

Research Article

Research on Abnormal Behavior Recognition of Buses Based on Improved Support Vector Machine

Yong Pan^{1,2}, Xiangmo Zhao,¹ Zhigang Xu,¹ Junwei Li,² Yifei Li,² and Li Liu²

¹School of Information Engineering, Chang'an University, Shaanxi, Xi'an 710064, China

²CHECC Data Co., LTD, Beijing 100089, China

Correspondence should be addressed to Yong Pan; 54820517@qq.com

Received 6 April 2021; Accepted 3 July 2021; Published 27 July 2021

Academic Editor: Qi-Zhou Hu

Copyright © 2021 Yong Pan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The abnormal driving behavior of buses brings about greater security threat. How to effectively identify the abnormal driving behavior of buses has become one of the problems of cracking public transportation safety. This paper constructs an abnormal behavior recognition model of buses based on improved support vector machine. Through the verification, the model has a high recognition rate, which provides an important means for further improving the safety of public transportation operations.

1. Introduction

The abnormal driving behavior of buses not only seriously affects the comfort and safety of passengers in the vehicle but also brings about great security risks to the surrounding vehicles. Therefore, through the research of abnormal driving behavior of buses, analyzing the temporal and spatial distribution characteristics of various behaviors, and establishing an effective mathematical model according to the characteristics of the corresponding driving data analysis, in this paper, through the training and identification of driving data, it can effectively identify the abnormal driving behavior of the bus and promptly remind the drivers of the bus and surrounding vehicles, so as to reduce the occurrence of traffic accidents and ensure the safety of road traffic.

The sudden acceleration and sudden braking behaviors of buses are the most common abnormal driving behavior [1]. In the relevant laws and regulations of the transportation industry, sudden acceleration and sudden braking behaviors are not clearly illegal. Therefore, the existing literature has less research on the sudden acceleration and sudden braking behavior. Literature [2, 3] proposed that the sudden acceleration and sudden braking behaviors of the vehicle belong to the abnormal longitudinal driving behavior, and its main data features are shown in the sudden change of

instantaneous longitudinal velocity and acceleration. Literature [4] studies the abnormal left-turning and right-turning behavior of buses, which is mainly reflected in the turning action at a relatively high speed when the vehicle is running, which results in the vehicle's posture to tilt to the left or right. Literature [5] investigates the abnormal left and right lateral driving behavior of buses, which is mainly reflected in the transformation of the vehicle's acceleration and roll angle. Literature [6, 7] extract features of vehicle speed information, including vehicle speed, GPS information, and triaxial acceleration information, so as to describe changes in vehicle position and driving status during operation.

This paper first introduces support vector machine algorithm; second, it builds an abnormal behavior recognition model of buses based on the improved support vector machine on this basis; finally, the model is verified, analyzed, and summarized by an example.

2. Support Vector Machine

SVM (support vector machine) is a machine learning algorithm for classification proposed by Corinna Cortes and Vapnik in 1995. Because it has shown many unique advantages in solving small sample and nonlinear problems, it

has been popular for more than ten years, and its classification ability can even surpass neural networks on some problems, and it is widely used in the classification and recognition of data samples. The principle of SVM is to find a hyperplane, in which as many data points as possible are correctly distributed on both sides of the hyperplane [8, 9]. According to the complexity of samples and the difficulty of classification, SVM can be considered as three types: linear separable SVM, approximately linear SVM, and nonlinear SVM.

2.1. Linear Separable Support Vector Machine. As shown in Figure 1, a simple sample set $\{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$ is constructed, and two kinds of sample data sets that need to be classified are represented by two kinds of graphs. The classified plane is set as H , and the selected hyperplanes are set as $H1$ and $H2$. In addition, $H1$ and $H2$ must be parallel to H and have the closest distance to the two types of sample data sets. When the classified plane H can allow the maximum distance between $H1$ and $H2$, the best classification hyperplane H is calculated at this time.

