Research on Driving Safety Monitoring Device and Urban Traffic Governance Based on Smart Mobility Big Data

Xiaohan Li, Jiale Gao, Jiaze Yuan, Yinghe Wang*

Suihua University, Suihua, 15200, China

*wyh971006@foxmail.com

Abstract. As a vital component of urban public safety management systems, the effectiveness of driving safety monitoring devices depends on accurate identification and real-time warning of traffic operation states. Based on the Smart Mobility Dataset, this study employs K-means clustering (optimal K=9, silhouette score=0.1952) with PCA dimensionality reduction to finely identify urban road network traffic patterns, and utilizes the Apriori algorithm (min_support=0.02, min_confidence=0.4) to extract key driving factors of high-energy-consumption traffic states. Nine typical traffic state patterns were identified, among which Cluster 0 (76.47% occupancy, 288.11 g/km emission), Cluster 1, Cluster 5, and Cluster 6 were classified as high-emission/high-energy risk clusters, accounting for 42.5% of the total samples. Association rule mining revealed that the combination of foggy weather and red traffic lights (support=0.03, confidence=0.44, lift=1.15) serves as a significant precursor signal for high-energy-consumption traffic states. The findings provide data-driven decision-making support for optimizing deployment strategies of driving safety monitoring devices, traffic emergency management by urban public administration departments, and risk warning mechanisms within public service networks.

Keywords: Driving Safety Monitoring Device; Smart Mobility; Urban Public Administration; Cluster Analysis; Association Rule Mining

1 Introduction

With the rapid advancement of urbanization in China, the total number of motor vehicles continues to grow, and driving safety monitoring devices (DMS), as the core sensing equipment of intelligent transportation systems, have a direct bearing on the level of urban public safety. However, existing DMS research has mostly concentrated on hardware design and signal processing, with insufficient attention paid to location optimization of device deployment and context-adaptability of warning thresholds [1-2]. In fact, the same monitoring device exhibits significantly different warning performance under varying traffic conditions, highlighting the urgent need for fine-grained identification of road network operation states based on traffic big data.

© The Author(s) 2026

B. P. Siba et al. (eds.), *Proceedings of the 2026 5th International Conference on Public Service, Economic Management and Sustainable Development (PESD 2026)*, Advances in Economics, Business and Management Research 411,
https://doi.org/10.2991/978-94-6239-825-2_11

In the field of traffic state identification, existing studies have utilized intersection license plate recognition data to classify network residual supply capacity using congestion indicators and the K-means clustering algorithm [3]; other scholars have employed K-means and DBSCAN clustering methods to identify the operation states of urban expressways, and further used BP neural networks for state assessment [4]. In terms of traffic accident risk mining, the Apriori association rule algorithm has been applied to freeway accident risk identification [5] and to cause analysis of traffic accidents in urban-rural junctions [6], confirming the effectiveness of this algorithm in extracting multi-dimensional association rules among risk factors. However, few studies have systematically linked traffic state identification results with deployment strategies of driving safety monitoring devices, nor have they explored differentiated setting of warning thresholds from the perspective of urban public administration.

Based on the Smart Mobility Dataset, this study employs K-means clustering and Apriori association rule mining to systematically analyze urban road network traffic patterns. The innovations of this paper are threefold: Introducing cluster analysis and association rule mining into deployment optimization research of driving safety monitoring devices, filling the methodological gap in this field from the perspective of public administration; Identifying 9 typical traffic state patterns and quantifying emission and energy risk levels for each pattern, providing data support for differentiated deployment of monitoring devices; Discovering the strong association rule "Fog∩Red Light→High Energy Consumption" as a precursor signal for dynamic threshold adjustment mechanisms of driving safety monitoring devices.

2 Data and Methods

2.1 Data Source and Description

This study employs the Smart Mobility Dataset, which integrates multi-source traffic information from Cyber-Physical Networks (CPN) and Social Networks (SN), comprising 5,000 valid records collected in March 2024. The dataset covers 15 fields, the spatiotemporal coverage spans typical urban road networks and is representative, including:

(1) Temporal and Spatial Dimensions: Timestamp, Latitude, Longitude.
(2) Traffic Flow Characteristics: Vehicle_Count, Traffic_Speed, Road_Occupancy.
(3) Environmental and State Characteristics: Traffic_Light_State, Weather_Condition, Accident_Report, Traffic_Condition.
(4) Public Service-related Characteristics: Sentiment_Score, Ride_Sharing_Demand, Parking_Availability.
(5) Performance Characteristics: Emission_Levels, Energy_Consumption

2.2 Feature Engineering

To enhance analytical effectiveness, the following temporal features were extracted from the Timestamp field in table 1:

Table 1. Feature Engineering Summary

Feature	Description
Hour	Hour of the day (0–23), capturing intra-day periodicity
DayOfWeek	Day of the week (0–6), capturing weekly periodicity
IsWeekend	Binary variable (1=weekend, 0=weekday)

The final feature set for clustering is: Vehicle_Count, Traffic_Speed, Road_Occupancy, Hour, IsWeekend. All numerical features were standardized using StandardScaler to eliminate scale effects.

