# One-Class Learning Based Algorithm for the Freeway Automatic Incident Detection

*Zhiyong Liu, Menghua Zhu and Keqing Fan,*

**Informatics School, Wuyi University, Jiangmen City, Guangdong Province, P. R. China**

**Summary**

According to the characteristic of the freeway, choosing three the most sensitive traffic parameters of AID(Automatic Incident Detection) as the feature vectors, an one-class classification based AID algorithm is proposed. The method establishes the regional distribution model of learning samples, and constructs the decision function in the feature space. When the detected data fall into the inner of the decision region, it will be judged as no incident happened, or else incidents happened. This method does not require any transcendental statistical hypothesis about the distribution of samples, even if the distribution of samples is non-convex and unconnected, it can gain better decision function. The experiment results show that the new algorithm can obtain a higher IR, and limit FIR effectively, so it is a new and potential method for AID.

***Key words:***

*AID ; One-class Classification ; SVM ; Freeway*

## 1. Introduction

Sudden incidents of the freeway commonly refer to the accidents such as impact each other, scratch, fire, turn over or explode and etc. happened in running vehicles on the freeway. These unpredictable traffic incidents will bring great expense, and affect badly normal running of the freeway. When sudden incidents happen on a section of freeway, during a short time, the traffic jam will occur, and the traffic speed will turn slow, and the traffic flow density will exceed the normal average value on the upstream; at the downstream, traffic flow will evacuate, and the traffic flow density will be less than the normal average value. This will lead traffic capacity to descend rapidly. If the capacity descends to less than traffic demand, there will come forth the accidental traffic jam, to lead to traffic paralysis.

Since later several years, many scholars have put forward using theory of artificial neural networks(ANN)[1] to detect traffic incidents with development of this theory. However, the application of this method on real-time detecting for traffic incidents will be limited because the ANN method will meet some obstacles such as slow learning speed and local convergence for mass traffic data. For traffic incidents detection of the freeway, the frequency of incidents occurring is generally relative small. There is no incident during the most of time, that is, for the sample data, a lot of samples will centralize into a tight region, which is called as normal samples; but a few samples of incident will scatter around the tight region according to the characteristic when the incidents occur. In order to detect the incidents, we need a method that can better classify these two kinds of samples. It may be found through investigation that one-class classification based on SVM can perfectly solve this problem among large numbers of classification methods. The method establishes the regional distribution model of learning samples, and constructs the decision function in the feature space. When the detection data fall into the inner of the decision region, it will be judged as no incident happened, or else incidents happened. This method does not require any transcendental statistical hypothesis about the distribution of samples, even if the distribution of samples is non-convex and unconnected, it can gain better decision function, and there is no problem of local convergence in the learning process. For problems of false negative and false positive when detecting incidents, it only need make decision function cover learning samples as tightly as possible.

## 2. One-class classification of incident detection

In one-class classification of incident detection on the freeway, the eigenvector may be selected according to the characteristic of each traffic parameter. Suppose that selected eigenvector $\boldsymbol{x}$ is denoted as $\boldsymbol{x} = [x_1, x_2, \cdots\cdots, x_s]$. It has been proved that samples set $\{\boldsymbol{x}_i\}$, $i = 1, 2, \cdots, N$ corresponding to normal conditional distributes in a quite tight region of special region $R^S$. What is called one-class classification is to find a cover $C(\boldsymbol{x})$ of the distribution region of the dot set $\{\boldsymbol{x}_i\}$, and construct a decision function $f(\boldsymbol{x})$, for an arbitrary eigenvector $x$, we have

$$\begin{cases} f(\boldsymbol{x}) \le 0 & \boldsymbol{x} \in \text{no incident happens} \\ f(\boldsymbol{x}) > 0 & \boldsymbol{x} \in \text{incidents happen} \end{cases} \tag{1}$$

That is, sample which falls into $C(\boldsymbol{x})$ is judged as no incident happened, or else is judged as incidents happened.

Manuscript received October 5, 2006.
Manuscript revised October 25, 2006.

Consider the situation that $C(\boldsymbol{x})$ is a hypersphere on $R^S$: suppose that the radius is $R$ and the epicenter of hypersphere lies on $\boldsymbol{a}$, then decision function is:

$$f(\boldsymbol{x}) = <(\boldsymbol{x} - \boldsymbol{a}) \cdot (\boldsymbol{x} - \boldsymbol{a})> - R^2 \tag{2}$$

Where, <·> means dot product of vectors. The covering error of $C(\boldsymbol{x})$ is defined as: if the sample $\boldsymbol{x}_i$ falls into $C(\boldsymbol{x})$, then the error is 0; if the sample falls outsides $C(\boldsymbol{x})$, then the error is the distance between the sample $\boldsymbol{x}_i$ and spherical surface. That is:

$$\xi(\boldsymbol{x}_i) = \begin{cases} 0 & \|\boldsymbol{x}_i - \boldsymbol{a}\| - R \le 0 \\ \|\boldsymbol{x}_i - \boldsymbol{a}\| - R & \|\boldsymbol{x}_i - \boldsymbol{a}\| - R > 0 \end{cases} \tag{3}$$

