Safety Science 118 (2019) 607–616

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Research on black spot identification of safety in urban traffic accidents based on machine learning method

Zhanyong Fan[a,c], Chun Liu[a], Dongjian Cai[b,c,*], Shun Yue[c]

[a] College of Surveying and Geo-information, Tongji University, Shanghai 200092, China
[b] School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China
[c] Suzhou Industrial Park Surveying Mapping and Geoinformation Co., Ltd., Suzhou 215027, China

ARTICLE INFO

Keywords:
Machine-learning
Support vector machine
Deep neural network
Black spot identification

ABSTRACT

With the rapid development of economy and urbanization, the number of urban motor vehicles keeps increasing. Urban travel is more convenient, but the traffic safety problems are increasingly prominent. Traffic accident data include not only time and place, but also people, roads, vehicles and the surrounding environment. Traffic accident black spot is the spatial location of traffic accident concentrated distribution. Most of the traditional traffic accident black spot identification only considers time and space factors, ignoring other factors. Based on the traffic accident data of Suzhou Industrial Park, this paper makes a fusion analysis of the multi-source influencing factors involved in traffic accident black spot. According to the structured association characteristics of urban traffic accident big data, a support vector machine method based on maximizing the classification interval is used to train the complex model and optimal learning of accident black spots in the study area. The accuracy of black spot identification is improved. At the same time, aiming at the rapid growth of traffic accident multi-source data, a black point identification algorithm based on deep neural network is proposed. The deep neural network of relevant data category information is established to verify the model's ability to identify accident black spots. A feature-based black spot identification method based on depth neural network is proposed. Furthermore, a dynamic adaptive machine learning architecture is built.

1. Introduction

With the rapid development of China's economy, urbanization is accelerating, travel volume has increased significantly, and the demand for transportation has grown steadily. To meet the travel demand, there has been a rapid increase in vehicle ownership, resulting in a sharp increase in traffic-accident data. With continuous technological development in computer networks, geographic information processing, information mining, spatial analysis and expression, and information systems, many enterprises, universities, and research institutes are analyzing and researching traffic-accident data (Saccomanno et al., 2001). There are many random factors in traffic accidents, but frequent traffic accidents of a similar type somewhere on a road indicate that one should consider the road characteristics and internal laws. These are usually called accident black spots (Shen et al., 2003).

Chi-Hung EvrlydWn of the United States proposed an analysis method based on accident statistics to accurately identify road black spots (Oppe, 1991). Jiang Yan used the density-based spatial clustering of applications with the noise (DBSCAN) algorithm to identify the black spots of an accident and developed a nuclear-density identification method (Shao, 2008). Wang Hongyao et al. introduced the density-based clustering (DENCLUE) algorithm, which implements arbitrary-length clustering to identify traffic-accident black points (Wang et al., 2013). Based on the ArcGIS platform, Ma Zhu used the equivalent accident method to realize an automatic search function for accident black spots (Ma et al., 2014). Wang Hai used the buffer method and the optimized kernel density clustering method to determine black spots for accident occurrence points, road sections, and regions (Debrabant et al., 2018). The identification method was used to carry out the example calculation and search location display on the GIS platform (Black, 2012), and a complete system for the identification and cause analysis of road traffic accidents. The black-point algorithm (Shao and Cai, 2009) for urban traffic accidents also includes quality-control, accident-frequency, accident-rate, equivalent total-accident-number, safety-factor, critical-rate, and integrated-matrix methods (Elvik, 2002).

Most research on black-spot identification is based on the spatial discreteness of traffic accidents, ignoring other factors in their occurrence (Mcguigan, 1981). We must integrate the multi-source features of

* Corresponding author.
E-mail address: caidj163@163.com (D. Cai).

https://doi.org/10.1016/j.ssci.2019.05.039
Received 19 April 2019; Received in revised form 17 May 2019; Accepted 24 May 2019
Available online 08 June 2019
0925-7535/

traffic accidents and accurately identify the location of black spots in traffic accidents (Sharif et al., 2014). The key influencing factors provide scientific support for the decision-making of traffic-management departments, which has become the research focus of accident black-spot identification (Chen et al., 2011; Almjewail and Almjewail, 2018).

Machine-learning aims to learn a large number of training sets composed of non-object samples, and use the learned template or classifier for object detection (Robinson et al., 2010). It enhances performance as experience increases when processing problems. Data-based machine-learning is an important aspect of modern intelligent technology. It studies laws based on observational data and uses them to predict future or unobservable data. The support vector machine is a new machine-learning method developed from statistics (Oskoei and Hu, 2008). It has many unique advantages in solving small-sample, nonlinear, and high-dimensional machine-learning problems. The support vector machine combines the techniques of statistical learning, machine-learning, and neural networks based on the principle of structural risk minimization instead of empirical risk minimization. Therefore, based on the characteristics of the support vector machine, this paper uses machine-learning to self-learn multi-source data from traffic accidents and apply it to Suzhou Industrial Park, enabling us to quickly and accurately obtain the distribution of black spots in the study area.

2. Support vector machine black-point identification

2.1. Support vector machine principles

The support vector machine proposed by Cherkassky (2002) is a class of methods based on maximizing the classification interval. It first extracts the support vectors located on the class boundaries, and uses them to construct the optimal classification hyperplane, which can minimize the probability that data points are misclassified (Chang and Lin, 2011). The known conditions and objectives of the support vector machine are as follows.

Known conditions: a training set given a sample point,

$$S = \{(x_1, y_1), (x_2, y_2), \cdots, (x_i, y_i), \cdots, (x_{l-1}, y_{l-1}), (x_l, y_l)\} \tag{1}$$

Among them, $x_i \in R^n (i = 1, 2, \cdots, l)$is the input to the sample, $y_i = \{-1, 1\}$is the class label corresponding to the samplex_i, and $l$is the data of the training sample. The goal of SVM training is to classify the samples into two categories while making the probability of misclassification as small as possible (Fauvel et al., 2008). The classification principle of support vector machine is shown in Fig. 1.

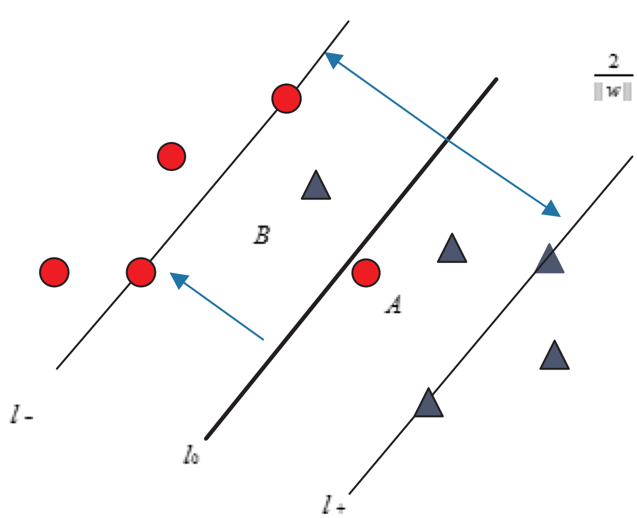

Fig. 1. Scenario of support vector machine processing line inseparable classification problem.

