A Study on the Location of Public Transport for General Studies Stops Based on DBSCAN Algorithm

Ziqi Wei*, Liangshan Shao, Weifeng Zhong

Liaoning Technical University, Huludao 125105, China
*156304113@qq.com

Abstract. Traffic congestion is an important problem that restricts the development of society, but traffic congestion does not happen all the time, there is a regularity, auxiliary public transport is to address this characteristic, complementary to the traditional public transport. For the bus system, the selection of the appropriate station location can meet the demand of the nearest passenger drop-off, improving the quality of services and increase the passenger load factor. In this paper, drawing on the improved DBSCAN algorithm, the idea of determining two parameters in the best clustering result of K-means algorithm is proposed to determine another parameter by distance constraints under the condition of considering walking distance constraints. Finally, the feasibility and applicability of the improved DBSCAN algorithm considering distance constraints are verified through the selection of the location of stops a public transport for general studies, which provides a basis for the setting of auxiliary bus stops.

Keywords: Paratransit, Station Selection, DBSCAN Algorithm, K-means Algorithm, Distance Constraints

1 INTRODUCTION

With the development of urbanisation and motorisation of traffic in China, traffic congestion has become an important problem restricting urban development. Therefore, the optimisation of the public transport system has become one of the important strategies to improve the traffic situation. The Outline for the Construction of a Strong Transportation State points out that it is necessary to "promote the transformation of transport development from pursuing speed and scale to paying more attention to quality and efficiency, from the relatively independent development of various modes of transport to paying more attention to the integrated and fused development, and from relying on the traditional factor-driven to paying more attention to the innovation-driven transformation"[1].Paratransit was created to meet the need for diversification of travel modes.

Auxiliary buses[2] are different from conventional buses in that their mode of operation is more diversified, their stops, routes and timetables are more flexible, and they are more economical and environmentally friendly compared to travelling by car. Looking at the results of domestic and international paratransit research, paratransit

V. Vasilev et al. (eds.), *Proceedings of the 2024 5th International Conference on Management Science and Engineering Management (ICMSEM 2024)*, Advances in Economics, Business and Management Research 306,
https://doi.org/10.2991/978-94-6463-570-6_12

can be broadly classified into seven modes of paratransit: dedicated paratransit lines, customised buses, flexible paratransit, dedicated paratransit, demand-responsive paratransit, public service paratransit for special groups such as people with disabilities and other paratransit modes for specific groups of passengers. In the United States, the paratransit system is relatively well-developed, with various modes of operation and a wide range of groups covered. Public transport in Hong Kong pays more attention to integrated development, and supplementary buses play an active role in connecting and interchanging public transport such as the MTR and buses.

Route planning and station selection are the core aspects of Public transport for general studies Bus operation and management. Therefore, the reasonable selection of bus stops is very important. Scientific optimisation of bus stops and better use of public transport resources can reduce the burden of urban traffic and help to solve the social problems of traffic congestion, energy shortage, air pollution and other problems.

2 STUDY ON IMPROVED DBSCAN ALGORITHM CONSIDERING DISTANCE CONSTRAINTS

2.1 Improved K-Means

The K-means algorithm requires the number of clusters K to be determined in advance, and the number of clusters will directly affect the clustering results[3]. In order to obtain an accurate value of K, the improved K-means algorithm uses the elbow rule
to determine the optimal number of clusters[4]. The sum of the squared distance errors between the prime of each cluster and the samples within the cluster is called the degree of aberration, and the greater the magnitude of the change in the degree of aberration, the better the clustering performance.

K-means algorithm：

Step 1: Enter the dataset M and enter the desired class clusters K; Step 2: K data points are randomly selected as initial cluster crowded heart on the dataset M;

Step 3: Label the points in the dataset M into the cluster class of the cluster crowded heart nearest to it;

Step 4: Update the cluster crowded heart, the barycenter of mass of all points in each cluster class is noted as the cluster crowded heart;

Step 5: Repeat steps 3, 4 and 5 until the barycenter of mass remains stable;

Step 6: Calculate and output profile coefficients and error sum of squares.

The sum of error squares is calculated as shown in equation (1).

$$\mathrm{SSE} = \sum_{i=1}^{K} \sum_{P \in C_i} |P - m_i|^2 \tag{1}$$

In the formula. C_i denotes the cluster of class i after clustering. P is a data point in C_i. M_i is the barycenter of mass of C_i and K is the number of clusters.

2.2 Improved DBSCAN Algorithm

The classical density-based clustering method, DBSCAN algorithm, needs to go ahead and determine Eps and MinPts, the values of which affect the clustering effect to some extent. The improved DBSCAN algorithm [5] an adaptive method to estimate some threshold parameters in the DBSCAN algorithm to avoid setting the parameters directly and improve the clustering effect. The improved DBSCAN algorithm [] uses an adaptive method to estimate some threshold parameters in the DBSCAN algorithm, avoiding the direct setting of parameters and improving the clustering effect. Through the improved K-means algorithm, after arriving at a suitable K value, the distance between any two samples is calculated and ranked in its most general cluster (the cluster with the largest number of samples), with the median denoted as Eps, and the number of points whose barycenter of mass of the calculated distance is not greater than the Eps point is denoted as MinPts.

DBSCAN Algorithm

Step 1: Input the dataset M. Label all the samples in the dataset M as unprocessed objects, noting D=M.

Step 2: Select any object P from D. If the number of points in M with distance to point P not greater than the value Eps is not less than the value MinPts, then P is recorded as the core point E. Otherwise, point P is recorded as a noise point and eliminated from D.

