ISSN 2616-5775 Vol. 7, Issue 10: 119-124, DOI: 10.25236/AJCIS.2024.071016

Accident Black Spots Identification of Electric Scooter Based on Self-Adaptive DBSCAN Algorithm

Jiaoxuan Chen[1], Xiaoqiang Zhang[1,2,3,a,*]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, 610031, China*
[2]*National Engineering Laboratory of Application Technology of Integrated Transportation Big Data, Southwest Jiaotong University, Chengdu, 610031, China*
[3]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu, 610031, China*
[a]*xqzhang@swjtu.edu.cn*
[*]*Corresponding author*

Abstract: *In recent years, electric scooters have become one of the emerging modes of short-distance transportation. As the number of people using electric scooters increases, so does the number of their traffic accidents. Therefore, it is extremely important and urgent to pay attention to the safety of electric scooters. In view of the fact that the traditional DBSCAN clustering algorithm needs to determine the Eps and MinPts values by experience, which is not accurate enough, this paper adopts a self-adaptive algorithm to determine the parameters of the DBSCAN algorithm. This algorithm determines the Eps and MinPts values by the distribution of the data itself, so as to improve the accuracy of the algorithm. This paper conducted a case study using data collected from the STATS19 road safety database as an example, and identified six accident blackspots. The clustering results were compared with the MDA-DBSCAN algorithm. The results show that the algorithm has better accuracy and adaptability in black spot identification, and provides more accurate data support and decision basis for urban traffic safety management.*

Keywords: *Electric scooters, DBSCAN, Black spots, Traffic safety*

1. Introduction

Traffic congestion, energy efficiency, and environmental concerns are fueling interest in light-duty electric vehicles. In particular, electric human-powered hybrid vehicles are notable for their low cost, ease of use, and extremely small footprint[1]. Electric scooters have conquered urban areas as a means of personal mobility and compete with other modes of transportation [2] and have been touted as a convenient, inexpensive solution for the last mile and other short trips[3]. The safety of pedestrians and e-scooter riders is a growing concern due to the lag in the development of transportation policies and regulations and the increasing number of e-scooter crashes[4].

Cluster analysis methods have been widely used in road safety research and accident-prone point identification. Cluster analysis methods can divide an unknown data set into several groups or classes with common attributes, and determine the similarity criterion of the division based on the distance between data objects. In the identification of traffic accident-prone points, there are mainly five clustering methods based on division results, grid-based, hierarchy-based, density-based and model-based [5]. Guo Lin[6] et al. established an improved K-means clustering algorithm to eliminate the influence of isolated points on the clustering results to identify and analyze the traffic accident black spots in Yinzhou District, Ningbo City. Lin Nanting[7] et al. created an accident black spot identification model based on time series clustering by taking a square with a side length of 1.5 km as a unit area and 3 months as a time step. Yingzhi Wang[8]et al. proposed a network spatio-temporal kernel density estimate based on traffic accident scenarios by forming accident spatio-temporal sub-segments through road network matching and road network cropping. Kernel density estimation (KDE+) is a current method that helps researchers and road managers in many countries/regions to quickly identify accident-prone locations in traffic networks[9–11]. Mao Che et al. analyzed the regional distribution characteristics of accidents using water traffic accident data from 2006-2010 under the jurisdiction of the Yangtze River Maritime Administration, and extracted 15 blackspot segments of water traffic accidents on the Yangtze River mainline using the DBSCAN algorithm . Geng Chao[12]and others

ISSN 2616-5775 Vol. 7, Issue 10: 119-124, DOI: 10.25236/AJCIS.2024.071016

combined dynamic segmentation and DBSCAN algorithm to identify the black spot sections. Zhang Yunfei[13]et al. proposed a method for identifying highway traffic accident black spot sections based on DBSCAN.