As shown in Fig. 1, the triangles and circles respectively represent the two types of samples, A and B, to be classified.l_+, l_- represent the boundary hyper planes of the two classes (corresponding to the straight line where the sample is two-dimensional), and $l_0$is the decision hyper plane. The following equations are satisfied:

$$l_-: wx + b = -1 \tag{2}$$

$$l_0: wx + b = 0 \tag{3}$$

$$l_+: wx + b = 1 \tag{4}$$

The objective function is

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} \xi_i \tag{5}$$

$$s.\ t. \quad y_i (wx_i + b) \geqslant 1 - \xi_i \quad i = 1, 2, \cdots, l \tag{6}$$

$$\xi_i \geqslant 0 \quad i = 1, 2, \cdots, l \tag{7}$$

where w is the normal vector of the decision hyper plane, $b$is a constant, $\xi = (\xi_1, \cdots, \xi_l)^T$ is a relaxation coefficient vector, and $C$is a penalty factor. This paper separates sample points A and B by determining hyper plane l_0.

2.2. Black-spot identification process based on support vector machine

The algorithm flowchart is shown in Fig. 2.

Black-point identification steps based on SVM:

(1) Preprocess traffic accident data: digitize and standardize;
(2) Divide the traffic accident data into training data and test data;
(3) Construct SVM parameters based on the training data;
(4) Solve for parameters according to the SMO algorithm;
(5) Find the interface parameters;
(6) Input the test data, and detect the identification accuracy of the interface parameters obtained by the training sample;
(7) Input all accident data into the parameters of the decomposition surface to determine whether it is black spots;
(8) The classification result is output, and the black-spot identification ends.

2.3. Black-spot identification analysis results based on SVM

This paper takes Suzhou Industrial Park in Jiangsu Province as a research area. Suzhou Industrial Park is located in the eastern part of Suzhou, and it has an administrative area of 278 square kilometers. It is a demonstration zone for cooperation between China and Singapore. It is the first comprehensive experimental area for open innovation in China. With the rapid development of the economy of Suzhou Industrial Park, car ownership continues to climb. By the end of 2017, Suzhou Industrial Park registered 288,000 vehicles, an increase of 16.6% compared with 2016, and private car ownership was 237,000. Suzhou Industrial Park is a typical example of a new urban area with rapid traffic development. The traffic accidents coincident with its development are suitable for analysis and research. We selected 6391 accident records from September 2016 to March 2018 as an analysis sample, and used the above model to realize the black-point prediction analysis and evaluation of traffic accidents. The original sample mainly included the five aspects of human, vehicle, road, weather, and spatial location. We first studied the accident distribution within the sample area by statistical analysis of spatial nuclear density, as shown in Fig. 3.

Based on Fig. 3, we counted the average accident nuclear density on each natural road section to obtain the spatial distribution of accident occurrence frequency. We defined a road segment with a high accident occurrence frequency as a “black segment.” We matched the spatial coordinates of an accident point with the road segment to determine whether the accident point occurred in the label of “black segment” and

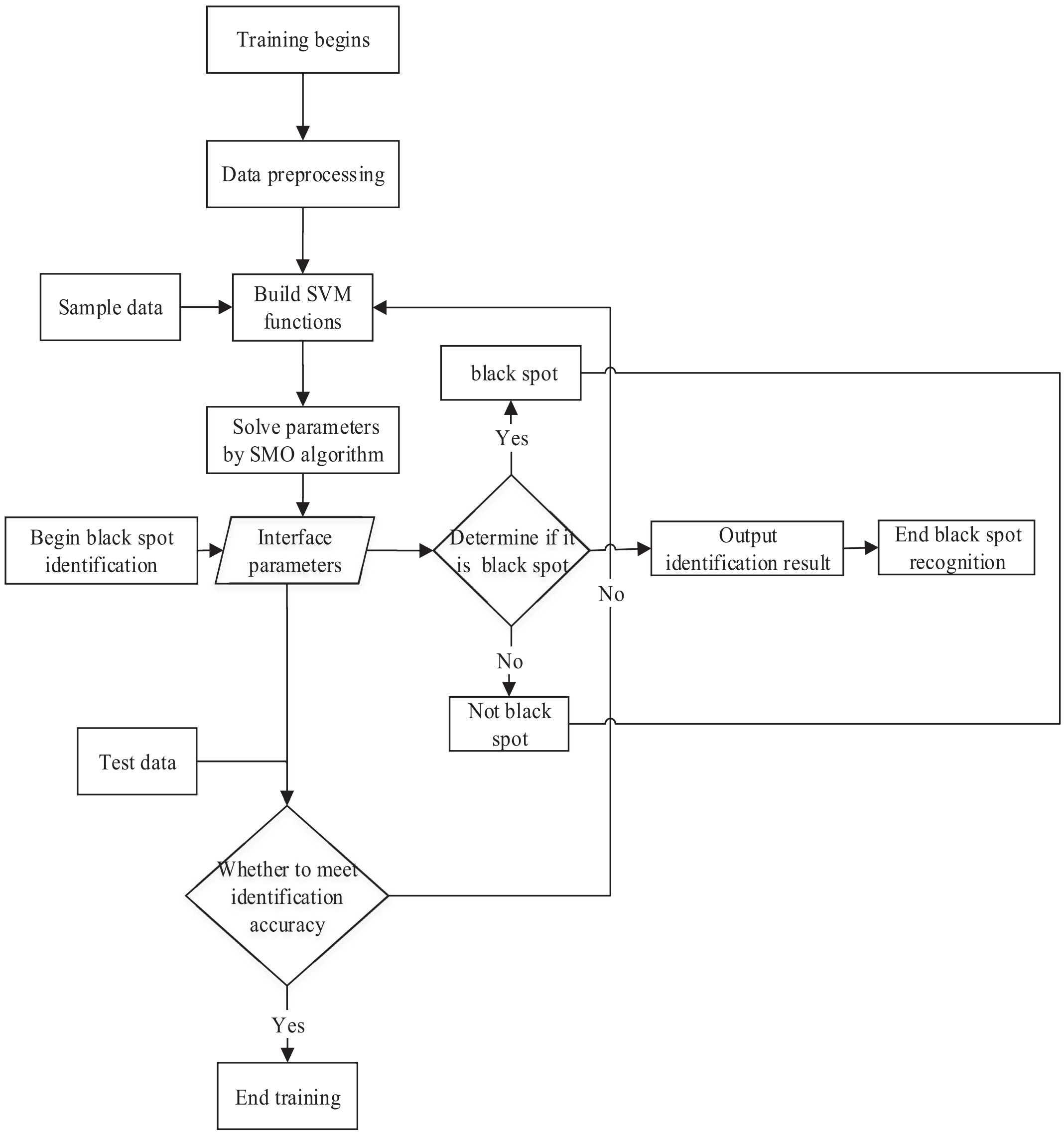

Fig. 2. Black-spot identification algorithm flowchart based on SVM.

will occur in the "black segment." The accident points on it were classified as black spots. The correlation between each original feature and the black point label of the accident was calculated by a χ^2 test, and the 10 features with the strongest correlation with the black-point feature were selected. These features are the x and y coordinates, weather conditions, accident type, cause of accident, type of road, road condition, ambient illumination, number of weeks, and month. Based on the above 10 features, we standardized continuous features and sparsely expressed discrete features to obtain 118 high-dimensional features. To improve the calculation speed of the model and reduce the influence of noise, we used the PCA to feature the dimension reduction of the high-dimensional features, and then used the radial basis kernel function (RBF) to map the sample space. The dimension of the data after data preprocessing in the paper was 26 dimensions. Finally, the SVM model was used to realize the black spots of the accident. In this experiment, 1083 of the 6391 samples were black spots. On this basis, we used a 25% test sample and 75% training sample. The original sample was divided into datasets. We ran the experiment 100 times, with prediction results as shown in Table 1.