The purpose of the calculation of SVM is to make the value of all sample points from the classification hyperplane greater than or equal to a value, which means that all sample points are located at their own type of support vector. This principle can be expressed by the following formula:

$$\max \gamma = \frac{y(w^T x + b)}{\|w\|_2} \quad (1)$$

$$\text{s.t. } y_i(w^T x_i + b) = \gamma'^{(i)} \geq \gamma', \quad (i = 1, 2, \dots, m).$$

In general, simplify the operation and set γ' as 1, so that the optimization function can be defined as

$$\max \frac{1}{\|w\|_2} \quad (2)$$

$$\text{s.t. } y_i(w^T x_i + b) \geq 1, \quad (i = 1, 2, \dots, m).$$

The meaning of formula (2) is to make the objective function $1/\|w\|_2$ get the maximum value within the specified target of $y_i(w^T x_i + b) \geq 1, (i = 1, 2, \dots, m)$. In addition, because the maximum value of objective function $1/\|w\|_2$ is equal to the minimum value of $(1/2)\|w\|_2^2$, the optimization function of SVM is equivalent to

$$\min \frac{1}{2}\|w\|_2^2 \quad (3)$$

$$\text{s.t. } y_i(w^T x_i + b) \geq 1, \quad (i = 1, 2, \dots, m).$$

Because the shape of the objective function $(1/2)\|w\|_2^2$ is a convex function and its specification requires an accurate representation to be an affine function, it can be obtained from the principle of convex optimization that the optimization requirement can be changed into an unspecified optimization function according to Lagrange function. To be precise, the optimization function is transformed into

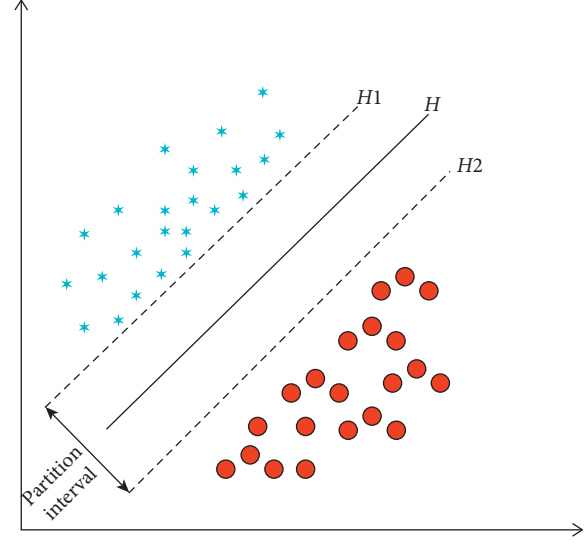


FIGURE 1: Linear separable support vector machine model.

$$L(w, b, \alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i + b) - 1] \quad (4)$$

$$\text{s.t. } \alpha_i \geq 0.$$

Due to the introduction of Lagrange multiplier, the optimization goal of SVM has become

$$\min_{w, b} \max_{\alpha_i \geq 0} L(w, b, \alpha). \quad (5)$$

It can be found from formula (5) that if we want to solve the optimization objective, we can first find the minimum value of the optimization function for W and B and then continue to find the maximum value of Lagrange multiplier α . For the minimum value of W and B , namely, $\min_{w, b} L(w, b, \alpha)$, this extreme value can be obtained by spherical partial derivatives of W and B , respectively:

$$\frac{\partial L}{\partial w} = 0 \implies w = \sum_{i=1}^m \alpha_i y_i x_i, \quad (6)$$

$$\frac{\partial L}{\partial b} = 0 \implies \sum_{i=1}^m \alpha_i y_i = 0.$$

From formula (6), the relationship between w and α can be obtained, so the parameters can be substituted into $L(w, b, \alpha)$ to eliminate w , and $\psi(\alpha)$ is defined as a function of α :

$$\psi(\alpha) = \min_{w, b} L(w, b, \alpha) = \frac{1}{2}\|w\|_2^2 - \sum_{i=1}^m \alpha_i [y_i(w^T x_i + b) - 1]. \quad (7)$$

Through the definition of norm and the operation of vector, W can be eliminated and the functional relationship between $\psi(\alpha)$ and α can be obtained:

$$\psi(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^m \alpha_i \alpha_j y_i y_j x_i^T x_j. \quad (8)$$