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm, as a density-based clustering method, shows a strong potential for application in the field of data mining and machine learning. However, one of its remarkable properties is also the high sensitivity to the input parameters Eps (neighborhood radius) and MinPts (minimum number of points required to form dense regions). The selection of these two parameters directly affects the quality and effectiveness of clustering, and improper settings may result in clustering results that are too loose and compact, or even fail to accurately recognize the natural structure in the data, especially when facing complex or unevenly distributed datasets, this sensitivity problem is particularly prominent.

In order to solve the uncertainty and inaccuracy of the DBSCAN algorithm due to manual parameter setting, this paper proposes a self-adaptive parameter-based K-Average Nearest Neighbor (K-ANN) algorithm. The algorithm aims to enhance the accuracy of clustering by automatically adjusting the key parameters to adapt to the intrinsic characteristics of different datasets.

2. Algorithmic Principle

2.1 Basic idea of DBSCAN algorithm

Principle of the density-based DBSCAN algorithm: each point in the data set is an object to be analyzed, from which point P is taken arbitrarily, if point P is a core point (the number of object points within the radius Eps of the neighborhood of P is greater than the density threshold MinPts), then point P is searched as a core point to find out the density of all the object points that are reachable from point P, and finally find out the largest set of interconnecting densities and label all object points within the set as the same cluster; if point P is not a core point, no other object points are reachable from point P, then point P is labeled as a noise point. Finally, a maximum set of interconnected densities is found, and all object points within the set are labeled as the same cluster; if point P is not a core point, and no other object points are densely reachable from point P, then point P is labeled as a noise point.

2.2 K-Average Nearest Neighbor (K-ANN) DBSCAN Algorithm[14]

The specific steps for the implementation of the KANN-DBSCAN algorithm are as follows:

Step 1: Calculate the distance distribution matrix of dataset:

$$D_{n\times n}=\{Dist(i,j)\ /\ 1\leq \mathrm{i}\leq n, 1\leq j\leq n\} \tag{1}$$

Where $D_{n\times n}$ is an $n\times n$ real symmetric matrix; n is the number of objects contained in the data set D; $Dist(i,j)$ is the distance from the ith object to the jth object in the data set D.

Step 2: Sort the elements of each row of the distance matrix $D_{n\times n}$ in ascending order, then the distance vector D_0 composed of the elements of the 1st column represents the distance from the object to itself, which is all 0. The elements of the Kth column form the K-nearest-neighbor distance vector for all data points.

Step 3: Average the elements in the vector D_K to obtain the K-averaged nearest neighbor distance $\overline{D_K}$,of the vector D_K and use it as a candidate Eps parameter. For all the values of K, the Eps parameter list $D_{Eps}$is obtained denoted as:

$$D_{Eps}=\{D_K\ /\ 1\leq K\leq n\} \tag{2}$$

Step 4: Generate the MinPts parameter list by mathematical expectation method. For a given list of Eps parameters, find the number of Eps neighboring objects corresponding to each Eps parameter in turn, and calculate the mathematical expectation of the number of Eps neighboring objects of all the objects as the neighborhood density threshold MinPts parameter of the dataset D, denoted as:

$$MinPts=\frac{1}{n}\sum_{i=1}^{n}P_i \tag{3}$$

where Pi is the number of Eps neighbors of the ith object, and n is the total number of objects in the dataset D.

Step 5: Find the candidate Eps parameter set D_{Eps} of dataset D according to the K-mean nearest neighbor method.

Step 6: Select the K-mean nearest neighbor distances corresponding to different K-values ($K = 1,2,\dots,n$), i.e., select the elements in the set D_{Eps} as the candidate Eps parameters and the MinPts parameter obtained by the formula in turn, and then input the DBSCAN algorithm to perform the cluster analysis of the dataset, and get the number of clusters generated under different K-values respectively. When the number of clusters generated is the same for three consecutive times, it is considered that the clustering result tends to be stabilized, and the number of clusters N is recorded as the optimal number of clusters.