Support vector machine accident black-point prediction achieved 88% accuracy, consisting of 92% for ordinary accidents and 63% for black spots. The prediction accuracy of black-spot accidents was lower than that of ordinary accidents, mainly because the sample proportion of black-spot accidents was relatively small, and the unbalanced sample distribution introduced bias into the learning model. The results also show that the x and y coordinates, weather conditions, accident type, accident cause, road type, road condition, environmental illumination, week, and month are closely related to whether an accident is a black spot. Considering the possible correlation between the accident and the implementation of traffic flow, we extracted the traffic volume of the road section of an accident 15, 30, 45, and 60 min before an accident, and used it as a dynamic traffic feature for traffic-accident prediction. The prediction results are shown in Table 2.

According to the data in Table 2, the real-time traffic flow significantly improves the prediction accuracy and recall rate of black-spot accidents, and finally achieves 95% black-spot accident prediction accuracy.

Based on the classical classification algorithm in machine-learning-support vector machine for black-spot identification of traffic accidents, the distribution of black spots in traffic accidents in Suzhou Industrial Park can be accurately identified. However, with the rapid growth of multi-source data in traffic accidents, the accuracy of the precision and recall rate of point-identification will be lower and lower, and the identification efficiency is not high. For this reason, the deep neural

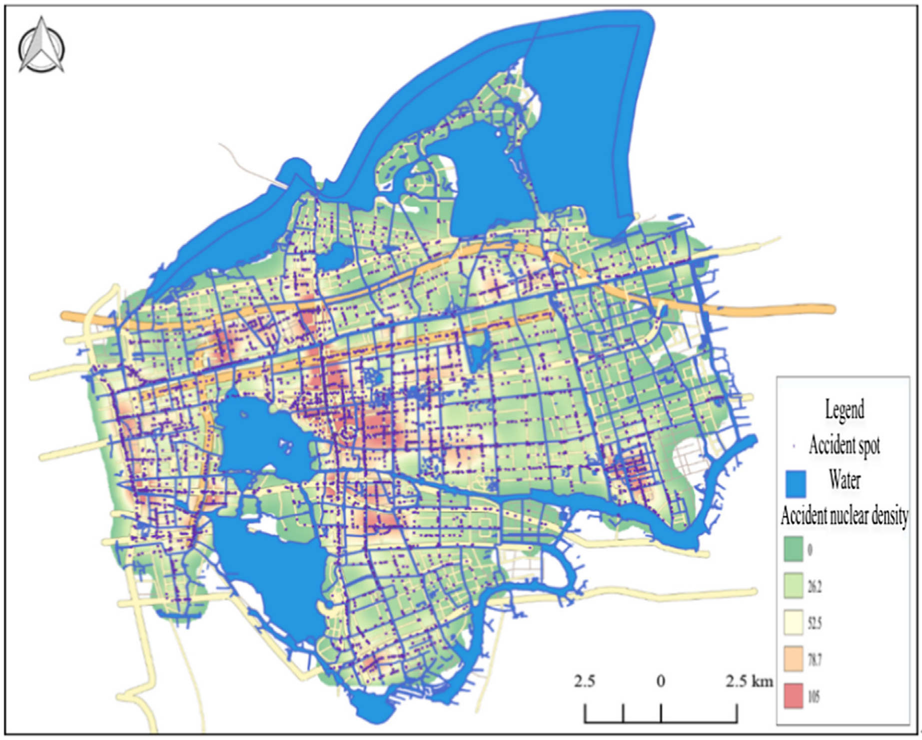

Fig. 3. Accident nuclear-density analysis and accidental spatial distribution.

Table 1
SVM accident black-spot prediction results.

	Precision rate	Recall rate	f_1 index	Support
Ordinary accident	92%	93%	93%	1338
Black spots accident	63%	61%	62%	260
Average value	88%	88%	88%	1598

Table 2
SVM accident black-spot prediction results after adding road-section traffic flow.

	Precision rate	Recall rate	f_1 index	Support
Ordinary accident	84%	98%	90%	86
Black spots accident	95%	69%	80%	51
Average value	88%	87%	86%	137

network algorithm in deep learning will be used to identify the black spots of traffic accidents in Suzhou Industrial Park.

3. Black-spot identification based on deep neural network

3.1. Black-spot identification process based on deep neural network

Deep neural network (DNN) algorithms have become a research hotspot in the field of machine-learning in industry and academia (Lecun et al., 2015). The DNN algorithm significantly improved the recognition rate (Li et al., 2017). We first establish a deep neural network that adds data-category information. Then the data to be identified are input to an input layer of the deep neural network generated based on the above data-category information. Finally, we obtain the category information of the data to be identified from the output layer of the deep neural network (Jin et al., 2017). The category information of the data to be identified is conveniently and quickly obtained by establishing a deep neural network based on the category information of the data (Jiang et al., 2015). Therefore, the category-recognition function of the deep neural network is realized, which is convenient for mining the deep regular pattern of the data to be identified (Cireşan et al., 2012), according to the category information of that data. The algorithm flowchart is shown in Fig. 4.

The steps in deep neural network black-spot identification are as follows:

(1) Establish an initial deep neural network;
(2) Generate a linear category analysis function after adding the data-category information in the locally saved initial linear analysis function according to the input training sample vector set;
(3) Obtain an optimization function of the initial depth neural network according to the locally stored unsupervised coding model optimization function and the linear class analysis function;
(4) Acquire parameters of the initial deep neural network according to the optimization function of the initial depth neural network;
(5) Establish a deep neural network according to a locally stored classification neural network, an initial deep neural network, and parameters of the initial deep neural network;
(6) Receive the input data to be identified;
(7) First, the data to be identified is input to the input layer of the deep neural network. Then we obtain the category information of the data to be identified from the deep neural network's output layer.