It can be seen from formula (8) that the optimization function $\psi(\alpha)$ only has α vectors as parameters, so as long as we can maximize $\psi(\alpha)$, we can find the corresponding α at this time, and then w, b can be obtained. The mathematical expression for maximizing $\psi(\alpha)$ is as follows:

$$\begin{aligned} \max_{\alpha} & -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^m \alpha_i \\ \text{s.t.} & \sum_{i=1}^m \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, m. \end{aligned} \quad (9)$$

Formula (9) is the problem of the extreme value of the quadratic function under the requirements of the inequality, and its optimal solution is set as $\alpha^* = (\alpha_1^*, \dots, \alpha_n^*)$. The initial optimal solution of the problem is w^* and b^* , and $w^* = \sum_{i=1}^n \alpha_i^* y_i x_i$ is required at the same time. In addition, because α^* is a Lagrange multiplier under the original optimization requirement, according to Kuhn–Tucker's requirement,

$$\alpha_i^* (y_i [(w \cdot x_i) + b^*] - 1) = 0, \quad i = 1, \dots, n. \quad (10)$$

Because most of samples α_i^* are equal to 0 and the only condition that makes it not equal to 0 is to make the samples (support vectors) on both sides of the optimal classification surface equal to each other, the mathematical expression of maximizing $\psi(\alpha)$ is brought into Kuhn–Tucker condition to get

$$b^* = y_j - \sum_{i=1}^n y_i \alpha_i^* (x_i x_j). \quad (11)$$

Therefore, the optimal hyperplane of the linear discriminant function in the two-dimensional space can be obtained by the following formula:

$$f(x) = \text{sgn}\{w^* x + b^*\} = \text{sgn}\left\{\sum_{i=1}^n \alpha_i^* y_i (x_i x) + b^*\right\}. \quad (12)$$

Because the value of the nonsupport vector is 0, formula (12) requires the calculation of the support vector.

2.2. Approximate Linear Support Vector Machine. As shown in Figure 2, there are some complex sample points that make the former linearly separable problem change into an approximate linearly separable problem. Therefore, the complex sample points can be regarded as noise points, which are similar to the behavior of turning left and changing lanes in the process of identifying abnormal driving behaviors of buses, and the degree of similarity is relatively high, which will lead to little difference between data samples when reflected in the data.

The existence of noise points will give rise to the data not to be completely linearly separable, so it is necessary to

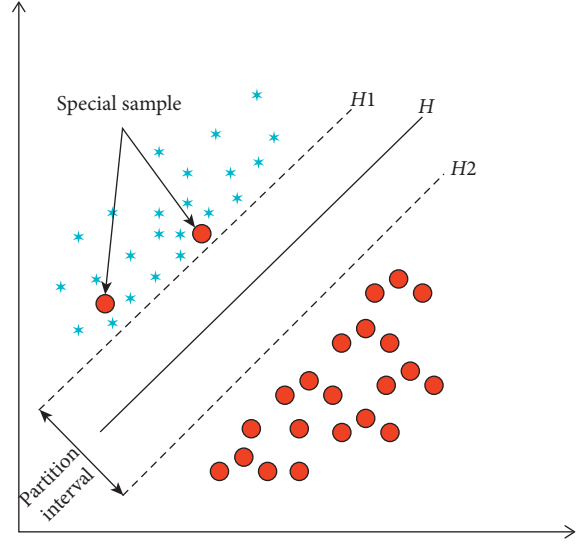


FIGURE 2: Principle of approximate linear support vector machine.

transform the problem to be approximately linearly separable. The relaxation variable β should be added to make the classification interval less than 1; then the optimal hyperplane problem can be transformed into

$$\max \frac{1}{\|w\|_2} \quad (13)$$

$$\text{s.t. } y_i (w^T x_i + b) \geq 1 - \beta, \quad (i = 1, 2, \dots, m).$$

The constraint condition of adding β can be transformed into

$$\min \frac{1}{2} \|w\|_2^2 + C \sum_i \beta_i \quad (14)$$

$$\text{s.t. } y_i [(w \cdot x_i) + b] \geq 1 - \beta_i, \beta_i > 0 \quad i = 1, 2, \dots, n.$$

If the difference between the representative nonlinear sample point and the set of the same categories is set as relaxation variable β , the distance will become larger and larger as the difference increases; and the missing difference between the outlier sample points and the objective function will increase with the increase of the penalty factor C . Among them, the penalty factor C will be obtained in the program of continuous optimization of all parameters. The first step is to specify a value to match the classifier, and then the test sample will be experimented. If the result is poor, change C and repeat the above process. This problem still belongs to the optimization problem, and its handle is the same as the previous optimization problems.