Step 7: Continue to perform step 2 until the number of clusters generated is no longer N, and select the maximum K-value corresponding to the number of clusters when the number of clusters is N as the optimal K-value. The K-mean nearest neighbor distance $\overline{D_K}$ corresponding to the optimal K value is the optimal Eps parameter, and the MinPts parameter corresponding to the optimal K value is the optimal MinPts parameter.

3. Case Study

This study uses e-scooter crash data extracted from the STATS19 database, which is reported by UK police forces and provided by the UK Department for Transport. The dataset collects a total of 3214 e-scooter crash data for the years 2020 to 2022, of which a total of 3213 valid messages were used in this study, and the data location information is shown in Figure 1, with the horizontal and vertical coordinates indicating latitude and longitude, respectively. Traffic accidents were collected from all over the UK, and at least one of the vehicles involved in the collisions in the accident sample used in this study was labeled as an electric scooter.

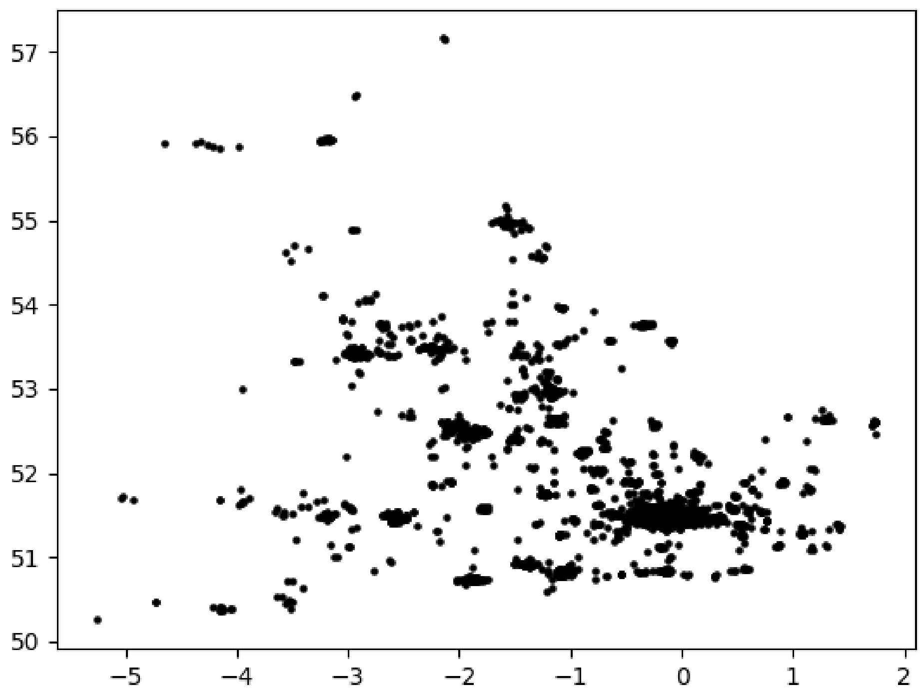

Figure 1: Distribution of accident point data.

3.1 Accident Black Spot Identification and Analysis

The output of all accident data sets using KANN-DBSCAN is shown in Table 1.According to the principle of the algorithm, the same number of clusters for three consecutive times outputs 6, and the maximum k is taken to be 9. At this time, the value of Eps is 0.059999404, and the value of MinPts is 79.38760897; whereas, MinPts is the smallest number of samples in the field of Eps, and it should be taken to be an integer, so the MinPts value is taken as 79.

ISSN 2616-5775 Vol. 7, Issue 10: 119-124, DOI: 10.25236/AJCIS.2024.071016

Table 1: Parameter output table.

