Physica A 517 (2019) 29–35

Contents lists available at ScienceDirect

Physica A

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

SVM and KNN ensemble learning for traffic incident detection

Jianli Xiao

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

HIGHLIGHTS

- This paper proposes a new ensemble learning method to detect traffic incidents.
- The proposed method has achieved satisfactory accuracy in traffic incident detection.
- More importantly, the proposed method has improved the robustness of traffic incident detection.

ARTICLE INFO

Article history:
Received 11 July 2018
Received in revised form 26 October 2018
Available online 5 November 2018

Keywords:
Traffic incident detection
SVM
KNN
Ensemble learning

ABSTRACT

Traffic incident detection is a very important research area of intelligent transportation systems. Many methods have obtained good performance in traffic incident detection. However, the robustness of these methods is not satisfactory. Namely, when one method is applied on another data set again, its performance is not always good, even it had obtained good performance on one data set once. In this paper, we propose an ensemble learning method to improve the robustness in traffic incident detection. The proposed method trains individual SVM and KNN models firstly. And then, it takes a strategy to combine them for better final output. Experimental results show that the propose method has achieved the best performance among all the compared methods. Also, the ensemble learning strategy in the proposed method has improved the robustness of the individual models.

1. Introduction

Traffic incident detection is a very important research area of intelligent transportation systems (ITS) [1–4]. The definition of traffic incidents are the nonrecurring events, for example, accidents, disabled vehicles, spilled loads, temporary maintenance and construction activities, signal and detector malfunctions, and other special and unusual events that disrupt the normal flow of traffic and cause motorist delay [5]. A lot of methods have been presented in the last few decades. According to the detectors they used, these methods can be classified into the methods used: video detectors, GPS detectors, infrared detectors, radar detectors, industrial loop detectors (ILDs), and so on.

Video cameras are usually installed to monitor the traffic state of the roads. Ren et al. [6] propose a video-based method to detect and position the traffic incidents, and then analyze the traffic states distribution of the road sections. In this way, the traffic flow, average travel speed and average space occupancy are obtained and adopted to detect traffic incidents. Yun et al. propose a new method to detect the traffic incidents by modeling of interaction among multiple moving objects. Traffic incident detection methods based on video detectors have achieved good performance. However, the videos from cameras are affected by weather and light easily. For example, the accuracy of the traffic incident detection accuracy will significantly reduced in the rainy and snowy days.

GPS detectors are also used for traffic incident detection. Asakura et al. [7] research the properties of traffic flow dynamics when the traffic incidents has arisen with GPS data, and propose an algorithm to predict the time and location of traffic

E-mail address: audyxiao@sjtu.edu.cn.

https://doi.org/10.1016/j.physa.2018.10.060
0378-4371/

congestion caused by an incident. Andrea and Marcelloni [8] propose an expert system for traffic incident detection which uses the real-time GPS data. This system is designed to support the traffic management for municipalities and citizens. Owing to the limitation of GPS data precision, researchers turn to the other methods to improve the accuracy of the traffic incident detection.

Infrared detector is a kind of non-contact sensor, which can be used to detect the traffic incidents conveniently. But it is not easy to overcome the influence of the weather for the infrared detector [9]. Radar detectors are applied on some specific occasions. As the high maintenance costs, they are not used widely. ILDs are popular traffic detectors, which are widely used in the world [10–12]. They can detect the traffic parameters accurately, and their performance is very robust even they have suffered from bad weather. More importantly, the price of the ILDs are very cheap. Thus, in this paper we focus on the traffic incident detection methods which use the data sampled by the ILDs. Many methods have been proposed for traffic incident detection with ILDS data. Most of them are based on machine learning [13,14]. Artificial neural networks is a classical method in machine learning. Ritchie et al. [15] adopt the artificial neural networks to detect traffic incidents and obtain satisfactory performance. However, the optimal parameters of the artificial neural networks are difficult to be obtained. For better performance, support vector machine (SVM) is used to detect traffic incidents by Fang et al. [16]. During to the trivial of selecting the optimal kernels and their parameters in SVM, the robustness of this method is not always satisfactory. Ensemble learning is a popular topic of machine learning [17,18]. It integrateds different individual models to improve the robustness [4]. In order to improve the robustness of the methods in traffic incident detection, we propose a new ensemble learning method, SVM and KNN ensemble learning, to detect the traffic incidents. The proposed method first trains individual SVM and k-Nearest Neighbor (KNN) models [19]. Then, it constructs the ensembles to detect the traffic incidents.

