.

Keywords: Electric charging station · Queuing theory · Clustering

1 Introduction

With the shortage of fossil energy and environmental pollution, the development of new alternative energy has been drawn attentions world-widely. By reducing carbon dioxide emissions and switching to electricity in the ground transport sector, China will achieve carbon neutral target by 2060 [1]. Therefore, the Chinese government has introduced policy measures and financial incentives to promote the development of electric vehicles [2]. Compared with traditional fuel vehicles, electric vehicles not only have lower noise, but also have stronger start dynamic. In addition, electric vehicles will not emit nitrogen oxides and other polluting gases [3], which play an important role in protecting the environment. With the dual support of national policy measures and market demand, the research and development of electric vehicles in China will become the future traffic industry main trend. More and more papers and related products have laid a solid foundation for the development of electric vehicles [4]. The literature [5] mainly discussed the comparison between the resource consumption of electric vehicles and that of traditional fuel vehicles in China. After analyzing the resource consumption of the production, using and recycling of the two kinds of vehicles, it is concluded that

L. Pan et al. (Eds.): BIC-TA 2021, CCIS 1566, pp. 282–292, 2022.
https://doi.org/10.1007/978-981-19-1253-5_21

the resource consumption of an electric vehicle is less than that of a fuel vehicle [6]. In order to encourage the public to use electric vehicles, China has introduced many specific measures. Relevant policy ran in 25 demonstration cities, which also had subsidies and policies [7, 8]. AI is also used in developing cruising strategy of electric taxis [9]. By then, the number of electric vehicles will be further increasing. The following electric vehicle facilities also need to be improved.

Electric charging station is an essential foundation facility for electric vehicles. The location and capacity of charging station play a very crucial role in improving the efficiency of charging service. Up to date, many researchers focused on this topic and conduct many works on this area. Wei [10] used a hybrid model combining Cellular Automata (CA) and Agent-Based-Modeling (ABM) to analyze the location of electric vehicle charging stations. In order to minimize the total cost, Xiao [11] established a charging queuing model considering the limited queue length and different location constraints, and determined the optimal location and capacity of electric vehicle charging facilities. Huang [12] proposed a robust deep K-Means model to exploit the hierarchical semantics of data in a layer wise way. Zhang and Yang [13] optimized the application of Simultaneous Heat Transfer Search (SHTS) in K-Means.

These literatures only carried out theoretical analysis without combination with the actual travel behavior of users and the city's street layout to make specific suggestions on the charging station location. At present, there are few literatures about the location of charging stations, though there are many benefits of large charging stations, such as safety, efficiency, manageability and high capacity. Due to large investments, the location of the large charging station is particularly important. Through reasonable location and unified management, large charging stations can not only satisfy and promote the popularization of electric vehicles, but also bring huge economic benefits to the society. To address these problems, in this paper we propose a reasonable charging station location proposal based on a specific operating vehicle trajectory in our research. In order to obtain the proposed location, this paper uses K-Means clustering algorithm to analyze urban driving trajectory data set of electric operating vehicles. Then, in order to optimize the charging piles' number of each station, this paper analyze the average queuing time of users using the queuing theory M/M/C model.

The rest of this paper is organized as follows. In Sect. 2, we come up with assumptions and describe a model. In Sect. 3, we briefly introduce a clustering algorithm and queuing theory. In Sect. 4, we take Chengdu City, China as an example to implement the algorithms and provide analysis on it. With the implementation, we find that the charging efficiency is improved and the queue length is reduced. Finally, we end this paper in Sect. 5 with conclusions and future works.

2 Model Establishment

In order to count the electric vehicle order data within a specific range, the origin of the order is taken as the geographic information data point, and the data point information includes the longitude and latitude of the point. The format of driving trajectory data set is shown in formula (1).

$$(X, Y) = \{(x_i, y_i) | i = 1, 2, \ldots, t, \ldots, n\} \tag{1}$$

To obtain the coordinate points of the charging station, it is necessary to first determine the number of the charging stations K in a limited area. The location of the charging station is expressed in formula (2).

$$\left(X', Y'\right) = \left\{\left(x_k^{'}, y_k^{'}\right)|k = 1, 2, \ldots, K\right\} \tag{2}$$

For each charging station, in order to meet the charging need of the majority of electric vehicles, it is also necessary to set the value of charging piles in each charging station as formula (3).

$$C = \{C_1, C_2, \ldots, C_K\} \tag{3}$$

All the processes of the model can be concluded in Table 1.