3.2. Black-point identification analysis results based on deep neural network

DNN has a more complex network structure and layers than a traditional neural network, which is important in deep learning. The number of layers of the neural network directly determines its ability to portray reality. DNN uses fewer neurons per layer to fit more complex functions to solve more complex problems (Siddiqui et al., 2018). A DNN consists of an input layer, hidden layers, and output layer. The hidden layer is composed of a number of neurons, and the number of hidden layers is usually greater than four. Because the original features of black spots in traffic accidents are mostly discrete, we used the

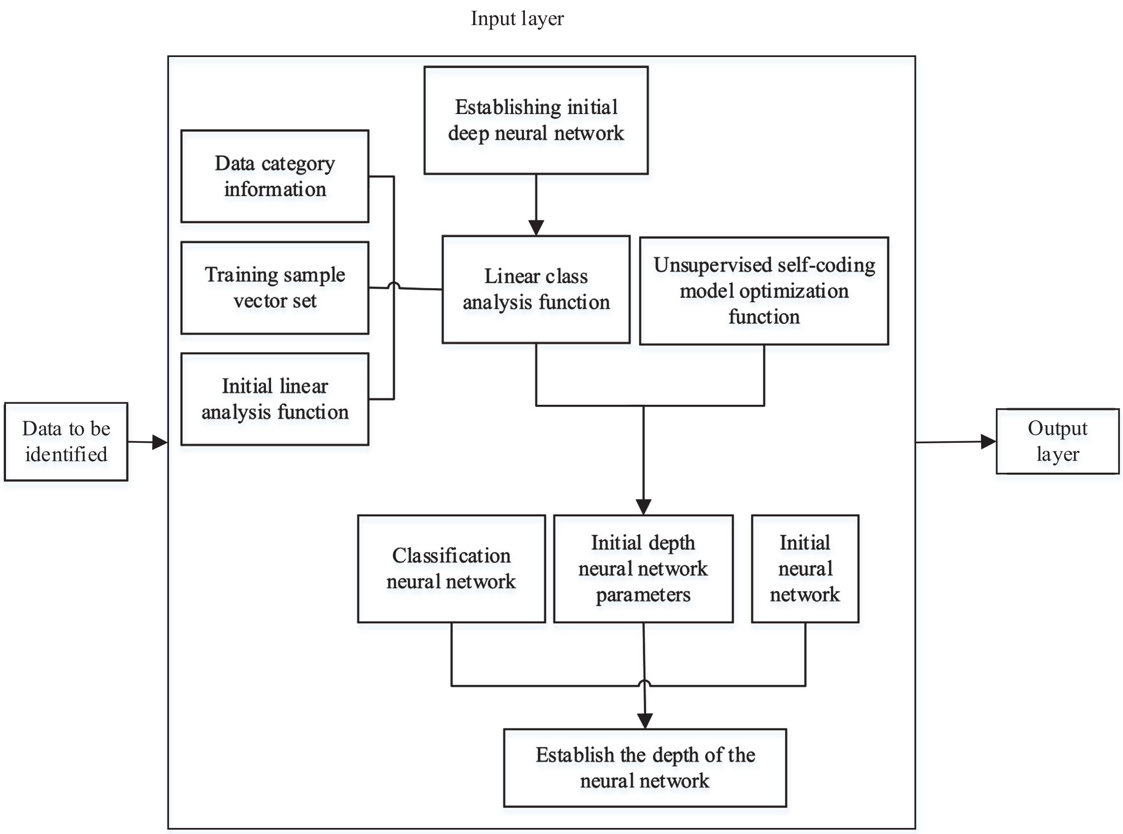

Fig. 4. Flowchart of algorithm for black-spot identification based on deep neural network.

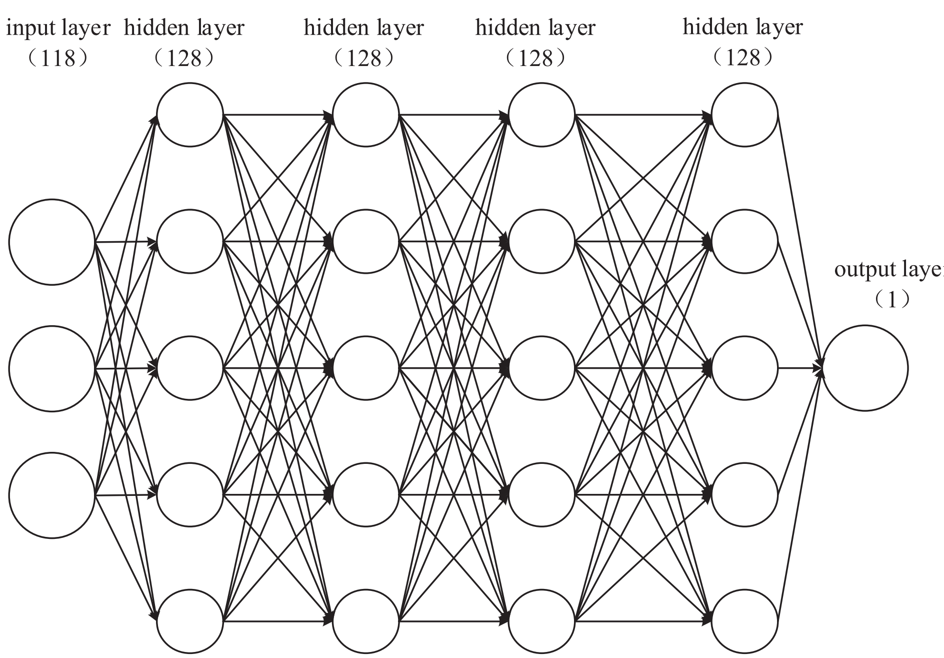

Fig. 5. DNN prediction-model structure of traffic-accident black spots.

Table 3
DNN traffic accident black-spot prediction results.

	Precision rate	Recall rate	f_1 index	Support
Ordinary accident	94%	93%	93%	1322
Black-spot accident	67%	70%	69%	276
Average value	89%	89%	89%	1598

network structure shown in Fig. 5 to construct a DNN prediction model for traffic-accident black spots (Gregoriades and Mouskos, 2013). The hidden-layer activation function used the ReLU function. To improve the overall discriminative ability of the model in traffic-accident black-point prediction, we define the loss function of the model as accuracy:

$$Accuracy = (\#TP + \#TN)/(\#TP + \#FP + \#TN + \#FN) \quad (8)$$

Table 4
Prediction results of DNN traffic-accident black spots introduced into section traffic flow.

	Precision rate	Recall rate	f_1 index	Support
Ordinary accident	91%	96%	93%	95
Black-spot accident	89%	79%	84%	42
Average value	90%	91%	90%	137

Consistent with the black-spot prediction experiment based on the SVM model, we used a 25% test sample and 75% training sample to divide the original sample into datasets, and we repeated the experiment 100 times. The prediction results of the DNN model are shown in Table 3.

It can be seen from the results that the prediction results of the DNN model were close to those of the SVM model. The average precision and average recall rate were both 89%. In terms of black-spot accidents, the DNN model showed a certain improvement compared to SVM. The black-spot accident precision rate was 67%, which was 4% higher than that of the SVM model; the recall rate was 70%, which was 9% higher than that of the SVM model. The model was again trained after introducing the section traffic flow of 15, 30, 45, and 60 min before the accident. The prediction results are shown in Table 4.