The remaining part of the article is structured as follows: Section 2 presents the individual SVM and KNN models. The SVM and KNN ensemble learning method is presented in Section 3. Section 4 is devoted to experiments. At last, the conclusion has been drawn in Section 5.

2. Individual SVM and KNN models

SVM is a kernel learning method [3], which uses the training set to construct a hyperplane for classifying the test samples, the function of the hyperplane can be described as:

$$f(x) = \omega x + b \tag{1}$$

where $f(x)$ is a hyperplane function. ω is a normal vector of the hyperplane and b is a variable.

Suppose the training set can be described as:

$$S = \{(x_1, y_1), (x_2, y_2), \dots, (x_h, y_h), \dots, (x_l, y_l)\} \tag{2}$$

where x_h denotes any one training sample, and $x_h \in R^n$, $h = 1, 2, \dots, l$. $y_h \in \{-1, 1\}$ is the class label for x_h. l is the number of the training sample.

With the training set S, the optimal ω and b can be obtained by solving the following optimal problem:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_h y_j \alpha_h \alpha_j K(x_h, x_j) - \sum_{h=1}^{l} \alpha_h \tag{3}$$

$$s.t. \quad \sum_{h=1}^{l} y_h \alpha_h = 0 \tag{4}$$

$$0 \le \alpha_h \le C \quad i = 1, 2, \dots, l \tag{5}$$

where $\alpha = (\alpha_1, \dots, \alpha_l)^T$ is the Lagrange multiplier vector, K is a kernel function and C is a penalty factor by manual setting. x_h and x_j are inputs of any two training samples, and y_h and y_j are their labels respectively. There is not a unified approach for selecting an optimal kernel function, which is chosen manually [3]. The common kernel functions include Linear Kernel, Polynomial Kernel, Gaussian Kernel, etc.

Suppose the optimal ω and b are ω^* and b^* solved from the optimal problem Eqs. (3)–(5), which are:

$$\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)^T \tag{6}$$

$$\omega^* = \sum_{h=1}^{l} \alpha_h^* y_h x_h \tag{7}$$

$$b^* = y_j - \sum_{i=h}^{l} \alpha_h^* y_h K(x_h, x_j) \tag{8}$$

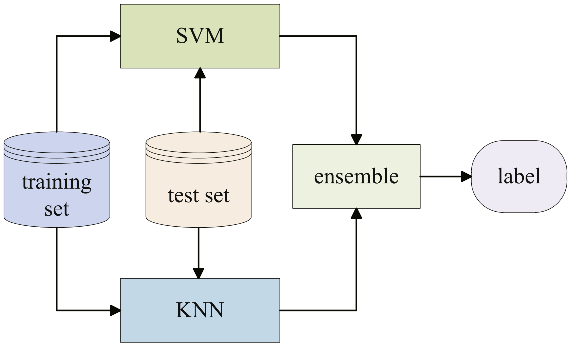

Fig. 1. The framework of the propose method.

Then $f(x)$ can be computed with the following formula:

$$f(x) = \sum_{h=1}^{l} y_h \alpha_h^* K(x_h, x) + b^* \tag{9}$$

For any one test sample, the class label can be computed with Eq. (9).