Table 1. Process of model

Input	The urban driving trajectory data set (X, Y)
Output	①The proposed number of charging stations K
	②The proposed locations of charging stations $\left(X', Y'\right)$
	③The proposed values of charging piles in each charging station C

In this paper, the M/M/C queuing theory model is used to model the charging pile design problem based on order distribution.

In the arrival process, there are C charging piles in the charging service system. Customers arrive at the system according to Poisson flow and queue up to receive service on the first come first served basis.

In the service process, the charging time of each vehicle is also different, but the overall service time follows a negative exponential distribution. Each charging pile can only charge one car at the same time, which does not interfere with each other.

We hope to optimize the cost of the model results and satisfy the queuing requirements in the same time. We come up with an evaluation variable in the form of formula (4).

$$U = min\{aK_iC_i + b(L_s)\} \tag{4}$$

Where U is the evaluation variable, a and b are two positive coefficients. In formula (4), the first term is the cost and the second term is queuing requirement. $[KC]$ and $[L_s]$ are dimensionless constants that have been processed separately. The queue length should be as short as possible. Therefore, we introduce some constraints (5) and (6) to our model.

$$\text{For each } K_i,\ C_i = \lim_{L_s \to \infty}[C] + 1 \tag{5}$$

Where $[C]$ is defined as the integer part of C, and they are all integers.

$$C \geq C_i, \quad C, C_i \in N^* \tag{6}$$

3 Brief of Methods and Algorithms

3.1 The Method of K-means Clustering Analysis

Clustering analysis is a method of data mining, which is often used to analyze a large number of sample data. Clustering analysis basing on a certain feature between various sample points (usually the distance between sample points) divided each sample point into different clusters, in order to simplify the sample data and find the same features of the data. There are many clustering algorithms that are widely used. The algorithm used in this paper is K-Means clustering algorithm based on Euclidean distance, namely K-Means clustering algorithm. The Advantage of K-Means clustering algorithm is that the closer the distance of the sample points is, the higher the similarity of the points is.

In K-Means clustering algorithm, the clustering results of a certain sample data set can be obtained through continuous iteration as long as the number of clusters K is given. Therefore, whether an appropriate K value is selected determines the quality of clustering results. In order to select an optimal K value, the sum of the squares of the distance from each sample point after iteration to the center of the cluster (the Sum of the Squares of the Errors, SSE) is often used as the evaluation index. The calculation formula is as (7).

$$SSE = \sum\nolimits_{k=1}^{K} \sum\nolimits_{p \in C_k} \left[dis(p, m_k)\right]^2 \tag{7}$$

Among the formula (7): C_k is the k-th cluster, p is the data point of C_k, m_k is the cluster center of C_k, $dis(p, m_k)$ is the distance between p and m_k.

The smaller the SSE is, the more convergent the cluster is, and the better the clustering result is. Obviously, smaller SSE does not mean better. This is because SSE is zero when K value is taken as the number of the sample points. On this occasion it certainly cannot achieve the purpose of clustering. In order to find the balance between smaller SSE and more reasonable K value, the elbow method is often used to determine the optimal K value. This method is outlined below.

As the value of K is gradually increased to a large enough value, calculate the SSE corresponding to every K value, then draw the broken line graph of SSE changing with K value. Obviously, it should be a line which keeps falling. The K value corresponding to the obvious inflection point can be selected as the optimal number of clusters.

The pseudo-codes of these processes above are given as follows.