The results in Table 4 show that the section traffic flow can improve the prediction accuracy of the DNN model for black-spot accidents, with 89% precision, but the improvement is not as good as that of the SVM model. At the same time, the recall rate of black-spot accidents has also significantly improved. Therefore, we believe the black spot of an accident is strongly correlated with the traffic flow at the pre-order time of the spatial position, and the introduction of the cross-sectional flow helps to improve the accuracy of the black spot of an accident.

3.3. Deep neural network black point identification based on features

(1) Accident black spots identification strategy under different weather conditions

Among the training samples, there were a total of 1265 samples of sunny accidents, which were evenly and densely distributed in space, and were also the most important type of accident point samples. The location of the traffic accident on sunny days was shown in Fig. 6.

It can be seen from Fig. 6 that the black spots of sunny accidents were concentrated in the core business circle east and west of Jinji Lake and the vicinity of the city railway station. These areas were also the areas with the most dense traffic. This indicates that in the case of good daily weather, black spots of accidents were directly related to the density of traffic activities and indirectly related to frequent economic activities.

It can be seen from Fig. 7 that there are 108 samples of cloudy accidents, which were sparsely distributed in space, mainly concentrated on the north-south trunk road on the east side of Jinji Lake. Its distribution pattern was significantly different from the cluster shape of the sunny day. It can be speculated that the black spots of the accidents in addition to the activity of the traffic activities may be strongly related to the environment of the road sections themselves.

It can be seen from Fig. 8 that there are 222 samples in the rainy day accident point, which were relatively evenly distributed in space and generally had a density. The distribution pattern of black spots in accidents was similar to that of sunny days, and was clustered in the core business district east of Jinji Lake. Unlike the black spots of accidents on sunny days, the black spots of rainy days were concentrated in a single core in space, which reflected the fact that bad weather conditions increase the probability of accidents in areas with high traffic accidents. Table 5 gave a comparison of the accuracy of black spots predictions for traffic accidents under different weather conditions.

It can be seen from Table 5 that the better the weather conditions, the higher the accuracy of black spot recognition. The main reason was that the number of samples was high when the weather conditions were good, and the paradox of the distribution of the samples on the weather conditions affects the recognition accuracy. It was worth noting that, in the case of poor weather conditions, even if the number of samples is small, the accuracy of black spot recognition of the accident remains at a high level, indicating that the characteristics of the accident black

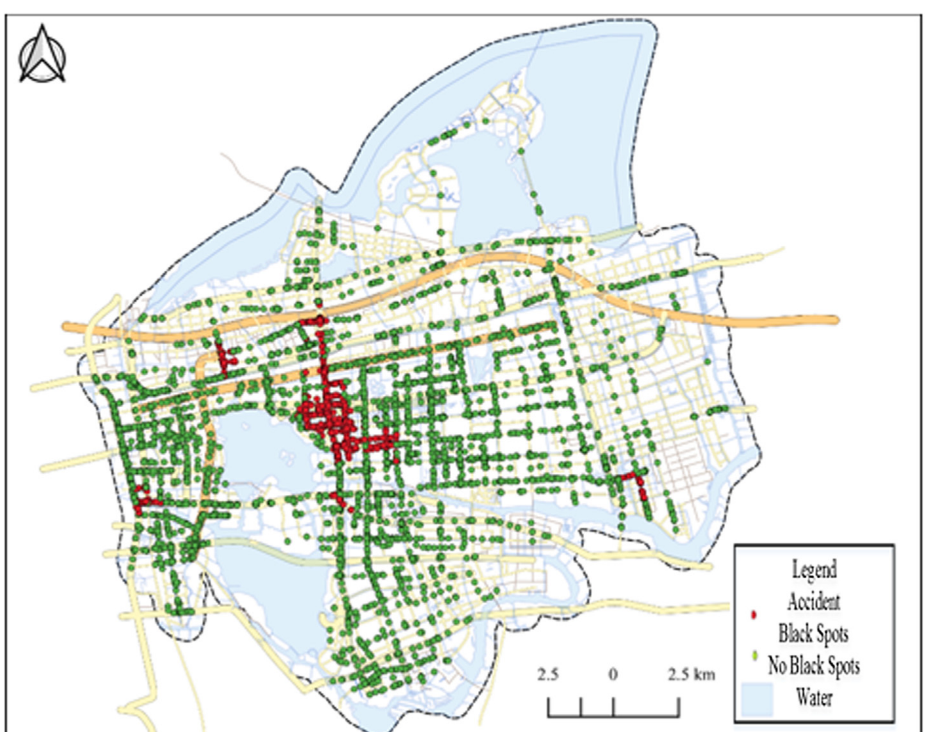

Fig. 6. The location of the traffic accident on sunny days.

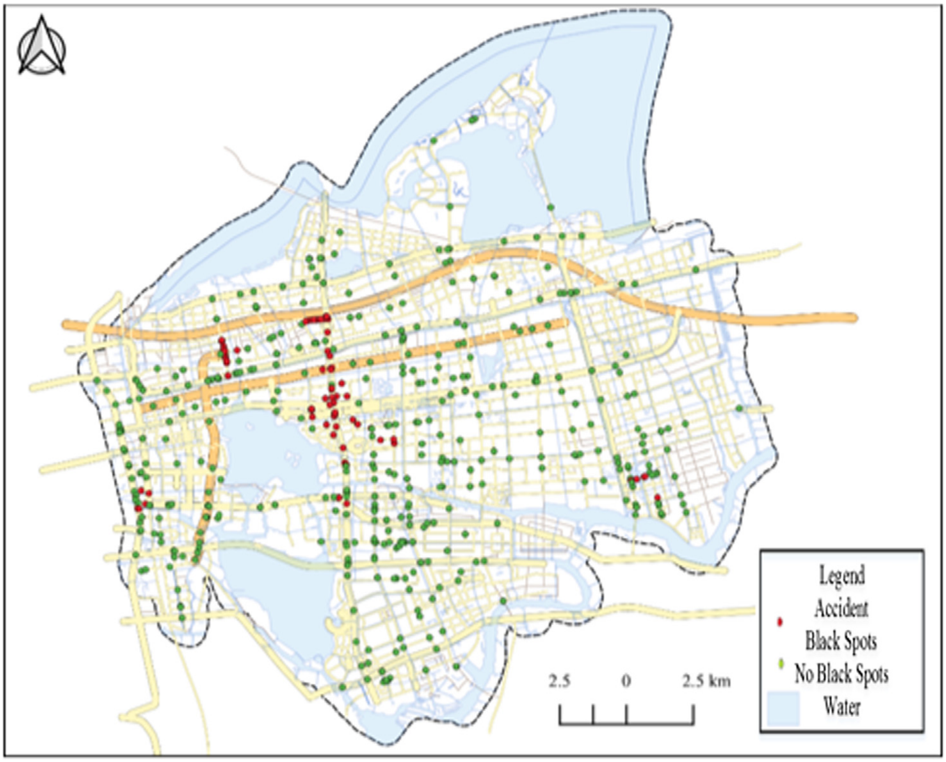

Fig. 7. The location of the traffic accidents on cloudy days.