2.3. Linear Indivisible Support Vector Machine. In the above linear separable support vector machines and approximate linear separable support vector machines, their data dimensions are not high, so the data set can be classified by using a low-dimensional hyperplane. However, the abnormal bus driving behavior data set contains a variety of

collected data and statistically calculated moral data, and the dimension of the data is significantly higher than that of the two-dimensional. There may not be a good hyperplane to separate the data set, so it is necessary to map the original data space to a high-dimensional space. If there is a hyperplane that can classify data in a high-dimensional space, it can solve the problem of identifying and confirming abnormal bus driving behavior.

The process of transforming the initial space R^d into the high-dimensional space H is set as $x \rightarrow \Phi(x)$; then the data set $\{(x_1, y_1), \dots, (x_n, y_n)\}$ in the initial space will be transformed into $\{(\Phi(x_1), y_1), \dots, (\Phi(x_n), y_n)\}$, so the problem of obtaining hyperplane is equivalent to

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\Phi(x_i) \Phi(x_j)), \quad C \geq \alpha_i \geq 0, \quad i = 1, \dots, m \\ \text{s.t.} \quad & \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned} \quad (15)$$

The equivalent classification function is converted into

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i (\Phi(x_i) \Phi(x)) + b^* \right\}. \quad (16)$$

Suppose that ϕ is an H mapping from low-dimensional input space χ (a subset or discrete set of Euclidean spaces) to the high-dimensional Hilbert space. Then if there is a function $K(x, z)$, for any $x, z \in \chi$, there is

$$K(x, z) = \phi(x_i) \cdot \phi(x_j). \quad (17)$$

Then, $K(x, z)$ is called kernel function. The advantage of this function is that it can transform features from low dimension to high dimension, and this kind of function can be calculated by relying on low dimensionality, and then the real classification results (after the inner product process) are displayed in high dimensionality. This approach can eliminate the difficulty of calculation in high-dimensional space and completely eliminate the difficulty of inseparable linearity of support vector machine. The commonly used kernel functions are shown in Table 1.

3. Abnormal Behavior Recognition Model of Buses Based on Improved Support Vector Machine Algorithm

Different abnormal driving behaviors of buses have different levels of complexity. As an example, changes in data characteristics of bus rapid acceleration behavior are mainly reflected in changes in vehicle speed and acceleration, and the duration of the change is short, mostly completed in two seconds. However, the abnormal lane-changing behavior of buses may be embodied as continuous lane-changing, and data feature changes are embodied in multidimensional data such as speed, X-axis acceleration, and Y-axis acceleration roll angle, which lasts for a long time, up to more than ten seconds. This leads to the inconsistency of the complexity and quantity of different types of samples when collecting

samples of abnormal bus driving behaviors. When the SVM algorithm classifies data with a large number of samples, the classification error is small; and when classifying data with fewer data samples, the error will be larger. Moreover, if there is a large amount of the same sample data, it will be regarded as new sample data to be calculated again, and the difficulty of calculation will increase greatly, which will lead to a significant increase in the time consumption of the algorithm. Therefore, this research adopts an improved support vector machine algorithm, weighted support vector machine algorithm. Compared with the traditional support vector machine algorithm, this algorithm increases the weight of data samples, reduces the impact of category differences, and speeds up the speed and accuracy of algorithm training.

The optimization problem of weighted support vector machine algorithm is described as

$$\min_{w,b,\xi_i} = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^m s_i d_i \xi_i^2, \quad \xi_i \geq 0, \quad i = 1, \dots, m \quad (18)$$

$$\text{s.t. } y_i (\langle W, X_i + b \rangle) \geq 1 - \xi_i.$$

In formula (18), s_i represents the weight added to each sample, and d_i represents the repetition factor of data points, so the term $s_i d_i \xi_i$ is the error loss of the misclassification of the data. The coefficient s_i is a constant between 0 and 1, which can be assigned different values according to the importance of the data samples. The coefficient s_i tends to 1 for the more important samples, and the coefficient s_i tends to 0 for the samples that need to be discarded, so as to overcome the impact of the algorithm on the imbalance of the number of samples in different categories.