Step 3: Find all the P density reachable points in M and record them in the cluster class, exclude the data in from D.

Step 4: Repeat step 2, step 3 and step 4 until all objects P are recorded in cluster class or marked as noise points.

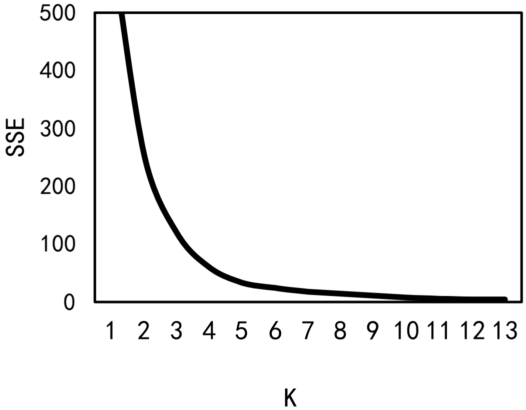
Step 5: Output the cluster class.

3 CASE STUDIES

In this paper, the home addresses of students in a grade in a primary school were investigated. Finally, 292 students were identified as the study population for the selection of sites for the auxiliary pass line, and the students were distributed in 55 residential neighbourhoods in the city, and the locations of common exits in each neighbourhood were investigated and converted to two-dimensional coordinates by the Gaussian projection method[6].

3.1 Elbow Method and Parameter Acquisition

Table 1. elbow method



Determination of the K-value of the K-means algorithm using the elbow rule, the curve of the sum of squares. of the errors as a function of the K-value, As shown in table 2.

As the number of clusters K increases, the sum of squared errors decreases, and there is a clear "inflection point" at K=4, so 4 is the optimal number of clusters for this clustering algorithm. In the case of the best clustering of the K-means algorithm, the distance between any two samples is calculated and ranked from the most general clusters (the clusters with the most samples), and the median is recorded as Eps, and the number of points whose distance barycenter of gravity is not greater than Eps is recorded as MinPts, which results in Eps = 670 and MinPts = 32.

3.2 Output Site Location

For the consideration of students' walking safety and the constraint of students' walking time, the artificial students' walking time to the bus stop is not more than 5 minutes, and through the reference students' walking speed in the city[7] is: the average walking speed across the street of primary school students is 1.15m/s, junior high school students is 1.27m/s, and senior high school students is 1.30m/s, and the distance of the students' walking to the bus stop is not more than 345m, so the parameter Eps=345 is obtained.

Clustering was performed with the modified DBSCN by the identified parameters Eps=345 and MinPts=26 to obtain clustering results excluding noise interference. The location of the barycenter of mass is calculated in each cluster to obtain the site locations as shown in Table 2.

Table 2. Site coordinates

Site	X-axis	Y-axis	Site	X-axis	Y-axis
1	1169.352	445.796	5	-818.911	2647.248
2	-1031.49	675.8607	6	-1821.85	3186.726
3	746.6284	706.877	7	347.4085	3257.473
4	2943.286	799.5493	8	1469.788	6809.631

4 CONCLUSION

This paper establishes a bus stop extraction method based on the improved DBSCAN algorithm on the basis of analysing the distance from passengers to the station. The improved DBSCAN algorithm mainly captures the values of two parameter parameters through the improved K-means algorithm, and the correct clustering results can be effectively obtained without specifying the values of all the parameters artificially in advance. Therefore, with the hint of the above parameter acquisition method, we can determine the other parameter while determining one parameter according to the actual constraints, and then the DBSCAN algorithm to arrive at a reasonable clustering. Finally, an empirical analysis of the location selection of auxiliary public transport school line stops has been carried out through the residential addresses of students of a certain grade in a school. The method proposed in this paper can efficiently and accurately select the alternative locations of bus stops according to the distance constraints from passengers to bus stops in different situations, which provides a basis and a new perspective for the setting of auxiliary bus stops.

REFERENCES

1. Editorial Board of the Chinese Journal of Highways. (2016). Review of Academic Research on Traffic Engineering in China.doi:10.19721/j.cnki.1001-7372.2016.06.001.
2. Lam Ka Wai. (2021-07-01). Foshan Municipal Transport Bureau party building to lead a new chapter of traffic first to promote development. Foshan Daily, 047. doi:10.28234/n.cnki.nfsbr.2021.001907.
3. Pan, Xinyue & Yang, Jiaqi. (2023). Research on classification of interior materials based on improved K-means clustering algorithm. doi:10.26914/c.cnkihy.2023.019457.
4. Zhang, Rui & Zhang, Li-Hong. (2024). Research on Customer Experience Management Optimisation Strategy Based on K-Means Clustering Algorithm. doi:10.20153/j.issn.2096-9759.2024.02.065.
5. Meng, Tao, Xiaoyong Wang & Shengli Hu. (2024). A deep mining method for learning data based on improved genetic algorithm and DBSCAN clustering). 1-7. doi:10.20171/j.cnki.23-1419/n.2024.01.001.
6. Wang, Huiran & Gao, Ziqiang. (2019). Errors in backcalculating the ellipsoidal area of plots using Gaussian projection maps.doi:10.19580/j.cnki.1007-3000.2019.06.023.
7. Wang, Huimin, Huiyi Liu & Canrui Lin(2023). Accessibility analysis of parks and green spaces in Liwan District, Guangzhou based on walking and regular public transport. doi:10.19892/j.cnki.csjz.2023.20.18.