KNN method [19] computes the class label for the test sample by the labels of the k nearest neighbors of the test sample. Suppose the distance metric has been defined (such as Euclidean distance, Mahalanobis distance, and so on). For any one test sample x, its k nearest neighbors can be found, and we use $N_k(x)$ to denote them. The class label of x is determined by the labels of the training samples in $N_k(x)$, which can be described as:

$$g(x) = arg \max_{C_j} \sum_{x_h \in N_k(x)} I(y_h = C_j) \tag{10}$$

where $h = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, M$. M is the number of the classes of the labels. x_h is a training sample in $N_k(x)$, and y_h is the class label of x_h. C_j is the label of the jth class. $g(x)$ is a decision function. I is an indicator function. For each class, it has a indicator function. Take the jth class as an example, its indicator function is:

$$I = \begin{cases} 1 & \text{if } y_h = C_j \\ 0 & \text{else} \end{cases} \tag{11}$$

In KNN model, the decision function $g(x)$ has different forms. In our research, we present the formula for $g(x)$ according to the voting strategy.

3. SVM and KNN ensemble learning

SVM is proposed by Cortes and Vapnik [20], which is widely used in classification and regression problems. KNN is another very successful model in machine learning, which is proposed by Cover and Hart [21]. As the superior performance of SVM and KNN, in this paper, we propose an SVM and KNN ensemble learning method, which employs individual SVM and KNN models to construct ensemble. As shown in Fig. 1, the proposed method uses the training set to train individual classification models: KNN and SVM models firstly. Then, the test set is assigned to the SVM and KNN models respectively, and the results of these two models are put into the ensemble model to generate the final output. Suppose the output of the SVM and KNN models are $f(x)$ and $g(x)$. The final output of the proposed method can be computed with the following formula:

$$e(x) = f(x) * P + g(x) * (1 - P) \tag{12}$$

where $e(x)$ is the final output of the ensemble model. P is a binary function. If the probability score of the incident generated by the SVM model is larger than the threshold T, the value of P is "1". Otherwise, the value is "0".

Algorithm 1: SVM and KNN Ensemble Learning.

(1) divide the whole data set into training and test sets
(2) use the training set to compute the optimal ω and b according to Eq. (3)–(5)
(3) obtain the decision function of the SVM model: $f(x) = \omega x + b$
(4) compute the label $g(x)$ of the test set with the formula Eq. (10)
(5) compute the final output generated by the ensemble with the equation: $e(x) = f(x) * P + g(x) * (1 - P)$

---

4. Experiments

In order to evaluate the performance of the proposed method, extensive experiments are performed. For comparison, Naive Bayes [22], SVM method [3] and KNN method [19] are also adopted to perform the experiments. For evaluation, we repeat each experiment for 30 times, and compute the averages and variances of the performance for each method.

4.1. Data set

Two data sets are employed to perform the experiments, which are I-880 data set [23] and PeMS data set. I-880 data set is download from the web site. This data set is collected at the I-880 freeway in the San Francisco Bay Area, California. The whole data set includes two parts: one part has incident samples and the other part has normal samples. The loop detectors are installed on the freeway from a 14.8-km segment of I-880 between the Marina and Wipple exits in both directions. There are 35 loop detector stations on the freeway segment. Data is collected at 30-s intervals including volume, speed, and occupancy.

PeMS data set is constructed by us, The raw data is obtained from a real-time archived data management system [24], which collects raw detector data from California freeways in real time. There are 8840 samples in the PeMS data set, including 1640 incident samples and 7200 nonincident samples. More importantly, we do not eliminate the noise data for testing the performance stability of the methods.

4.2. Performance criteria

In our research, three existing criteria : detection rate (DR), false alarm rate (FAR), and classification rate (CR) are taken to evaluate the performance of each method from Ref. [3].

DR is the ratio of number of detected incidents to the total number of incidents. DR indicates the accuracy towards detecting incident cases. A larger value of DR indicates that the method is more sensitive to the traffic incidents and can detect the traffic incidents more accurately. DR is computed by the following formula [3]:

$$DR = \frac{\text{number of incident cases detected}}{\text{total number of incident cases}} \times 100\% \tag{13}$$

FAR is the ratio of the false alarm cases to the total number of non-incident instances . For traffic incident detection, the smaller the FAR is , the better the performance is. The following equation is used to compute FAR.

$$FAR = \frac{\text{number of false alarm cases}}{\text{total number of non-incident cases}} \times 100\% \tag{14}$$

CR is the proportion of instances that were correctly classified based on total instances in data set. CR can be computed as:

$$CR = \frac{\text{number of instances correctly classified}}{\text{total number of instances}} \times 100\% \tag{15}$$

For traffic incident detection, the larger the CR is, the better the performance of the method is.