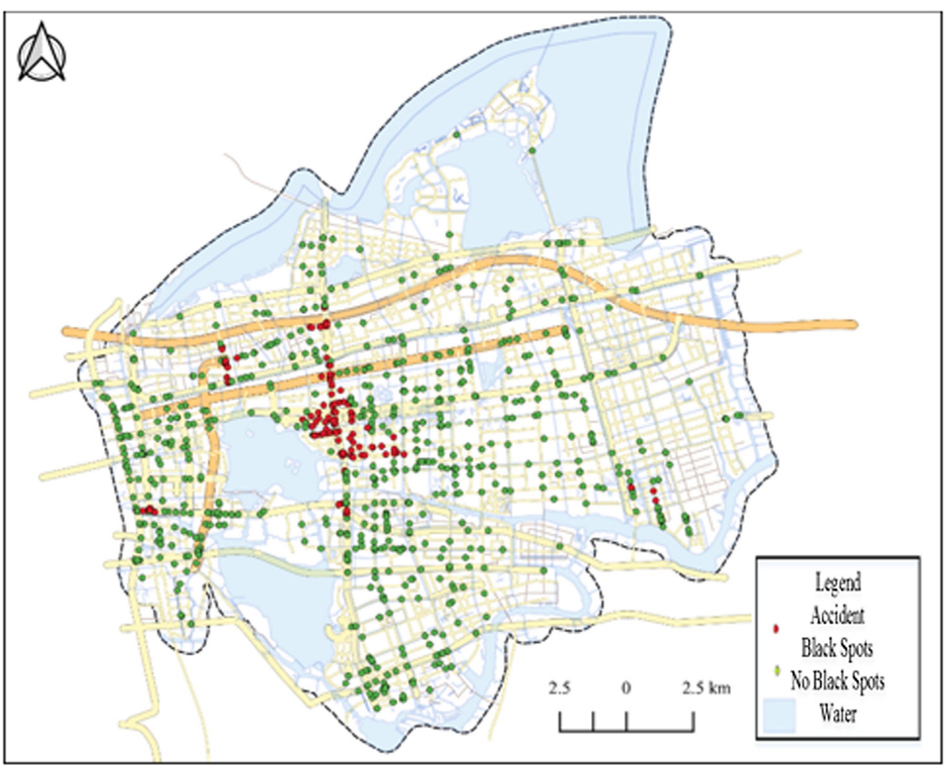

Fig. 8. The location of the traffic accidents on rainy days.

spots are more obvious, and the more easy to identify.

(2) Accident black point identification strategy under different time conditions

According to the daily commuter statistics of Suzhou City, traffic travel has a relatively dense time in the morning and evening. This article will divide the daily 7:30–8:30 into the morning and 17:00 to 18:00 evening peak hours. The distribution of black spots in the morning and evening peak hours was analyzed separately. According to

Table 5
Black out prediction results of traffic accidents under different weather conditions.

	Sunny days	Cloudy days	Rainy days
Black point recognition accuracy	88.70%	85.19%	83.33%
Accident sample size	1265	108	222

Fig. 9, the accident points at the morning peak were distributed continuously on the main roads in the east-west direction of the city, indicating that the probability of accidents in the east-west traffic flow was higher, which was related to the daily law of the east-west commuting of the morning peak in Suzhou. At the same time, the distribution of black spots in the accident was mainly along the eastern coastal area of Jinji Lake. Fig. 10 showed that the accident distribution at the late peak was relatively random and the density was relatively low. The distribution of black spots in the accident was mainly concentrated on the north-south trunk line east of Jinji Lake. The distribution of the accident point showed that the purpose of the late peak travel was relatively weaker than the commuter traffic phenomenon of the early peak tension. The side of the morning peak that reflected the tension law has a significantly higher probability of accidents than the late peak traffic

Table 6 showed the prediction accuracy of black spots in the morning peak and evening peak hours. The results showed that the evening peaks hours had a higher prediction accuracy when the number of samples were smaller. This paper concluded that under the pressure of commuting under the peak of the morning peak, the factors of accidents were more complicated, and the characteristic performance was not concentrated at night, so it was more difficult to accurately identify the black spots of the accident.

Based on the above analysis, the feature-based deep neural network dynamically displayed the spatial distribution of traffic accident black spots. At the same time, in order to get more feature-based deep learning neural network black point identification results, this paper can divide the dynamic learning model into six identical independent models, including: sunny morning peak, sunny night peak, cloudy morning peak, cloudy night peaks, rainy morning peaks, rainy days and evening peaks, respectively, continue to add corresponding sample data for dynamic learning, in order to gradually improve the model prediction accuracy, and provide an effective data reference for practical engineering applications.

3.4. Traffic accident black spots real-time online deep learning framework

Based on the spatial and temporal distribution of data and the error of the observation field and background field, this model proposed a dynamic deep neural network black spots identification method. New observation data was fused during the dynamic operation of the deep neural network black point identification model. It was within the dynamic training framework of the learning model. The online deep learning framework and the data assimilation strategy continuously integrated the discretely distributed observation sample information in time and space to automatically adjust the model parameters to improve the estimation accuracy of the dynamic learning model and improve the model prediction ability. Considering the different paranoia of sample distribution under different time and space conditions, according to the conclusions of the test in 2.3, the weather conditions and the morning and evening peak hours were taken as important factors affecting the sample distribution, and the samples are divided twice, and the division is used separately. The post-sample dynamic training deep neural network black point recognition model. Learning deep neural networks faces many challenges, such as gradient disappearance, local minima, a large number of parameters that need to be adjusted, difficulties in regularization methods, selection of hyperparameters, and so on. Current advances in these areas are primarily designed to address the specific problems encountered in optimizing deep neural networks, and most of these methods assume that deep neural networks are trained in batch learning settings that require. All training data sets are prepared before the learning task begins. This is not true for many

