A Highway Abnormal Event Discrimination and Detection Method

Lei Zhou [1, a], Yi Lin [2, b], Haochen Bai [3, c]

[1] Xi'an High Tech Zone Housing and Urban Rural Development Bureau, China

[2] School of transportation engineering, Chang'an University, Xi'an 71000, China

[3] School of Highway, Chang'an University; CCCC First Highway Consultants. Co., Ltd., China

[a] 914966928@qq.com, [b] 2939045700@qq.com, [c] 37698212@qq.com

Abstract. With the continuous increase in highway traffic volume, abnormal events occur frequently, seriously affecting traffic fluidity and safety. This study aims to develop an efficient abnormal event detection method. Based on simulation data generated by the VISSIM software, the K-means clustering algorithm is applied, with speed, density, and occupancy rate as key feature parameters, aiming to effectively identify abnormal events on highways. The research results show that, after comparing the performance of models using different combinations of feature parameters, the K-means model using speed and occupancy as input features exhibits the best performance, achieving an accuracy of 90.4% and a false negative rate of only 6.47%. The model proposed in this study can effectively identify abnormal events on highways, thereby providing strong support for improving the safety management level of highways.

Keywords: Highway, Blind Spots, Abnormal events, characteristic parameter, k-means clustering.

1. Introduction

Since the 1960s, many scholars abroad have proposed the development of highway event detection systems. In 1961, the New York State government established one of the earliest automatic detection systems [1]. Between 1965 and 1970, the California Department of Transportation introduced the classic California Algorithm [2]. From 1970 to 1975, the Texas Transportation Institute (TTI) developed the Standard Deviation Method [3] for the Houston highways, which is used to determine whether an abnormal event has occurred. Levin [4] proposed the Bayesian Algorithm, similar to the California Algorithm but utilizing conditional probabilities based on occupancy changes to identify congestion states. Thom introduced the McMaster Algorithm [5], which not only distinguishes between congested and non-congested conditions but also between incident-related and recurring congestion. However, as a single-section detection method, it lacks spatial consideration, resulting in lower accuracy and difficulty in parameter calibration.

With advances in artificial intelligence technology, improvements in computational capabilities, and the accumulation of large datasets, some transportation researchers have begun to apply AI techniques to address traffic problems. Zhang et al. [6] utilized Bayesian Networks (BN) to quantitatively model the causal relationships between traffic events and traffic state parameters. Wan Binbin [7] and colleagues proposed a detection method based on Multiple Input Multiple Output (MIMO) fuzzy theory and incremental discriminant analysis. Aliari et al. [8] employed various neural network models, establishing a sliding time window for continuous monitoring of speed sequences within the window. This approach helped reduce false alarms in imbalanced datasets where congestion data are much less prevalent than normal data. Akira Kinoshita [9] used floating car data to compare the probability deviation between "current" and "historical" states, using the degree of deviation to distinguish between recurrent and incidental congestion. Qin Tao [10] combined extracted feature parameters with Backpropagation (BP) neural networks to effectively discriminate between incidental and recurrent local congestion in urban areas. Chen et al. [11] conducted real-time traffic state recognition through traffic software simulation and the Fuzzy C-Means (FCM) algorithm, enhancing the real-time performance of recognition. Du [12] improved the traditional FCM algorithm by using entropy weighting for feature selection and applied the improved algorithm to train a multi-class Support Vector Machine (SVM). Pi et al. [13] implemented traffic congestion event identification using Hard C-Means (HCM) and FCM algorithms, providing a high-precision method for real-time traffic state recognition. Li [14] utilized the FCM algorithm to identify different traffic state indicators, demonstrating the high feasibility of the algorithm.

This study is based on highway section simulation data generated using the VISSIM software. Speed, density, and occupancy are selected as feature parameters, and the K-means clustering algorithm is applied to discern the occurrence of abnormal events. Several evaluation indicators are used to compare models and validate the effectiveness of the algorithm, providing a method for detecting abnormal events on highways.