Similar to the general support vector machine algorithm, the weighted support vector machine algorithm also applies the Lagrange multiplier method to obtain the best optimization plane:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\Phi(x_i) \Phi(x_j)), \\ & C s_i d_i \geq \alpha_i \geq 0, \quad i = 1, \dots, m \end{aligned} \quad (19)$$

$$\text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0.$$

The corresponding decision function is

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i (\Phi(x_i) \Phi(x)) + b^* \right\}. \quad (20)$$

Through the comparison of the process of solving the optimal hyperplane of the support vector machine algorithm, it is found that although the results of the two calculation methods are similar, the limitation of weighted support vector machine algorithm requires that the maximum value of α_i be changed to the multiplier of C and $s_i d_i$.

Since the support vector machine algorithm is essentially a two-category algorithm but the problem of abnormal

TABLE 1: Common kernel functions.

Name	Expression	Parameter
Polynomial kernel	$K(x_i, x_j) = x_i^T x_j$	$d \geq 1$ is polynomial degree
Gaussian kernel	$K(x_i, x_j) = (x_i^T x_j)^d$	$\sigma > 0$ is the bandwidth of Gaussian kernel
Linear kernel	$K(x_i, x_j) = \exp(-(\ x_i - x_j\ ^2 / 2\sigma^2))$	$\sigma > 0$
Laplace kernel	$K(x_i, x_j) = \exp(-(\ x_i - x_j\ / \sigma))$	$\sigma > 0$
Sigmoid kernel	$K(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta)$	$\beta > 0, \theta < 0$, \tanh is a hyperbolic tangent function

driving behavior recognition of bus is a multicategory problem, in this research, the support vector machine algorithm and binary tree algorithm are combined to convert the multicategory problem into a series of two-category problems, so as to realize the recognition of various abnormal driving behaviors. As shown in Figure 3, in the binary tree algorithm, firstly, the vehicle motion is divided into longitudinal motion and lateral motion according to the direction of the change of the vehicle operating state. In the longitudinal motion, the vehicle motion is divided into reverse and forward according to the forward direction of the vehicle; in the forward motion, the vehicle motion is divided into constant speed and variable speed according to the change of the vehicle speed; in variable speed motion, it is divided into acceleration motion and braking according to the change of acceleration. In the lateral movement, it is divided into leftward movement and rightward movement according to the lateral movement direction of the vehicle and divided into left turn and left tilt and right turn and right tilt according to the change of vehicle posture.

4. Model Validation

In the experimental verification, the data acquisition device integration module (composed of GPS, gyroscope, three-axis accelerometer, magnetometer, and other sensors) is fixed at the center of mass of the bus, and the measurement coordinate system of the sensor is matched with the running direction of the vehicle, so that the positive direction of the accelerometer and gyroscope is consistent with the heading direction of the vehicle. After the module is installed, the parameters of each sensor need to be calibrated to reduce the impact of data error.

In the data collection link, the bus is driven to continuously perform operations such as starting, going straight, turning, changing lanes, and braking, as well as collecting the various sensors' data for the vehicle's acceleration, deceleration, left turn, right turn, left lane change, and right lane change. Through the collection of data, a large number of data samples of various abnormal driving behaviors were obtained. In this research, 200 sets of various abnormal driving behavior data were intercepted as the training data set, and each data segment was marked as a certain kind of abnormal driving in the training data set. For the label of the behavior, 100 groups are used as verification data, the verification data does not need to be labeled, and it is a data fragment directly intercepted from the collected raw data.