Eps	Minpts	ClusterNumber	K
0.01217019	7.196450809	41	1
0.020992808	16.18462017	12	2
0.028065634	25.15161893	10	3
0.033819087	33.49097136	9	4
0.038776313	41.03518057	8	5
0.04386376	49.46419676	7	6
0.050320784	61.03268991	6	7
0.055219318	70.23879203	6	8
0.059999404	79.38760897	6	9
0.064566525	88.35709838	5	10
0.068829674	96.84464508	5	11
0.073470046	105.9797634	4	12

The values of Eps and MinPts are brought into the DBSCAN algorithm to get Fig. 2, black is the noise point and the colored points are the accident black spots, and 6 accident black spots are obtained for a total of 1,639 accidents, and Fig. 3 learns the distribution of the accident black spots in the UK map, and it can be seen that the accident black spots are basically concentrated in the city center area.

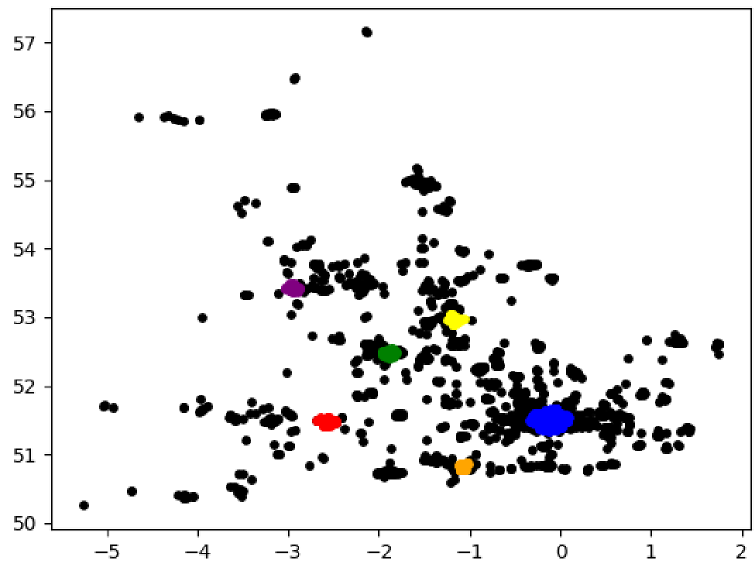

Figure 2: Clustering results of KANN-DBSCAN algorithm.

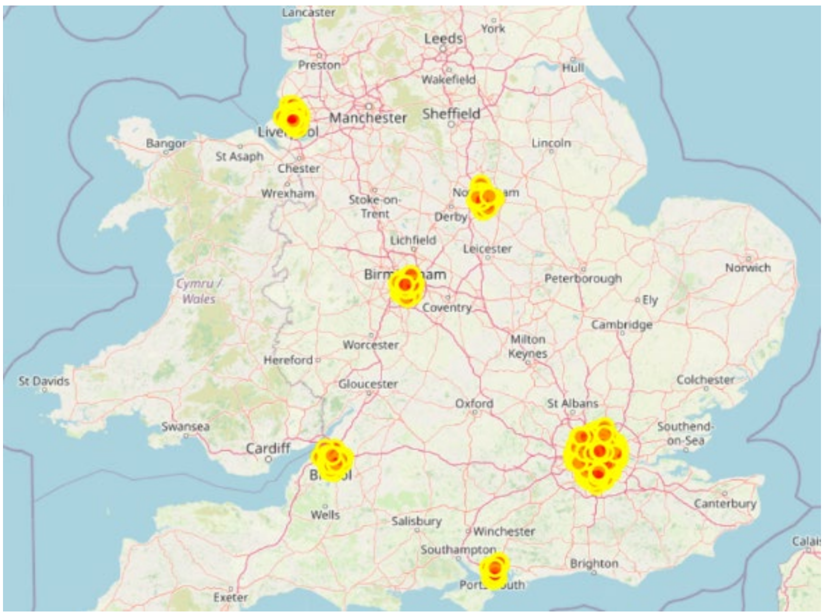

Figure 3: Black spot map distribution.