Procedure: K-Means clustering algorithm and the elbow method
Input: One coordinate data set $(X, Y) = \{(x_i, y_i)|i = 1,2, \ldots, t, \ldots, n\}$
Output: The count of clusters *K*, the coordinates of final cluster centers (X', Y')
1: Initialize the coordinates of cluster centers
$(X', Y') = \{(x'_k, y'_k)|k = 1,2, \ldots, K\} = \{(x'_1, y'_1), (x'_2, y'_2), \ldots, (x'_K, y'_K)\}$
2: *K* is the count of clusters
3: ***for** K **in range (1, t) do***
4: ***while*** the elements in (X', Y') do not change any more
5: ***for** i **in range (1, n) do***
6: Assign the point (x_i, y_i) to the nearest cluster center (x'_k, y'_k)
7: ***end for***
8: ***for*** each (x'_k, y'_k) ***in*** (X', Y') ***do***
9: Take the average coordinate of all data points that belongs to this cluster center (x'_k, y'_k) as the new cluster center point coordinate (x'_k, y'_k)
10: ***end for***
11: ***end while***
12: Calculate the *SSE*
13: ***end for***
14: Draw the broken line graph of SSE changing with *K* value
15: Select the *K* value corresponding to the obvious inflection point as output

3.2 The Method of M/M/C Queuing Theory

Queuing theory, a branch of operational research, also known as stochastic service system theory, is a mathematical theory and method to study the phenomenon of random aggregation and dispersion of systems and the working process of stochastic service systems. In this paper, we use queuing theory to get the quantity index of service objects, so as to determine and optimize the parameters of related systems, and so as to make the target service system play the best benefits.

The mathematical model of queuing theory established in this paper is M/M/C type. The definition of this kind of model is that the customer flow obeys Poisson distribution of the parameter λ. The service time obeys the exponential distribution of the parameter μ. The number of service desks is constant *C*. And the services of each service desk are independent of each other.

Without considering other relevant constraints, the operation indexes can be obtained from the existing data to facilitate the evaluation of the system. When the expectation of a certain operation index has been obtained, the reasonable range of several important parameters can also be deduced.

After the clustering points are obtained, a mathematical model related to queuing theory can be applied to analyze the number C of charging piles. In the absence of constraints such as cost or achievability, we hope that the number of charging piles of each charging station corresponding to each clustering point can exactly meet the customers' needs, which means that the system service intensity $\rho = \frac{\lambda}{C\mu} \leq 1$. From this formula we'll get a constrain about C, which is $C \geq \frac{\lambda}{\mu}$.

From the mean value of the Poisson distribution $E[N(t)] = \lambda t$, it can be seen that the number of cars arriving per hour is the value of the Poisson index in actual situations. Therefore, we multiply the total number of orders with a proportional coefficient f to get the K we need. In this article, we take f as 1/25.

As mentioned earlier, the average charging time μ is determined to be 1 h as an inherent attribute of the entire service system.

The formula (8) is the specific formula used to calculate the queue length index.

$$L_s = \frac{1}{C!}\frac{(C\rho)^C \rho}{(1-\rho)^2}P_0 + \frac{\lambda}{\mu} \tag{8}$$

where P_0 is calculated as formula (9).

$$P_0 = \left[\left(\sum_{k=0}^{C-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k\right) + \frac{1}{C!}\frac{1}{(1-\rho)}\left(\frac{\lambda}{\mu}\right)^C\right]^{-1} \tag{9}$$

3.3 The Application Processes

The main procedures of the practical application in a specific city are as follows.

Firstly, cutting the longitude and latitude of the city into x segments respectively, or rather, dividing the urban area of the city into a set number of small grids.

Secondly, counting the numbers of electric vehicles' start points in each grid as the outdegree of this grid. The result can be expressed by a x-order matrix.

Thirdly, replacing all electric vehicles' starting points in every grid with the central coordinates of this grid.

Fourthly, selecting the top $y\%$ outdegree of grids as the aforementioned urban driving trajectory data set (X, Y).

Fifthly, using K-Means clustering algorithm and the elbow method to determine the proposed locations of charging stations (X', Y') and the proposed number of charging stations K.

Sixthly, using the M/M/C queuing theory to determine the proposed values of charging piles in each charging station C.

In these procedures, x and y are undermined constants.

4 Practical Application and Results Analysis

It is the city of Chengdu, China that was chosen as the research object in this paper. The number of the longitude and latitude segments x is taken as 100, which means the urban area of Chengdu is divided into 10,000 small grids. The percentage of the top outdegree of grids $y\%$ is taken as 5%.