In our research, we propose a new comprehensive criterion, performance index (PI), to evaluate the performance of the methods. PI has combined the criteria DR, FAR, and CR, which can evaluate the performance of the classifier more comprehensively, and can be computed with the following formula:

$$PI = w_{DR}DR + w_{FAR} \cdot (1 - FAR) + w_{CR}CR \tag{16}$$

where w_{DR}, w_{FAR}, and w_{CR} are the weights of DR, FAR, and CR respectively. In our research, the values for w_{DR}, w_{FAR}, w_{CR} are all $\frac{1}{3}$. A larger value of PI indicates that the method has better performance in traffic incident detection.

4.3. Experimental results

The experimental results on data sets I-880 and PeMS are drawn in Figs. 2, 3, Tables 1 and 2. From Fig. 2 and Table 1 we can see that the proposed method has achieved the best performance on the criterion PI. As PI is a comprehensive criterion which combines the criteria DR, FAR, and CR, if one method has the largest value on PI, it has the best comprehensive performance. Thus, the proposed method has achieved the best comprehensive performance on I-880 data set. Besides that, the performance of KNN is very close to the proposed method. As the noise data in I-880 data set has been processed well, this data set cannot recognize whose performance is much better between the performance of KNN and the proposed method. In order to evaluate the performance well, we use a data set including noise data, PeMS data set, to perform the experiments.

The experimental results on PeMS data set have been presented in Fig. 3 and Table 2. These results indicate that the proposed method obtains the best performance, as it has the highest value on PI. Because the propose method has the smallest variance on each criterion, the robustness of the proposed method is the best. Also, we can see that the robustness of SVM and Naive Bayes is much worse than KNN and the proposed method. This can be indicated by Fig. 3. Meanwhile, we find that the performance of the propose method is much better than KNN method. As the propose method combines KNN and SVM models to detect traffic incidents, it has the advantages of KNN and SVM simultaneously, and can make the performance more robust. Thus the propose method can overcome the influence arisen by the noise data well.