2. Methodology

2.1. Data Collection

Abnormal events on highways refer to sudden situations or anomalous behaviors that affect normal traffic flow, such as vehicle breakdowns or extreme weather conditions leading to traffic congestion or accidents. In this chapter, we use the VISSIM simulation software to simulate event scenarios, simulating both the occurrence and non-occurrence of abnormal events. The road is set up with six lanes in each direction, and the simulation starts at 10:00 and ends at 12:00, with the abnormal event occurring at 10:30. A total of 100 abnormal events and 20 non-abnormal events under different conditions were simulated. Data collection points a and b were located 50 meters upstream and downstream of the location where the event occurred, respectively. The simulation duration was set to 2 hours, with data collected every 5 minutes. Other simulation parameters were set according to Table 1. After the simulation, the data were obtained from the data collection results option. The simulation parameters for the occurrence of abnormal events are shown in Table 1. An example of the VISSIM simulation and data collection is illustrated in Fig. 1.

Table 1. Parameters for the simulation of abnormal events occurrence

Parameters	Range	Units
Upstream traffic volume	5500, 5900, ..., 4000	veh/h
Number of lanes	3, 2 (one lane closed)	lanes
Proportion of large vehicles	10%, 15%, ..., 30%	%
Length of lane affected by the event	300	m
Speed of small vehicles	80-110	km/h
Speed of large vehicles	70-90	km/h

Timeint	Datacolloc	Dist	Speedavgarith(all)	Speedavgharm(all)	OCCUPRATE
0-300	1	234.41	89.433841	89.121643	31.78%
0-300	2	235.18	86.212498	86.081703	33.16%
300-600	1	583.79	85.541501	85.308392	29.99%
300-600	2	234.69	89.142844	88.991141	32.10%
600-900	1	234.85	86.266673	86.099989	32.40%
600-900	2	584.41	89.056931	88.657762	31.10%
900-1200	1	233.4	79.672388	78.83847	34.60%
900-1200	2	234.19	84.203878	83.923388	30.51%
1200-1500	1	584.45	84.798236	84.259266	25.13%
1200-1500	2	584.76	90.073675	90.000708	28.45%
1500-1800	1	583.47	86.542524	86.067919	30.01%
1500-1800	2	233.23	82.989944	82.652039	29.69%
1800-2100	1	584.76	81.048041	80.951493	30.33%
1800-2100	2	584.76	70.667362	68.734514	31.50%

Figure 1. Example of VISSIM simulation and data acquisition

2.2. Abnormal Events Occurrence Discrimination based on K-means Clustering Algorithm

2.2.1. Algorithm Selection

This study employs a clustering method, which fundamentally assesses the similarity among data objects using multivariate statistical techniques and groups them into distinct clusters. Common clustering algorithms include k-means clustering, density-based clustering, and hierarchical clustering. Given the primary objective of discerning the occurrence of traffic abnormal events, only two clusters are required: normal and abnormal states. K-means clustering is suitable for scenarios with a predetermined number of clusters and is straightforward and computationally efficient. By setting the number of clusters to 2, this method effectively distinguishes between normal and abnormal traffic conditions, enabling the detection of abnormal events.

In cluster analysis, distance measures are typically used to quantify the similarity or dissimilarity between data objects, serving as the basis for grouping. Euclidean distance, due to its intuitive nature and ease of computation, is widely used in practice. Its mathematical expression is given by:

$$d(x,y)=\sqrt{\sum_{i=1}^{p}\left|\left(x_i-y_i\right)\right|^2} \tag{1}$$

2.2.2. Algorithmic Indicators Selection

Given the multidimensional nature of traffic flow on highways, we selected speed, density, and occupancy as key parameters for feature input. Compared to flow, these three indicators—speed, density, and occupancy—more directly reflect the instantaneous state and changing trends of traffic flow, making them more sensitive to abnormal events such as traffic accidents and congestion. Consequently, by leveraging these three parameters, the K-means clustering algorithm can more accurately capture changes in traffic conditions, thereby effectively identifying the occurrence of abnormalevents.

2.2.3. Discrimination Steps Based on the K-means Clustering Algorithm for Abnormal Events

The specific process for discriminating abnormal events occurrence based on the K-means clustering algorithm is as follows:

Step 1: Obtain traffic data for both the non-occurrence and occurrence of an abnormal event through VISSIM simulation.