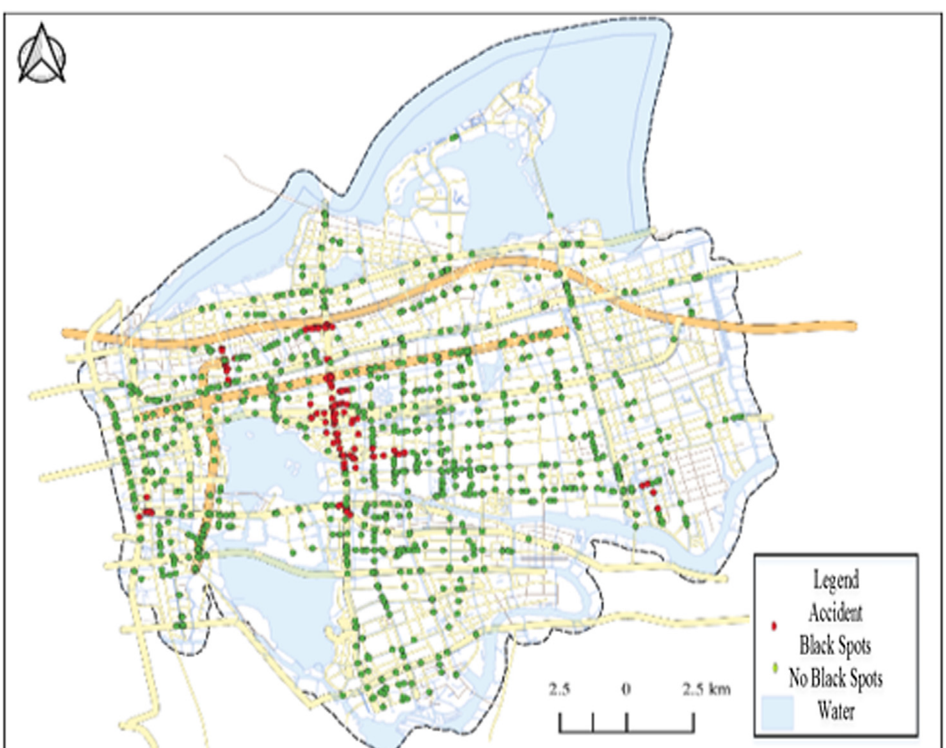

Fig. 9. Location distribution of early peak traffic accidents.

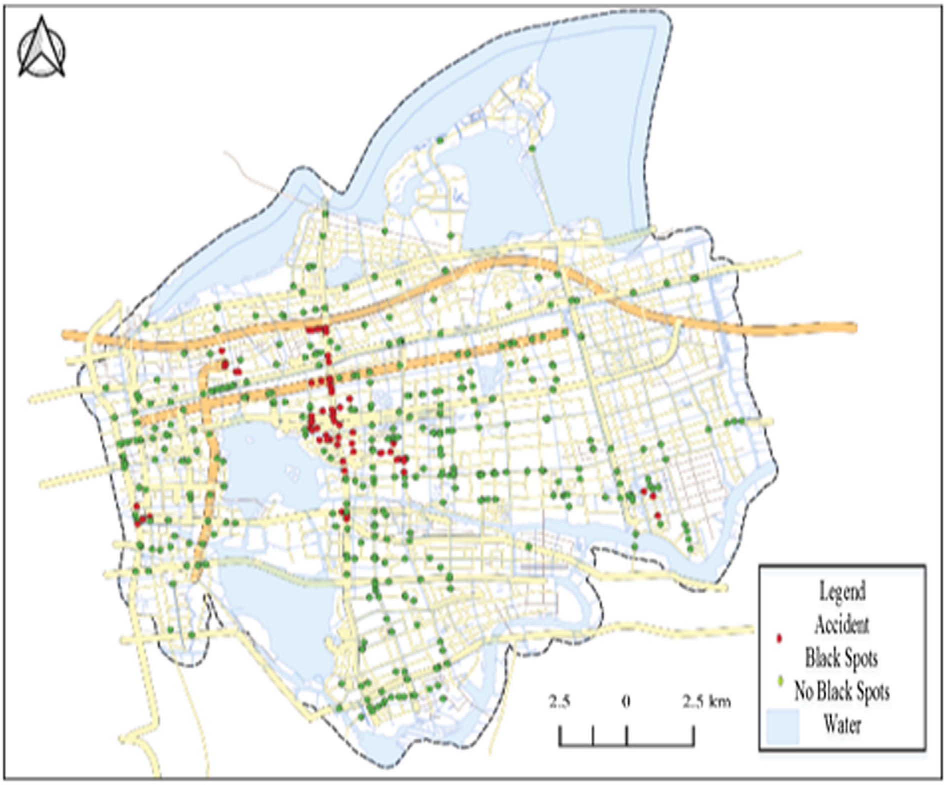

Fig. 10. Location distribution of evening peak hour traffic accidents.

Table 6
Black spots prediction results of traffic accidents under different weather conditions.

	Morning peak hours	Evening peak hours
Black spots recognition accuracy	86.72%	88.22%
Accident sample size	262	111

tasks in which real-time data arrives in the form of streams, and there may not be enough memory to store. The characteristics of traffic accident data vary with time and space. This continuous change poses a great challenge to maintain model performance. In addition, the data may also exhibit conceptual drift (Jo et al., 2014). Therefore, in further practical applications, consider establishing an online learning model to continuously train the model by continuously receiving new data.

Some researchers have proposed the basic idea of online learning algorithms (Zinkevich, 2003). This learning algorithm learns the optimal prediction model in the case of sequentially arriving data streams, making online learning more scalable and higher memory utility. However, most of the existing online learning algorithms are designed to use online convex optimization to learn shallow models (for example, linear methods or kernel methods), and they cannot learn nonlinear functions in complex application scenarios.

This paper used the algorithm of hedging backpropagation proposed by Sahoo et al. (2017) This algorithm was different from the standard backpropagation. In each round of online learning, the performance of each classifier was evaluated and passed. The backpropagation algorithm was extended using classifiers with different depths in the hedging algorithm (Freund et al., 1997) to train the deep neural network online, which helps to share information between the shallow network

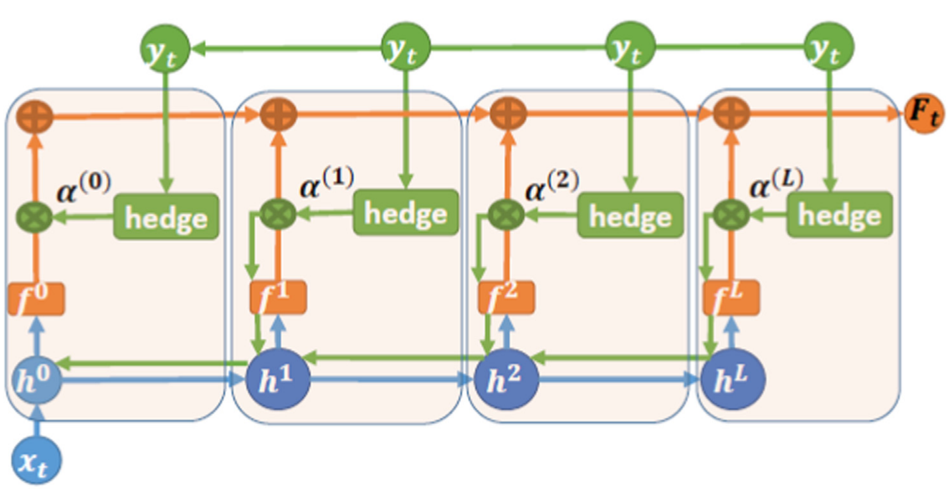

Fig. 11. Online deep learning framework using Hedging Back Propagation (HBP).

and the deep network (see Fig. 11).

The blue line represented the feedforward information flow when calculating hidden layer features. The orange line indicated the softmax output (representing the prediction) after the hedging combination. The green line indicated the update stream when the hedging back-propagation algorithm was used. Through algorithmic transformation based on this method, our model can finally learn the deep neural network model from the traffic accident data flow that arrives sequentially. More importantly, over time, it can adaptively expand the model capacity from simple to complex from time to time, combining the advantages of online learning and deep learning.