ISSN 2616-5775 Vol. 7, Issue 10: 119-124, DOI: 10.25236/AJCIS.2024.071016

The MDA-DBSCAN algorithm was used to get the optimal Eps value of 0.08622762181988879, the optimal MinPts value was taken as 92, and the five accidental blackspots totaling 1720 blackspots were obtained by substituting into the DBSCAN algorithm, as shown in Fig. 4.

The KANN-DBSCAN algorithm is compared with the MDA-DBSCAN algorithm, in which the MDA-DBSCAN also selects the optimal parameters, and the number of accidental blackpoints accidents are compared as in Table 2, and some limitations of the MDA-DBSCAN algorithm in the process of recognition can be clearly observed through careful comparison. It can be seen that MDA-DBSCAN ignores the accident blackspot of Portsmouth, moreover, even among the successfully recognized accident blackspots, the range of blackspots defined by the MDA-DBSCAN algorithm shows a certain degree of imprecision. This may be due to the fact that the algorithm adopts a more conservative or less flexible strategy when dealing with data density variations, noise point exclusion, or clustering boundary determination.The KANN-DBSCAN algorithm not only accurately captures all important accident blackspots including Portsmouth, but also defines the range of blackspots in a much more precise way, which effectively reduces false alarms and omissions. It can be seen that KANN-DBSCAN outperforms the MDA-DBSCAN algorithm in the accuracy of accident blackspot identification.

Table 2: Parameter output table.

Accident black spot area	Number of accidents output by KANN-DBSCAN	Number of accidents output by MDA-DBSCAN
London	1002	1133
Portsmouth	80	0
Bristol	196	199
Birmingham	109	127
Nottingham	142	146
Liverpool	108	115

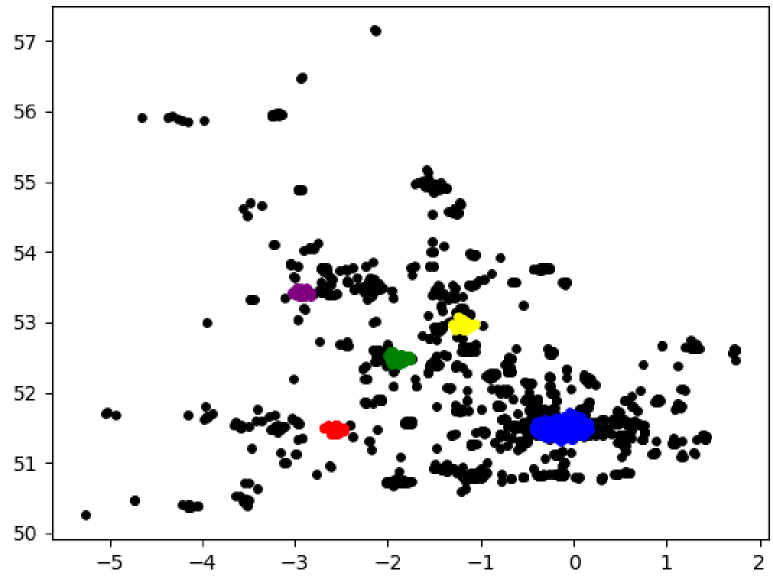

Figure 4: Clustering results of MDA-DBSCAN algorithm.

4. Conclusion

With the continuous expansion of the modern road traffic network and the acceleration of the intelligentization process, the use of electric scooters, as a convenient and environmentally friendly short-distance travel tool, is rapidly expanding globally, especially in large and medium-sized cities, and has become the new darling of the public's daily commuting and recreation. However, with the surge in the use of electric scooters, their safety issues have become more and more prominent. In view of this situation, this paper discusses how to improve the safety of electric scooters through technical means, and adopts the KANN-DBSCAN algorithm, which, compared with the MDA-DBSCAN algorithm, shows a significant advantage in data processing accuracy and efficiency, and successfully identifies and locates six accident black spots of electric scooter accidents from the huge amount of

ISSN 2616-5775 Vol. 7, Issue 10: 119-124, DOI: 10.25236/AJCIS.2024.071016

data.