2.3 Analytical Methods

2.3.1 K-means Clustering.

The K-means algorithm partitions data by minimizing the sum of squared distances from samples to cluster centers (SSE). This study employs the Elbow Method (SSE) in conjunction with the Silhouette Score to determine the optimal K value. The silhouette score is calculated as:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{1}$$

where $a(i)$ is the average distance from sample i to other samples in the same cluster, and $b(i)$ is the average distance from sample i to the nearest other cluster. The silhouette score ranges from -1 to 1, with higher values indicating better clustering. Considering K=2 as overly coarse, this study selects the K with the highest silhouette score within the range of K=3 to K=10.

2.3.2 PCA Dimensionality Reduction Visualization and.

Principal Component Analysis (PCA) is employed to reduce the high-dimensional clustering results to two dimensions for visualization, facilitating intuitive understanding of traffic state distribution patterns by public administration departments.

2.3.3 Apriori Association Rule Mining.

The Apriori algorithm is used to discover frequent patterns and association rules among traffic features. Key metrics include Support Confidence and Lift, This study sets minimum support at 0.02, minimum confidence at 0.4, and maximum itemset length at 4.

3 Results and Analysis

3.1 Optimal Cluster Number Determination

Evaluation results for K=3 to K=10 are as follows in table 1, K=9 achieves the highest silhouette score (0.1952) and is determined as the optimal number of clusters. The SSE

curve continues to decline with increasing K, with a noticeable inflection point at K=9, validating the rationality of K=9 in Fig 1.

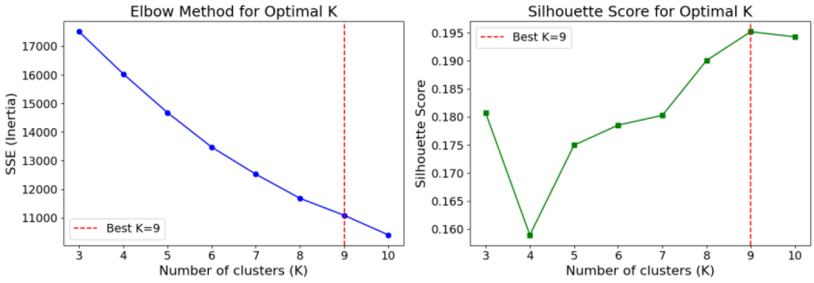


Fig. 1. Table captions should be placed above the tables

3.2 Clustering Results and Interpretation of Typical Traffic Patterns

The sample distribution and center characteristics for each of the 9 clusters are as follows in table 2, Clusters 0, 1, 5, and 6 are high-emission clusters (emission levels above the overall mean of 272.17 g/km), while Clusters 1, 3, 5, and 7 are high-energy-consumption clusters (energy consumption above the overall mean of 17.34 L/h). Overlapping clusters are 1 and 5 in table 2.

Table 2. Table captions should be placed above the tables.

Cluster	n	Vehicle_Count	Speed (km/h)	Occupancy (%)	Hour	Weekend	Emission (g/km)	Energy (L/h)
0	420	168.64	41.74	76.47	4.75	Yes	288.11	17.06
1	604	78.46	59.38	55.39	16.34	No	274.00	17.51
2	692	211.91	56.83	68.89	5.90	No	270.53	17.10
3	652	105.08	34.20	33.76	5.14	No	264.85	17.68
4	410	134.46	45.21	76.20	18.03	Yes	267.94	17.18
5	626	225.02	40.66	36.06	16.76	No	272.85	17.40
6	468	170.82	21.82	35.74	12.86	Yes	279.38	17.04
7	699	139.75	21.55	78.54	12.93	No	268.01	17.57
8	429	145.94	61.22	30.42	10.41	Yes	269.79	17.32

Based on the feature combinations of the 9 clusters, the following typical traffic patterns can be summarized:

Pattern A: High Occupancy/Early Morning Type (Cluster 0): Occupancy of 76.47%, occurring in early morning around 4:45 on weekends, with moderately high vehicle volume. This pattern reflects special travel patterns on weekend mornings, with the highest emission level (288.11 g/km). Driving safety monitoring devices should increase monitoring density on expressways during such periods.

Pattern B: Low Vehicle Volume/High Speed/Afternoon Type (Cluster 1):

Only 78.46 vehicles but speed reaches 59.38 km/h, occurring at 16:34 on weekdays. This pattern represents free-flow conditions on uncongested sections, yet energy consumption is the highest, indicating energy efficiency issues under high-speed conditions.