After using LIBSVM, the computing power of support vector calculation method is increased. LIBSVM is a flexible and efficient support vector machine software package

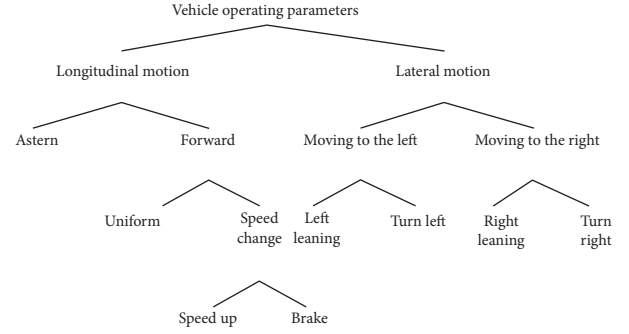


FIGURE 3: Binary tree algorithm.

developed by National Taiwan University. It provides executable files and source code, which can solve various support vector machine algorithms and algorithm expansion problem. 200 sets of training data are used to train the built support vector machine algorithm to stabilize the various parameters of the algorithm, and then 100 sets of verification data are used to verify the improved support vector machine algorithm to test the recognition rate of the algorithm. During the verification process, the receiver operating characteristic (ROC) curve was drawn as shown in Figure 4.

By analyzing the ROC curve graphs of nine different behaviors, it can be found that each curve can reach the top quickly, and the area under the curve is much higher than the area under the diagonal. It shows that the improved support vector machine algorithm has a better ability to discriminate abnormal bus driving behavior. Among them, the real experimental results are presented in Table 2.

Through the analysis of the experimental results, it can be concluded that the improved support vector machine algorithm has high recognition rate for driving behaviors with long duration and large data variation (such as reversing, acceleration, uniform speed, and deceleration) and poor recognition rate for driving behaviors with short duration and no obvious data variation (such as left leaning and right leaning).

5. Summary

This paper selects the data that best reflects the abnormal driving behavior as the research data and then studies the support vector machine algorithm in the machine learning algorithm, analyzes the structure of the algorithm, and improves the support vector machine algorithm according to the characteristics of the abnormal driving behavior of public transportation vehicles. Finally, a bus abnormal driving behavior data set containing 200 sets of training data