Step 2: Select density, speed, and occupancy as feature parameters and establish a dataset *D*D.

Step 3: Preprocess the data, and divide the preprocessed dataset containing various sample data into training and testing sets.

Step 4: Choose the number of cluster centers to be 2, and randomly initialize the cluster centers.

Step 5: Follow the basic steps of the K-means clustering algorithm, iteratively updating the cluster centers until the clustering result converges and the cluster centers are obtained.

Step 6: Combining the cluster centers from Step 5 for both the non-occurrence and occurrence of an abnormal event, calculate the Euclidean distance l_1 between the input feature data and the cluster center for the non-occurrence of an abnormal event and the Euclidean distance l_2 between the input feature data and the cluster center for the occurrence of an abnormal event. If $l_1<l_2$ then it is determined that an abnormal event has not occurred; otherwise, it is determined that an abnormal event has occurred, and the discrimination process ends (where the Euclidean distance is calculated according to Equation (1)).

In Step 2 of data preprocessing, normalization was applied to scale the data linearly to a specified range, enhancing the convergence speed and accuracy of the clustering algorithm. The formula for normalization is as follows:

$$x^*=\frac{x-x_{min}}{x_{max}-x_{min}} \tag{2}$$

The cluster centers obtained in Step 5 are in normalized form. To obtain the original values of the cluster centers, an inverse normalization process is required, which is given by the following formula:

$$x=x^{*}(x_{max}-x_{min})+x_{min} \tag{3}$$

Based on the above steps, two-cluster analyses were conducted for each pair of the three feature parameters: speed, density, and occupancy. The clustering results are illustrated in Fig. 2, Fig. 3, and Fig. 4. Using Equation (3), the normalized cluster centers were transformed back to their original scales to obtain the cluster centers for both the occurrence and non-occurrence of abnormal events, as shown in Table 2, Table 3, and Table 4.

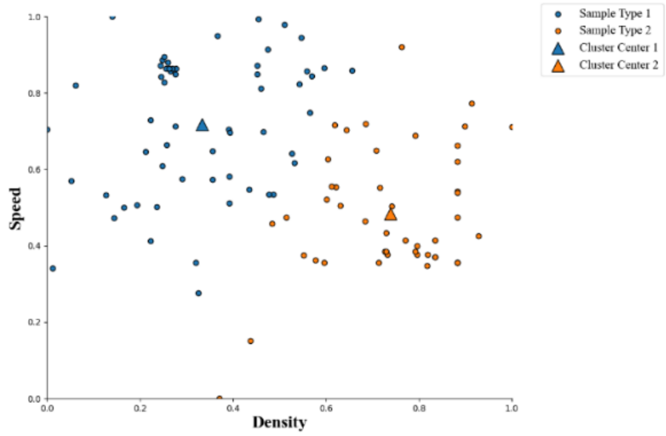

Figure 2. Speed-Density two feature parameters clustering effect diagram

Table 2. Normalized speed-density two-feature parameter clustering centers

Traffic State	Speed (km/h)	Density (veh/km)
Abnormal Events Occurrence	32.4	123.6
Non-Occurrence of Abnormal Events	72.1	50.1

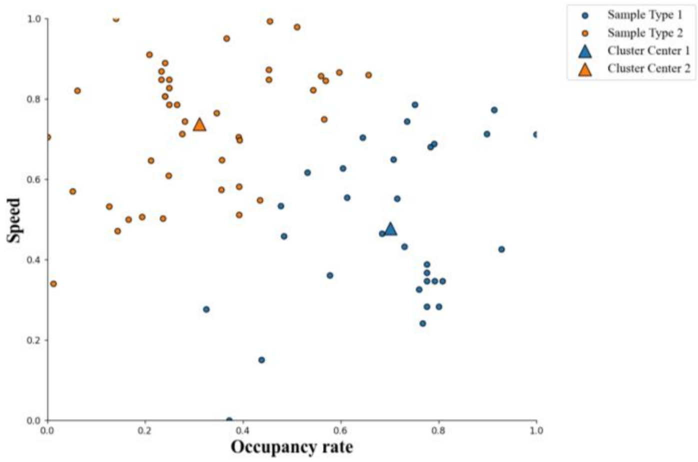

Figure 3. Speed-Occupancy Rate two feature parameters clustering effect diagram

Table 3. Normalized speed-occupancy rate two-feature parameter clustering centers