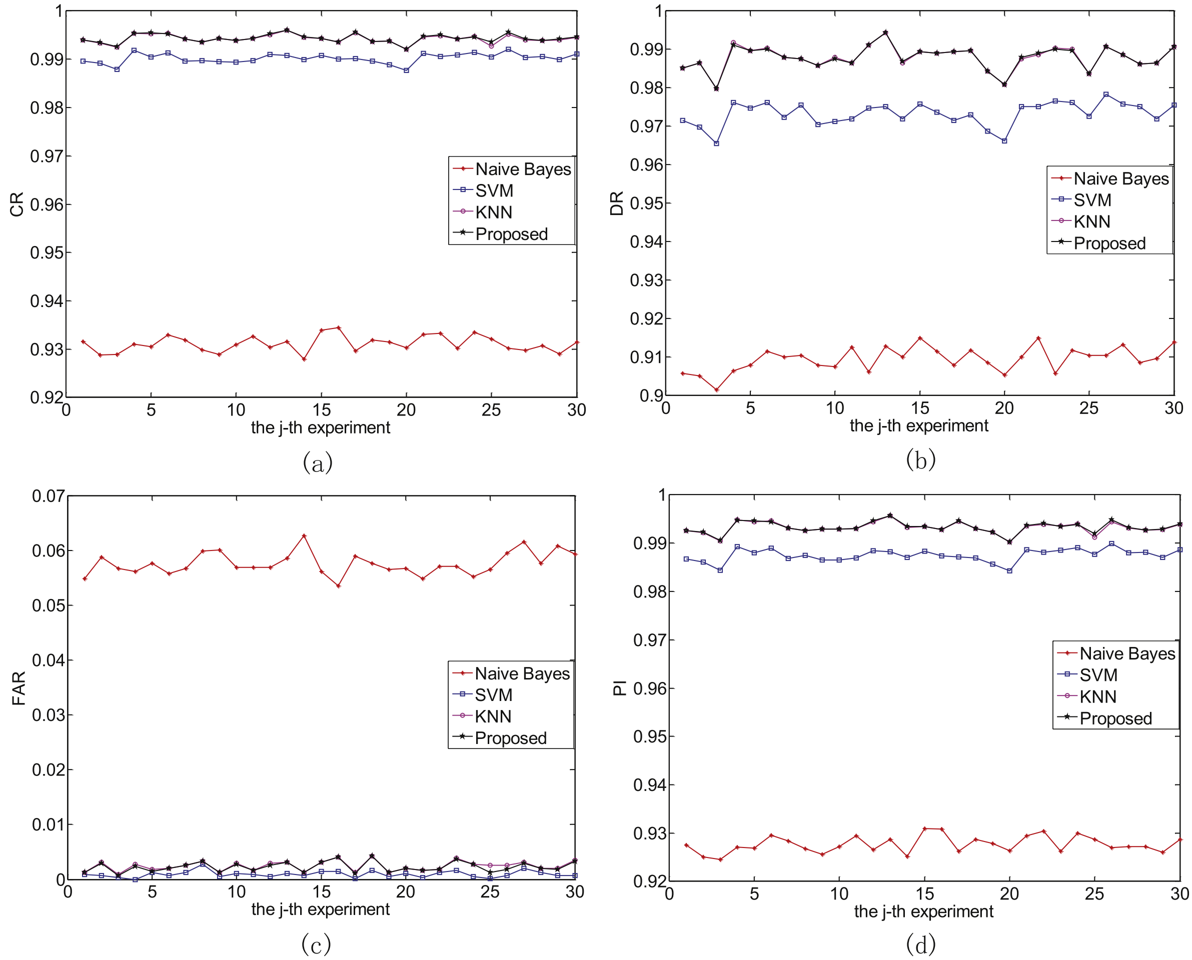

Fig. 2. Experimental results of Naive Bayes, SVM, KNN, and the proposed method on I-880 data set: performance on (a) CR, (b) DR, (c) FAR, (d) PI.

Table 1
Experimental results of all four algorithms based on I-880 data set (the performances are presented with the form *average* ± *variance*).

Algorithm	CR(%)	DR(%)	FAR	PI
Naive Bayes	93.11 ± 0.17	90.95 ± 0.32	5.76 ± 0.21	92.77 ± 0.17
SVM	99.02 ± 0.10	97.33 ± 0.31	0.10 ± 0.06	98.75 ± 0.13
KNN	99.42 ± 0.09	98.79 ± 0.31	0.25 ± 0.09	99.32 ± 0.12
Proposed	99.43 ± 0.09	98.78 ± 0.31	0.23 ± 0.09	**99.33 ± 0.12**

Table 2
Experimental results of all four algorithms based on PeMS data set (the performances are presented with the form *average* ± *variance*).

Algorithm	CR(%)	DR(%)	FAR	PI
Naive Bayes	82.94 ± 0.63	49.45 ± 3.33	10.72 ± 0.61	73.89 ± 1.24
SVM	89.68 ± 0.38	52.49 ± 4.31	3.27 ± 0.71	79.63 ± 1.32
KNN	88.48 ± 0.36	59.54 ± 1.84	6.03 ± 0.42	80.66 ± 0.65
Proposed	88.64 ± 0.36	59.64 ± 1.85	2.86 ± 0.44	**80.81 ± 0.64**

5. Conclusions

This paper proposes a new ensemble learning method in traffic incident detection. It employs the individual SVM and KNN models firstly, and then it takes one ensemble learning strategy to combine them for better final output. The experimental results show that the propose method has achieved the best performance among all the compared methods. More importantly, the ensemble learning strategy has also improved the robustness of the individual models.