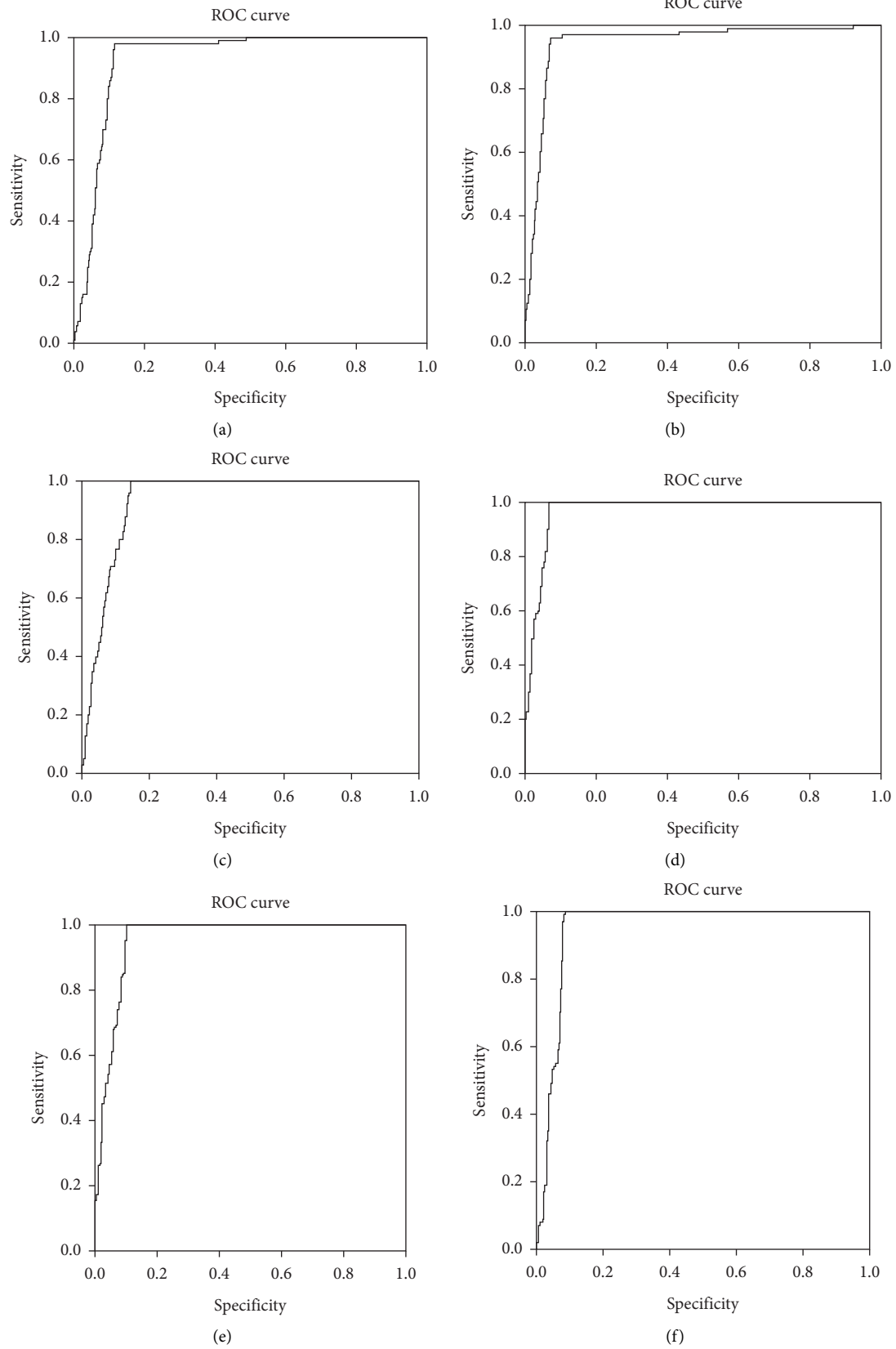


FIGURE 4: Continued.