Traffic State	Speed (km/h)	Occupancy Rate (%)
Abnormal Events Occurrence	31.6	42.7
Non-Occurrence of Abnormal Events	79.8	29.3

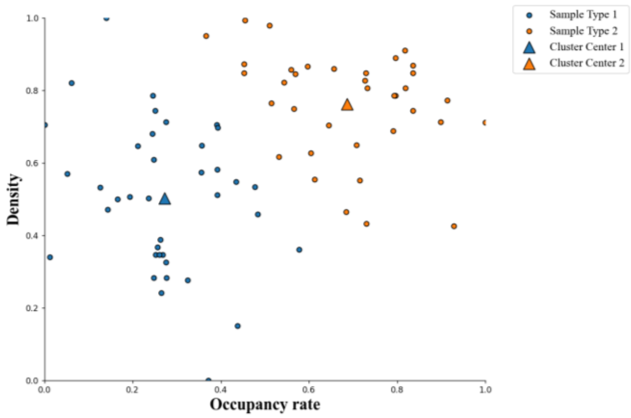

Figure 4. Density-Occupancy Rate two feature parameters clustering effect diagram

Table 4. Normalized density-occupancy rate two-feature parameter clustering centers

Traffic State	Density (veh/km)	Occupancy Rate (%)
Abnormal Events Occurrence	114.2	43.9
Non-Occurrence of Abnormal Events	41.0	31.4

3. Result

An abnormal event occurrence sequence was input, and two-cluster analyses were performed for each pair of the three feature parameters: speed, density, and occupancy. The abnormal event detection was carried out following the specific steps outlined in Section 2.3.3. Fig. 5 presents the discrimination results of the three models, where "1" indicates the discrimination result as an abnormal event occurrence, "2" indicates no abnormal event occurrence, Model 1 is the clustering result with speed and density as feature parameters, Model 2 is the clustering result with speed and occupancy as feature parameters, and Model 3 is the clustering result with density and occupancy as feature parameters. These results were compared with the actual occurrences.

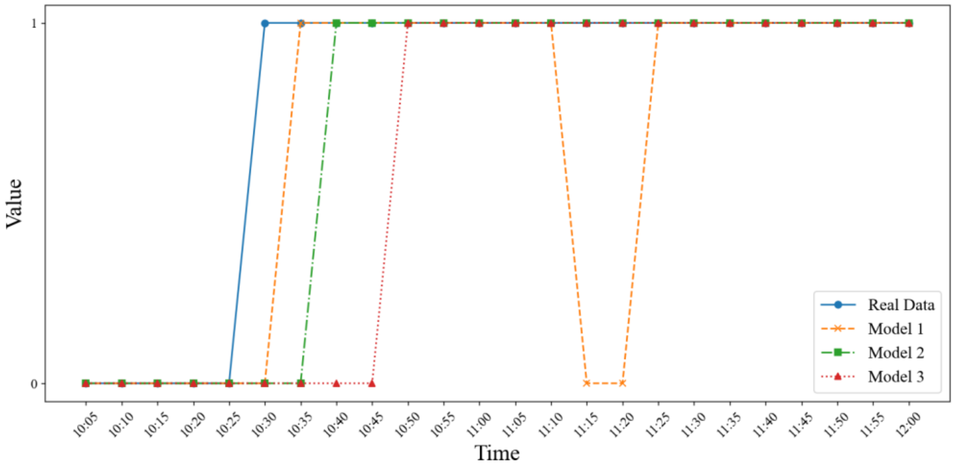

Figure 5. Comparison of three model discriminations

For training, 90 abnormal events that occurred and 18 that did not occur were selected, while for testing, the remaining 10 abnormal events that occurred and 2 that did not occur were used. Cross-validation was employed to assess the model's performance through multiple splits of the training and testing sets, effectively reducing model bias due to uneven data distribution. To evaluate the performance of different models, the precision and missed detection rate of the abnormal event occurrence discrimination algorithms were chosen as performance indicators. The comparison results are presented in Table 5.