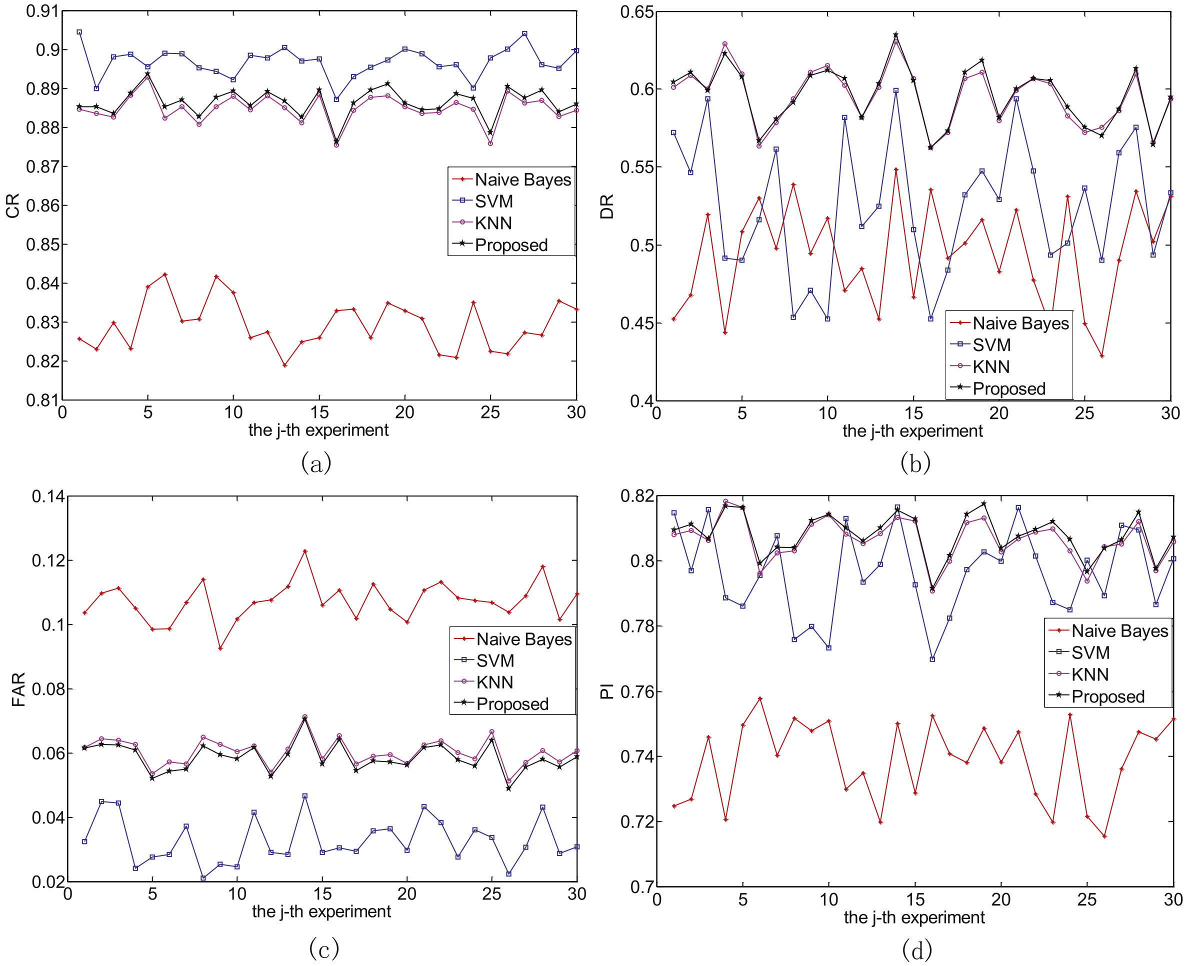

Fig. 3. Experimental results of Naive Bayes, SVM, KNN, and the proposed method on PeMS data set: performance on (a) CR, (b) DR, (c) FAR, (d) PI.

Acknowledgment

This work is supported by China NSFC Program under Grant NO. 61603257.