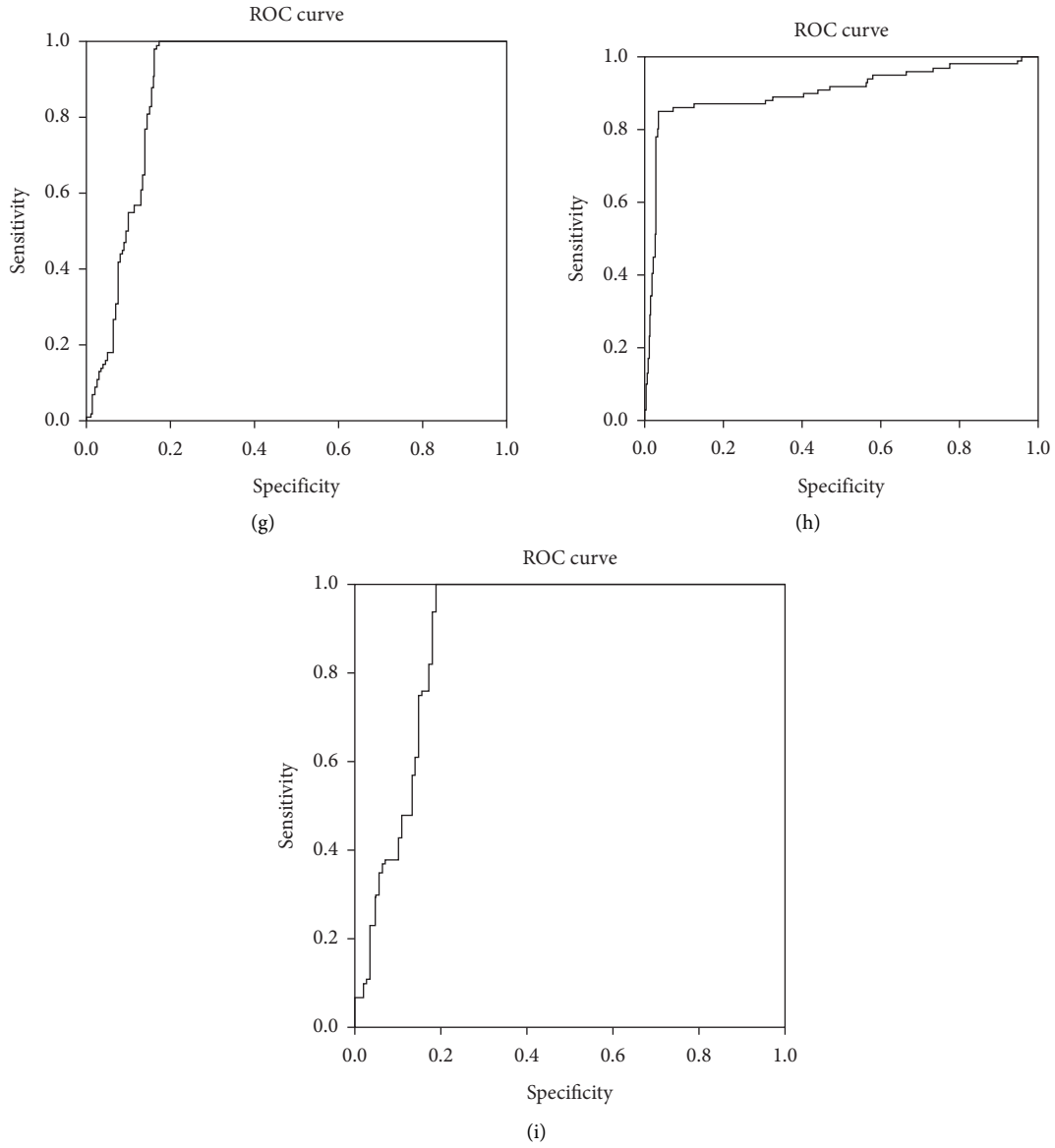


FIGURE 4: The ROC curves are verified by algorithm. (a) Reverse moment. (b) Acceleration moment. (c) Uniform moment. (d) Deceleration moment. (e) Left turn moment. (f) Right turn moment. (g) Brake moment. (h) Left leaning moment. (i) Right leaning moment.

TABLE 2: Bus driving behavior identification results.

Vehicle behavior moment	Total number of samples	Number of error samples	Recognition rate (%)	Results of area under curve (AUC)
Reversing	100	15	85	0.929
Accelerating	100	11	89	0.946
Uniform speed	100	14	86	0.935
Slow down	100	8	92	0.970
Turn left	100	12	88	0.956
Turn right	100	13	87	0.950
Braking	100	22	78	0.901
Left leaning	100	19	81	0.904
Right leaning	100	21	79	0.894

and 100 sets of verification data was established, and the improved support vector machine algorithm was used to train and recognize the data set. The experimental verification results prove that the improved support vector machine algorithm has high recognition rate for driving behaviors with long duration and large data variation (such as reversing, acceleration, uniform speed, and deceleration), which will help to enhance the safety supervision of bus drivers and improve the safety management level of public transport.

Data Availability

The data are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] C. Miyajima, H. Ukai, A. Naito et al., "Driver risk evaluation based on acceleration, deceleration, and steering behavior," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1829–1832, IEEE, Prague, Czech, May 2011.
- [2] S. R. G. Christopoulos, S. Kanarachos, and A. Chroneos, "Learning driver braking behavior using smartphones, neural networks and the sliding correlation coefficient: road anomaly case study," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 65–74, 2018.
- [3] D. Kang and C. Wang, "Acceleration characteristics analysis of medium and small bus," *Vehicle Utility Technology*, vol. 2, pp. 94–97, 2016.
- [4] A. Ma, J. Mao, and J. Shen, "Research progress in driver steering behavior modeling," *Journal of Automotive Engineering*, vol. 3, no. 6, pp. 391–399, 2013.
- [5] Z. Ouyang, J. Niu, and M. Guizani, "Improved vehicle steering pattern recognition by using selected sensor data," *IEEE Transactions on Mobile Computing*, vol. 17, no. 6, pp. 1383–1396, 2018.
- [6] X. Xia, Z. Wang, H. Zhang, and W. Xiong, "GPS real-time attitude measurement system based on MEMS inertial devices," *Sensors and Microsystems*, vol. 29, no. 8, pp. 141–144, 2010.
- [7] B. Zou, H. Zhang, and J. Jiang, "Multisensor information fusion improved extended Kalman filter attitude determination," *Computer Application Research*, vol. 31, no. 4, pp. 1035–1042, 2014.
- [8] L. F. LengQiangkui and Y. Qin, "A multi-class support vector machine classification algorithm based on hybrid binary tree structure," *Computer Science*, vol. 45, no. 5, 2018.
- [9] H. Zhang, D. Qian, and C. Shao, "Driver distraction discriminant support vector machine model optimization algorithm," *Transportation System Engineering and Information*, vol. 18, no. 1, pp. 127–132, 2018.