4. Conclusions

Frequent traffic accidents cause great harm. Therefore, quick and efficient identification of black spots of traffic accidents was an urgent problem. This paper first introduces the relationship between machine-learning and the SVM, explains the basic principle of the SVM, and proposed a black-point identification algorithm based on SVM. We then proposed a black-point analysis process based on a deep neural network. Machine-learning and deep-learning training were carried out on the traffic-accident data of Suzhou Industrial Park, and the exact locations of traffic black spots were predicted.

The conclusions of this paper were as follows:

(1) Based on nuclear density black-spot identification, combined with 10 characteristic data items, such as accidents, persons, vehicles, and road and environment, we selected the algorithm-support vector machine in machine-learning to predict and analyze the black spots of traffic accidents, and we obtained 63% precision and a 61% recall rate.
(2) For the increase of traffic-accident data, the prediction accuracy of SVM will continually decrease. For this reason, based on a deep neural network, the prediction and analysis of traffic accident black spots will achieve 67% precision and a 70% recall rate.
(3) In the training prediction of accident data, the accident road section flow of four time periods was added. The precision of SVM and deep neural network increased to 95% and 89%, respectively, and the respective recall rates increased to 69% and 79%. The results showed that the road-section flow of traffic accidents can identify black spots.
(4) Based on the characteristics of black spots weather, morning and evening peak distribution, the feature-based deep neural network black point identification was proposed. In order to improve the learning ability and prediction accuracy of deep learning, this paper proposed a real-time online deep learning framework, which provided a new research method for the rapid dynamic identification of traffic accident features based on massive traffic data.
(5) The characteristics of traffic accident data varied with time and space. This continuous change poses a challenge to maintaining model performance. In further practical applications, we will consider an online learning model to continuously improve the black-edge recognition accuracy based on new data.
(6) Traffic accident characteristics are becoming more and more complex and changeable. It is difficult to determine the cause of black spots. With the increase of traffic accident data, the training of black spot identification model becomes more complex. How to improve the accuracy, real-time and accurately find the cause of black spot identification model will become the focus of research.

Acknowledgements

We would like to thank Accdon (www.accdon.com) for providing linguistic assistance during the preparation of this manuscript. We are grateful for the support of the Jiangsu Provincial Surveying Mapping Geoinformation Research Project Fund, project number JSCHKY201712.

References

Almjewail. A., Almjewail, A., Alsenaydi, S., Alsudairy, H., Al-Turaiki, I., 2018. Analysis of Traffic Accident in Riyadh Using Clustering Algorithms.
Chang, Chih-Chung, Lin, Chih-Jen, 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3), 27.
Chen, H., 2012. Black spot determination of traffic accident locations and its spatial association characteristic analysis based on GIS. J. Geographic Inform. Syst. 4 (6), 608–617.
Chen, Y.W., Ren-De, Y.U., Sun, L.C., 2011. Discrimination of traffic accident black-spot based on clustering analysis method. Transport Stand.
Cherkassky, V., 2002. The nature of statistical learning theory. IEEE Trans. Neural Netw. 38 (4), 409.
Cireşan, D., Meier, U., Masci, J., Schmidhuber, J., 2012. Multi-column deep neural network for traffic sign classification. Neural Netw. 32 (1), 333–338.
Debrabant, B., Halekoh, U., Bonat, W.H., Hansen, D.L., Hjelmborg, J., Lauritsen, J., 2018. Identifying traffic accident black spots with Poisson-Tweedie models. Accid. Anal. Prev. 111, 147–154.
Elvik, R., 2002. The importance of confounding in observational before-and-after studies of road safety measures. Accid. Anal. Prevent. 34 (5), 631–635.
Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans. Geosci. Rem. Sens. 46 (11), 3804–3814.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 23–37.
Gregoriades, A., Mouskos, K.C., 2013. Black spots identification through a Bayesian Networks quantification of accident risk index. Transp. Res. Part C Emerg. Technol. 28 (3), 28–43.
Jiang, B., Wu, T., Zheng, C., Wong, W.H., 2015. Learning summary statistic for approximate bayesian computation via deep neural network. Statistics 27 (4).
Jin, K.H., Mccann, M.T., Froustey, E., Deep, Unser M., 2017. Convolutional Neural Network for Inverse Problems in Imaging.[J].IEEE. TRANSACTIONS ON IMAGE PROCESSING 26 (9), 4509–4522.
Jo, Gama O, Bifet, A., Pechenizkiy, M., 2014. Bouchachia A.A survey on concept drift adaptation. ACM Comput. Surveys (CSUR) 46 (4), 1–37.
Lecun Y,Bengio Y,Hinton G.Deep learning[J].NATURE,2015,521(7553): 436.
Li, Y., Xie, W., Li, H., 2017. Hyperspectral image reconstruction by deep convolutional neural network for classification[J].PATTERN. RECOGNITION 63, 371–383.
Ma, Z., Chen, Y., Zhang, L., 2014. Influence factors of accident severity for urban road. J. Chongqing Jiaotong Univ. (Natl. Sci.) 33 (1), 111–114.
Mcguigan, D., 1981. The use of relationships between road accidents and traffic flow in “black-spot” identification. Traffic Eng. Control 22.
Oppe, S., 1991. Development of traffic and traffic safety: global trends and incidental fluctuations. Accident Anal. Prevent. 23 (5), 413.
Oskoei, M.A., Hu, H., 2008. Support vector machine-based classification scheme for myoelectric control applied to upper limb. IEEE Trans. Biomed. Eng. 55 (8), 1956–1965.
Robinson, E.C., Hammers, A., Ericsson, A., Edwards, A.D., Rueckert, D., 2010. Identifying population differences in whole-brain structural networks: a machine learning approach. Neuroimage 50 (3), 910–919.
Saccomanno, F.F., Grossi, R., Greco, D., Mehmood, A., 2001. Identifying black spots along highway SS107 in Southern Italy using two models. J. Transp. Eng. 127 (6), 515–522.
Sahoo, D., Pham, Q., Lu, J., Hoi, S.C.H., 2017. Online deep learning: learning deep neural. Netw. Fly.
Shao, Z., 2008. Reviews on identifying methods of traffic accident black-spot. Road Traffic Safety.
Shao, Z., Cai, J., 2009. Problems in practice of identifying traffic accident macula and its countermaesures. J. Hubei Univ. Police (02), 83–87.
Sharif, R.E.K.S. Prioritization of accident black spots using GIS2014.
Shen, X., Guo, X.C., Song, J.M., 2003. Study on road traffic accident black spot identification method. J. Highway Transp. Res. Dev.
Siddiqui, S.A., Salman, A., Malik, I., Shafait, F., Mian, A., Shortis, M., Harvey, E., 2018. Automatic fish species classification in underwater videos: exploiting pretrained deep neural network models to compensate for limited labelled data. ICES J. Marine Sci. 75 (1).
Wang, H., Sun, L., You, K., 2013. Accident-prone location identification method based on DENCLUE clustering algorithm. J. Transp. Eng. Inform. (2), 5–10.
Zinkevich, M., 2003. Online convex programming and generalized infinitesimal gradient ascent. Icml 928–936.