Conflicts of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] B. Debrabant, U. Halekoh, W.H. Bonat, D.L. Hansen, J. Hjelmborg, J. Lauritsen, Identifying traffic accident black spots with poisson-tweedie models, Accid. Anal. Prev. 111 (2018) 147–154.
[2] J.B. Sheu, A composite traffic flow modeling approach for incident-responsive network traffic assignment, Physica A 367 (2006) 461–478.
[3] J. Xiao, Y. Liu, Traffic incident detection using multiple-kernel support vector machine, Transp. Res. Rec.: J. Transp. Res. Board (2324) (2012) 44–52.
[4] J. Xiao, X. Gao, Q.J. Kong, Y. Liu, More robust and better: a multiple kernel support vector machine ensemble approach for traffic incident detection, J. Adv. Transp. 48 (7) (2014) 858–875.
[5] J. Xiao, Y. Liu, Traffic incident detection by multiple kernel support vector machine ensemble, in: Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on, IEEE, 2012, pp. 1669–1673.
[6] J. Ren, Y. Chen, L. Xin, J. Shi, B. Li, Y. Liu, Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment, IET Intel. Transport Syst. 10 (6) (2016) 428–437.
[7] Y. Asakura, T. Kusakabe, L.X. Nguyen, T. Ushiki, Incident detection methods using probe vehicles with on-board gps equipment, Transp. Res. C 81 (2017) 330–341.
[8] E. D'Andrea, F. Marcelloni, Detection of traffic congestion and incidents from gps trace analysis, Expert Syst. Appl. 73 (2017) 43–56.
[9] R. Weil, J. Wootton, A. Garcia-Ortiz, Traffic incident detection: sensors and algorithms, Math. Comput. Model. 27 (9–11) (1998) 257–291.
[10] D.H. Lee, S.T. Jeng, P. Chandrasekar, Applying data mining techniques for traffic incident analysis, J. Inst. Eng. 44 (2) (2004).

[11] L. Sun, Z. Lin, W. Li, Y. Xiang, Freeway incident detection based on set theory and short-range communication, Transp. Lett. (2018) 1–12.
[12] A.B. Nikolaev, Y.S. Sapego, A.M. Ivakhnenko, T.E. Mel'nikova, V.Y. Stroganov, Analysis of the incident detection technologies and algorithms in intelligent transport systems, Int. J. Appl. Eng. Res. 12 (15) (2017) 4765–4774.
[13] M. Chong, A. Abraham, M. Paprzycki, Traffic accident analysis using machine learning paradigms, Informatica 29 (1) (2005).
[14] S. Chen, W. Wang, H. Van Zuylen, Construct support vector machine ensemble to detect traffic incident, Expert Syst. Appl. 36 (8) (2009) 10976–10986.
[15] S.G. Ritchie, R.L. Cheu, Simulation of freeway incident detection using artificial neural networks, Transp. Res. C 1 (3) (1993) 203–217.
[16] F. Yuan, R.L. Cheu, Incident detection using support vector machines, Transp. Res. C 11 (3–4) (2003) 309–328.
[17] T.G. Dietterichl, Ensemble learning, 2002.
[18] M.M. Bejani, M. Ghatee, A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data, Transp. Res. C 89 (2018) 303–320.
[19] P. Soucy, G.W. Mineau, A simple knn algorithm for text categorization, in: Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on, IEEE, 2001, pp. 647–648.
[20] C. Cortes, V. Vapnik, Support vector machine, Mach. Learn. 20 (3) (1995) 273–297.
[21] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.
[22] H. Huang, M. Abdel-Aty, Multilevel data and bayesian analysis in traffic safety, Accid. Anal. Prev. 42 (6) (2010) 1556–1565.
[23] K.F. Petty, H. Noeimi, K. Sanwal, D. Rydzewski, A. Skabardonis, P. Varaiya, H. Al-Deek, The freeway service patrol evaluation project: database support programs, and accessibility, Transp. Res. C 4 (2) (1996) 71–85.
[24] T. Choe, A. Skabardonis, P. Varaiya, Freeway performance measurement system: operational analysis tool, Transp. Res. Rec.: J. Transp. Res. Board (1811) (2002) 67–75.