The data all comes from Didi company, which published in GAIYA plan (https://outreach.didichuxing.com/research/opendata/). In the data source, it concludes more than 6,000,000 electric operating vehicle orders in 30 days in November 2016. The northeast corner of the vehicle operation area is Qilong Temple with geographic coordinates of

Table 2. Data format

Timestamp of departure	Arrival timestamp	Longitude of beginning	Latitude of beginning	Longitude of ending	Latitude of ending
1478091677	1478092890	104.00816	30.70622	104.064147	30.685848
……	……	……	……	……	……

$30.734868^{\circ}N$, $104.180312^{\circ}E$, while the southwest is the Joy Park with coordinates of $30.60068^{\circ}N$, $103.983974^{\circ}E$. The Table 2 describes the data format.

In the actual experiment, we researched the urban driving trajectory data of 30 days in November 2016. In this paper the result on November 2^{nd} is taken as an example. It concludes about 200,000 orders.

The coordinates of the top 5% outdegree of those grids are selected as the urban driving trajectory data set (X, Y).

In the fifth step, the *SSE-K* image obtained by elbow method is shown in the Fig. 1.

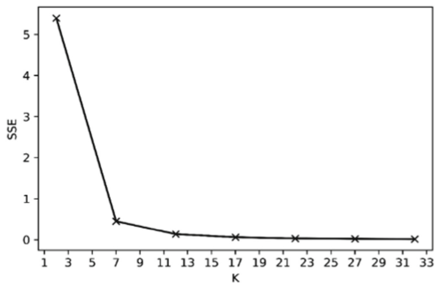

Fig. 1. The elbow method

It can be clearly seen from the figure that $K = 7$ is the obvious inflection point, so according to the elbow method mentioned earlier the proper cluster number K should be approximately equal to 7.

Here we list the results of clustering with K of 7, 9, 11, 13 in Fig. 2.

In the calculation, we calculate the total number of orders contained in a certain cluster center corresponding to each K value for different cluster K values. Since the data and coordinate position corresponding to this point are relatively stable, there is an obvious trend of change, which is convenient for analysis. In the subsequent processing, the data at this point shall prevail. When this analysis method is applied to the processing of other points, it needs to be modified (such as adding a correction coefficient) to meet the actual situation.

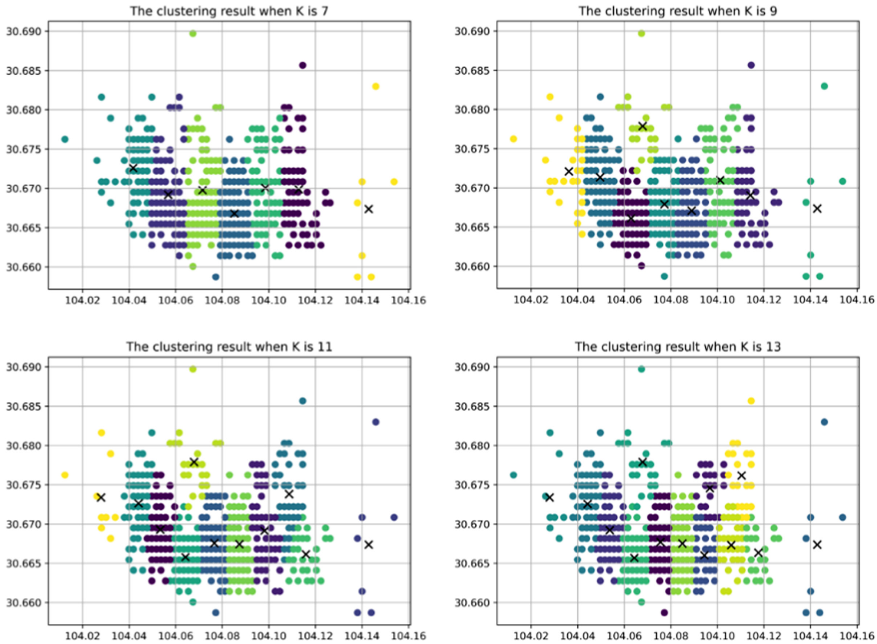

Fig. 2. The results of clustering with K of 7, 9, 11, 13

First, we need to compare the value of the queue length L_s corresponding to each K value when the number of charging piles is different. According to the value of the total number of orders and the principle that the overall average service intensity ρ is as close as possible to 1, we take the value of C from 35 to 51. Because some data cannot correctly reflect the status of the queuing system, we deleted the redundant data and put the data that can visually express the length of the queue in the Table 3.