Table 5. Comparison of discrimination precison and leakage rate of three models

Evaluation Indicator	Model 1	Model 2	Model 3
Precision	79.5%	90.4%	86.3%
False Negative Rate	7.67%	6.47%	7.91%

From Table 5, considering both precision and missed detection rate, Model 2 outperforms Models 1 and 3. It not only has the highest precision but also the lowest missed detection rate. Therefore, the model obtained by applying the K-means clustering algorithm with both speed and occupancy as feature parameters is most effective for abnormal event detection. This model ensures high accuracy in detecting abnormal events while minimizing the possibility of false alarms.

4. Conclusion

This study has the following limitations in detecting anomalous highway events based on K-means clustering algorithm:

(1) This study is based on simulation data generated using VISSIM, with simulation parameters set to default values typical of general scenarios. However, there may still be deviations from real-world scenarios, and the quality and quantity of the data might contain certain inaccuracies.

(2) The parameter selection in this study primarily relies on subjective experience without detailed analysis.

(3) The K-means algorithm is sensitive to the initial choice of cluster centroids, which can lead to instability in the results. Additionally, the K-means algorithm performs poorly in handling noise and outlier data.

Future research should focus on optimizing data quality, refining parameter selection methods, and improving clustering algorithms to enhance the precision of abnormal event detection.

References

[1] Payne, Howard J, Tignor, Samuel C. Freeway incident-detection algorithms based on decision trees with states [C]. Annual Meeting of the Transportation Research Board.1978:30-37.

[2] Cohen S, Ketselidou Z. Microcomputers in Transportation. A Calibration process for Automatic Incident Detection Algorithms [C]. International Conference on Microcomputers in Transportation. 1993:506-515.

[3] C.L. Dudek, C.J. Messer, and N.B. Nuckles. Incident Detection on Urban Freeways [C]. Transportation Research Record. Washington DC: TRB, 1974.12-24.

[4] Martin P T, Perrin J, Hansen B, et al. Incident detection algorithm evaluation [J]. Prepared for Utah Department of Transportation, 2001.

[5] Srinivasan D, Cheu R L, Poh Y P, et al. Development of an intelligent technique for traffic network incident detection [J]. Engineering Applications of Artificial Intelligence, 2000, 13(3): 311-322.

[6] Zhang K, Taylor M A P. Effective arterial road incident detection: a Bayesian network based algorithm [J]. Transportation Research Part C: Emerging Technologies, 2006, 14(6): 403-417.

[7] Wan Binbin. Research on Highway Traffic Abnormal Event Detection Methods Based on Fuzzy Theory and Incremental Discriminant Method [D]. Kunming University of Science and Technology, 2014.

[8] Aliari S, Sadabadi K F. Automatic detection of major freeway congestion events using wireless traffic sensor data: machine learning approach [J]. Transportation Research Record, 2019, 2673(7): 436-442.

[9] Kinoshita A, Takasu A, Adachi J. Real-time traffic incident detection using probe-car data on the Tokyo Metropolitan Expressway [C], Big Data 2014 IEEE International Conference on IEEE, 2014: 43-45.

[10] Qin Tao. Study on Rapid Recognition Methods for Incidental Local Congestion on Urban Roads [D]. Changsha University of Science & Technology, 2014.

[11] Chen Huiru. Research and Application of Real-Time Traffic State Discrimination Methods for Basic Sections of Expressways [D]. Chang'an University, 2015.

[12] Du Chong. Study on Real-Time Traffic State Discrimination Methods for Expressways [D]. Beijing Jiaotong University, 2017.

[13] Pi Xiaoliang, Wang Zheng, Han Hao, et al. Application Research on Traffic State Classification Method Based on Information Collected by Inductive Loop Detectors [J]. Journal of Highway and Transportation Research and Development, 2006, (04): 115-119.

[14] Li Xiaolu. Traffic State Discrimination Based on Fuzzy C-Means Clustering for Urban Road Traffic [J]. Journal of Transportation Science and Economics, 2016, 18 (04): 32-36+42. 2016.04.009.