Pattern C: High Vehicle Volume/High Speed/Early Morning Type (Cluster 2): 211.91 vehicles at 56.83 km/h, occurring around 5:90 on weekdays. This pattern represents concentrated traffic flow on urban arterials in the pre-rush-hour period.

Pattern D: Low Speed/Night/Weekend Type (Clusters 6 and 7): Speeds of only 21.82-21.55 km/h, occurring at night, with occupancy rates of 35.74% and 78.54% respectively. This pattern corresponds to off-peak nighttime periods with persistent local congestion, representing low-speed high-risk scenarios requiring attention from driving safety monitoring devices.

3.3 Association Rule Mining Results

In table 3, For high-energy-consumption association rules, 10 effective strong association rules were identified (confidence ≥0.43, lift ≥1.12), ranked by lift:

Table 3. Table captions should be placed above the tables.

Rank	Antecedents	Consequents	Support	Confidence	Lift
1	Fog ∩Red Light	High Energy	0.03	0.44	1.15
2	Fog ∩ Low Occupancy	High Energy	0.03	0.44	1.15
3	Rain ∩ High Emission	High Energy	0.02	0.44	1.15
4	Fog ∩ Low Speed	High Energy	0.03	0.44	1.15
5	Rain ∩ Red Light	High Energy	0.04	0.44	1.13
6	Rain ∩ Very Low Speed	High Energy	0.02	0.44	1.13
7	Fog ∩ Evening	High Energy	0.02	0.44	1.13
8	Red Light ∩ Medium Occupancy	High Energy	0.05	0.44	1.13
9	High Speed ∩ Evening	High Energy	0.02	0.43	1.13
10	Very Low Speed ∩ Morning	High Energy	0.02	0.43	1.12

The combination of foggy weather and red traffic lights is the strongest signal for predicting high-energy-consumption traffic states (lift=1.15). This finding has significant reference value for setting warning thresholds of driving safety monitoring devices-under foggy conditions with red lights, monitoring devices should lower the threshold for risk warnings.

No strong association rules were found for high emissions, indicating that emission levels are influenced by multiple interacting factors and cannot be effectively predicted through simple rules, requiring real-time monitoring.

4 Conclusions

Based on the Smart Mobility Dataset, this study employs K-means clustering (optimal K=9, silhouette score=0.1952) and Apriori association rule mining to systematically identify traffic states and analyze risks in urban road networks. Nine typical traffic patterns were identified, among which Cluster 0 (76.47% occupancy, 288.11 g/km emission), Cluster 1, Cluster 5, and Cluster 6 were classified as high-emission/high-energy risk clusters, accounting for 42.5% of total samples. Association rule mining confirmed that "Fog ∩ Red Light → High Energy Consumption" is the strongest risk signal (lift=1.15), which can serve as a direct input factor for dynamically adjusting warning thresholds of driving safety monitoring devices under adverse weather conditions. The findings provide data-driven decision-making support for urban public administration departments in optimizing deployment of driving safety monitoring devices and improving traffic risk warning mechanisms.

This study has the following limitations: the data spans only one month, failing to cover seasonal traffic pattern variations; the impact of new energy vehicle penetration on emission factors has not been considered; clustering features do not include infrastructure variables such as road class and number of lanes. Future research can extend to multi-season data collection cycles, incorporate new energy vehicle penetration parameters, and integrate road network structure features to build a more generalizable decision-making model for deployment of driving safety monitoring devices, providing more systematic theoretical support for the intelligent transformation of urban public administration.

Acknowledgments

This work was supported by the 2026 University-level College Students' Innovation and Entrepreneurship Training Program of Suihua University (Project Title: "Intelligent Eye In-vehicle DMS: A Driving Safety Monitoring Device Fusing Facial Features and Vehicle Conditions").

References

1. Tim J, Darren S, Michaele V, et al. Cars and distraction: How to address the limits of Driver Monitoring Systems and improve safety benefits using evidence from German young drivers[J]. Technology in Society, 2021, 66: 101628.
2. Ming S, Ronggui Z, Chengwu J. Analysis of HAZMAT truck driver fatigue and distracted driving with warning-based data and association rules mining[J]. Journal of Traffic and Transportation Engineering (English Edition), 2023, 10(1): 132-142.
3. Li Maosheng, Li Hangcong. Network traffic state classification method based on intersection license plate recognition data[J]. Journal of Harbin Institute of Technology, 2023, 55(11): 82-90.
4. Yu Lelan, Shao Zixuan, Xu Cheng. Research on identification and evaluation method of urban expressway traffic operation situation[J]. Police Technology, 2024(2): 81-85.

5. Li Meiling, Li Zihui, Chen Xuehui, et al. Traffic accident risk identification on expressways based on association rules[J]. Journal of Shandong Jianzhu University, 2024, 39(3): 99-106.
6. Wu Biao, Wang Xingyu, Liu Tuo, et al. Cause identification of traffic accidents in urban-rural junction based on association analysis[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2022, 46(6): 948-952.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