There are negative numbers in Table 3. The reason is that the function of the queue length is a hyperbolic function, which is close to negative infinity on the left side of the asymptote. If a negative number appears in the table, we can consider that the queue length is close to infinity here. The C value corresponding to the number marked in gray is the value of C_i in formula (5), (6). This value is a limit, which means that when C is smaller than this value, the queue length is closed to infinity.

Use KC as the independent variable and L_s as the dependent variable to draw the picture. The result is shown in Fig. 3.

Table 3. Calculation result

Table 3.1 (K=7, λ=36.5)

C	35	37	39	41	43	45	47	49	51
L_s	4.07	102.51	45.10	39.47	37.72	37.02	36.72	36.59	36.54

Table 3.2 (K=9, λ=30.3)

C	29	31	33	35	37	39	41	43	45
L_s	-0.31	67.37	36.26	32.31	31.08	30.61	30.42	30.35	30.32

Table 3.3 (K=11, λ=24.6)

C	23	25	27	29	31	33	35	37	39
L_s	2.43	80.41	30.12	26.27	25.19	24.82	24.68	24.63	24.61

Table 3.4 (K=13, λ=23.9)

C	23	25	27	29	31	33	35	37	39
L_s	-8.96	40.37	27.26	25.00	24.29	24.04	23.95	23.92	23.90

Table 3.5 (K=15, λ=18.6)

C	17	19	21	23	25	27	29	31	33
L_s	0.99	60.21	22.40	19.64	18.93	18.70	18.63	18.61	18.60

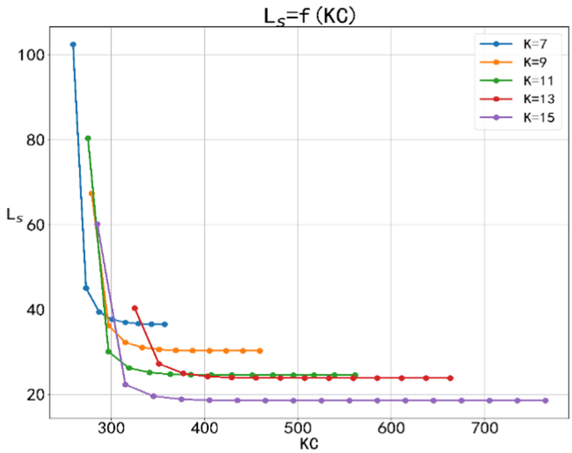

Fig. 3. The function relationship between KC and L_s

In the previous analysis, we got the mathematical relationship between KC and L_s, that is, cost and queue length. On this basis, we hope to construct an evaluation variable that can contain the function of KC and L_s at the same time, and use the function value and the current curve to optimize the whole method. In the formula (4), we take $a = 0.1$ and $b = 0.8$ as examples and draw the image of U in Fig. 4 for optimization analysis.

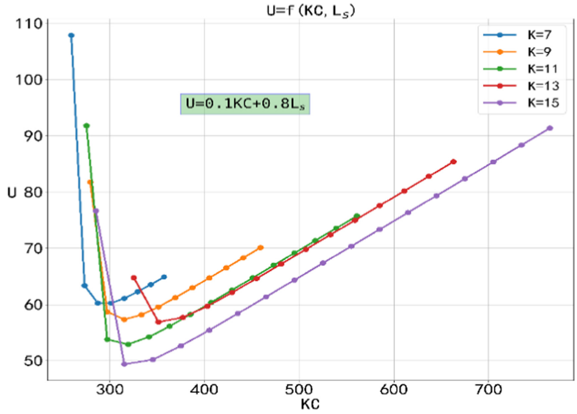

Fig. 4. The function relationship between U and KC, L_s

5 Conclusions

In this paper, we mainly use M/M/C queuing theory and K-Means algorithm to calculate the number of the electric vehicle charging stations and to complete the positioning of electric vehicle charging station. We take the actual social environment as background, and believe that these results can provide some help for the actual construction of charging stations. We use the selected hot points data to simulate the operation of the charging stations. All in all, the contributions of this paper are summarized as follows:

1. To establish stations, we employ K-Means algorithm by real data to analyze customers features of driving behavior/tracing. In this way, we cluster a large number of scattered points into hot spots for analysis.
2. In the modeling, we consider the queue length for each vehicle by queuing theory, so that we can obtain a reasonable number of charging piles to optimize the customer queuing time.

References

1. Zhang, R., Hanaoka, T.: Deployment of electric vehicles in China to meet the carbon neutral target by 2060: provincial disparities in energy systems, CO_2 emissions, and cost effectiveness. Resour. Conser. Recycl. **170**, 105622 (2021). https://doi.org/10.1016/j.resconrec.2021.105622
2. Yang Andrew, W.: A review of evolutionary policy incentives for sustainable development of electric vehicles in China: Strategic implications. Energy Policy **148**, 111983 (2021). https://doi.org/10.1016/j.enpol.2020.111983
3. Schnell, J.L., Peters, D.R., Wong, D.C., Lu, X., Guo, H., Zhang, H., Kinney, P.L., Horton, D.E.: Potential for electric vehicle adoption to mitigate extreme air quality events in China. Earth's Future **9**(2), e2020EF001788 (2021). https://doi.org/10.1029/2020EF001788
4. Hsieh, Y.L., Pan, M.S., Green, W.H.: Transition to electric vehicles in China: Implications for private motorization rate and battery market. Energy Policy, **144** (2020)

5. Zhao, L.: The life cycle environmental rebound effect of battery electric vehicles in China. A provincial level analysis. Appl. Econ. **53**(25), 2888–2904 (2021)
6. Li, J., Liang, M., Cheng, W., Wang, S.: Life cycle cost of conventional, battery electric, and fuel cell electric vehicles considering traffic and environmental policies in China. Int. J. Hydrog. Energy **46**(14), 9553–9566 (2021). https://doi.org/10.1016/j.ijhydene.2020.12.100
7. Graham, J.D., Belton, K.B., Xia, S.: How china beat the US in electric vehicle manufacturing. Iss. Sci. Technol. **37**(2), 72–79 (2021)
8. Zhou, M.: Characterizing the motivational mechanism behind taxi driver's adoption of electric vehicles for living: Insights from China. Transp. Res. Part A **144**, 134–152 (2021)
9. Hua, Z., Li, D., Guo, W.: A deep q-learning network based reinforcement strategy for smart city taxi cruising. In: Zhang, H., Yang, Z., Zhang, Z., Wu, Z., Hao, T. (eds.) NCAA 2021. CCIS, vol. 1449, pp. 59–70. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-5188-5_5
10. Wei, G., Lin, R., Jianping, L., Jiang, W., Wei, C.: The generalized dice similarity measures for probabilistic uncertain linguistic MAGDM and its application to location planning of electric vehicle charging stations. Int. J. Fuzzy Syst. **54**, 1–16 (2021). https://doi.org/10.1007/s40815-021-01084-z
11. Xiao, D., An, S., Cai, H., Wang, J., Cai, H.: An optimization model for electric vehicle charging infrastructure planning considering queuing behavior with finite queue length. J. Energy Stor. **29**, 101317 (2020). https://doi.org/10.1016/j.est.2020.101317
12. Huang, S., Kang, Z., Zenglin, X., Liu, Q.: Robust deep k-means: an effective and simple method for data clustering. Patt. Recog. **117**, 107996 (2021). https://doi.org/10.1016/j.patcog.2021.107996
13. Zhang, X., Yang, F.: Information guiding and sharing enhanced simultaneous heat transfer search and its application to k-means optimization. Appl. Soft Comput. **109**, 107476 (2021). https://doi.org/10.1016/j.asoc.2021.107476