These accident blackspots are generally located in the city center, which is the center of economic activities and population gathering, with huge daily traffic flow and frequent road congestion. In this context, the electric scooter with its small and flexible, easy to operate characteristics, has become the first choice of many citizens to avoid congestion, fast passage. However, it is also due to its fast speed, poor stability and some users of traffic rules and other factors, resulting in a dense flow of people, complex road conditions in the city center, electric scooter accidents are frequent, to the public safety can not be ignored hidden danger.

In view of the above problems, this paper further analyzes the deep-seated reasons for the high incidence of electric scooter accidents, and points out that in addition to the improvement of the technical level, it is more necessary to start from the policy and management level. On the one hand, the relevant departments should speed up the development and improvement of laws and regulations for electric scooters, clarify their road standards, driving norms and violation of penalties, and enhance the safety awareness and legal literacy of users; on the other hand, they should increase the monitoring efforts on the accident black spots of electric scooters, increase the number of warning signs and safety facilities, and optimize the timing of traffic signals, so as to reduce the likelihood of accidents.

References

[1] Corno M, Savaresi S M. Design and Control of an All-in-the-Wheel Assisted Kick Scooter. IEEE/ASME Transactions on Mechatronics, 2016, 21(4): 1858~1867

[2] Kopplin C S, Brand B M, Reichenberger Y. Consumer acceptance of shared e-scooters for urban and short-distance mobility. Transportation Research Part D: Transport and Environment, 2021, 91: 102680

[3] Sanders R L, Branion-Calles M, Nelson T A. To scoot or not to scoot: Findings from a recent survey about the benefits and barriers of using E-scooters for riders and non-riders. Transportation Research Part A: Policy and Practice, 2020, 139: 217~227

[4] Lee K J, Yun C H, Yun M H. Contextual risk factors in the use of electric kick scooters: An episode sampling inquiry. Safety Science, 2021, 139: 105233

[5] Dong C, Chang N, Dong C, et al. Overview of the identification of traffic accident-prone locations driven by big data. d, 2023, 2(1): 67~76

[6] Guo Ling, Zhou Jibiao, Dong sheng, et al. Analysis of Urban Road Traffic Accidents Based onImproved K-means Algorithm. China J. Highw. Transp., 2018, 31(4): 270~279

[7] Lin Nanting, Hu Lin, Lin Miao, et al. Black spot identification and analysis of traffic accidents based on time series clustering[J].J Changsha Univ Sci Tech(Nat Sci),2023,20(2):45-54.

[8] Wang Yingzhi, Wang Lijun .An Identification Method of Traffic Accident Black Point Based on Street-Network Spatial-Temporal Kernel Density Estimation. Scientia Geographica Sinica ,2019 ,39(8): 1238 -1245.

[9] Bíl M, Andrášik R, Janoška Z. Identification of hazardous road locations of traffic accidents by means of kernel density estimation and cluster significance evaluation. Accident Analysis & Prevention, 2013, 55: 265~273

[10] Bíl M, Andrášik R, Nezval V, et al. Identifying locations along railway networks with the highest tree fall hazard. Applied Geography, 2017, 87: 45~53

[11] Elvik R. A survey of operational definitions of hazardous road locations in some European countries. Accident Analysis & Prevention, 2008, 40(6): 1830~1835

[12] Geng Chao, Peng Yuhua. Identification method of traffic accident black spots based ondynamic segmentation and DBSCAN algorithm.Journal of Chang an University(Natural Science Edition), 2018, 38(5): 131~138

[13] Zhang Yunfei, Zhang Zhexu, Zhu Fangqi. Identification of highway accident black spots based on spatio-temporal density clustering. Bulletin of Surveying and Mapping, 2022(10): 73~79

[14] Li Wenjie, Yan Shiqiang, Jiang ying, et al. Research on Method of Self-Adaptive Determination of DBSCAN Algorithm Parameters. Computer Engineering and Applications, 2019, 55(5): 1-7+148