Apparently, that probability of false negative and false positive is as small as possible means the samples covered by $C(\boldsymbol{x})$ are as much as possible, but the volume of hypersphere should be as small as possible. It can be depicted as a constrained optimization problem, which is[3]:

$$\begin{aligned} &\min\ R^2 + c\sum_i \xi_i \\ &\text{s.t.} \quad <(\boldsymbol{x}_i - \boldsymbol{a}) \cdot (\boldsymbol{x}_i - \boldsymbol{a})> \le R^2 + \xi_i \\ &\qquad\ \ \xi_i \ge 0 \end{aligned} \tag{4}$$

Where, $\sum_i \xi_i$ is the learning error, $c$ is a weight coefficient, which gives the tradeoff between the volume surrounded by decision region and the learning errors.

For constrained optimization problem given by formula (4), introduce Lagrange multipliers $\{\beta_1, \beta_2, \cdots, \beta_S\}$, and by using K-T(Kuhn Tucker) conditions the constrained optimization problem can be written as:

$$\left.\begin{aligned} &\min \sum \beta_i <\boldsymbol{x_i} \cdot \boldsymbol{x_i}> - \sum_i \sum_j \beta_i \beta_j <\boldsymbol{x_i} \cdot \boldsymbol{x_j}> \\ &\text{s.t.} \ \ \sum_i \beta_i = 1 \qquad 0 \le \beta_i \le c \end{aligned}\right\} \tag{5}$$

In the formula above, $\beta_i$ is Lagrange multiplier of the corresponding sample $x_i$. If $\beta_p$ of the corresponding sample $\boldsymbol{x}_p \in \{\boldsymbol{x}_i\}$ satisfies $0 < \beta_p < c$, then $\boldsymbol{x}_p$ exactly lies on spherical surface. These samples make up of the support of $C(\boldsymbol{x})$, called "Support Vector Set" of $C(\boldsymbol{x})$, which is marked as $\{\boldsymbol{x}_p\}$. It can be proved that center $\boldsymbol{a}$ and radius $R$ of $C(\boldsymbol{x})$ can be written as

$$\begin{cases} \boldsymbol{a} = \sum_i \beta_i \cdot \boldsymbol{x}_i \\ R = \|\boldsymbol{x}_k - \boldsymbol{a}\|, \quad \boldsymbol{x}_k \in \{\boldsymbol{x}_p\} \end{cases} \tag{6}$$

The cover $C(\boldsymbol{x})$ obtained by the way above is a hypersphere. If the set $\{\boldsymbol{x}_i\}$ does not obey the "sphericity" distribution, then it implies that $C(\boldsymbol{x})$ is not "tight". Therefore, suppose that there is a mapping $\Phi$: $R^S \to R^L$, which will map $\{\boldsymbol{x}_i\}$ to the "sphericity" distribution $\{\Phi(\boldsymbol{x}_i)\}$ on $R^L$, then learning process of formula (5) carried on $R^L$ will get a more rational decision function. Because it only involves dot product of variables in the formula (2), (4) and (5), a kernel function $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = <\Phi(\boldsymbol{x_i}) \cdot \Phi(\boldsymbol{x_j})>$ is introduced to replace dot product $<\boldsymbol{x_i} \cdot \boldsymbol{x_j}>$ on $R^S$ in order to obviate the difficulty of determining the mapping $\Phi(\cdot)$ on $R^L$ directly. Then the new decision function can be described as:

$$\begin{aligned} f(\boldsymbol{x}) = k(\boldsymbol{x}, \boldsymbol{x}) - 2\sum_i \beta_i k(\boldsymbol{x}, \boldsymbol{x}_i) - k(\boldsymbol{x_k}, \boldsymbol{x_k}) \\ + 2\sum_i \beta_i k(\boldsymbol{x_k}, \boldsymbol{x}_i) \end{aligned}$$

(7)

$\{\Phi(\boldsymbol{x}_p)\}$ is support vector on $R^L$, $\boldsymbol{x}_k \in \{\boldsymbol{x}_p\}$. $C(\boldsymbol{x})$ gained from this way is hypersphere on $R^L$, and its shape on $R^s$ is determined by the choice of kernel function $k(\boldsymbol{x_i}, \boldsymbol{x_j})$.

## 3. Algorithm for incident detection

One-class classification based incident detection algorithm for the freeway can be described as follows:

**Step1.** Select kernel function $k(\boldsymbol{x_i}, \boldsymbol{x_j})$; According to the conclusion from the equation (7), a radial basis function is chosen as the kernel function:

$$k(\boldsymbol{x_i}, \boldsymbol{x_j}) = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}} \tag{8}$$

Now, let $k(\boldsymbol{x}_i, \boldsymbol{x}_j) \mid_{i=j} = 1$, the equation (7) boils down to[6]:

$$f(\boldsymbol{x}) = 2\sum_i \beta_i [k(\boldsymbol{x_i}, \boldsymbol{x_k}) - k(\boldsymbol{x_i}, \boldsymbol{x})] \tag{9}$$

That is, on the condition of adopting radial basis function, $f(\boldsymbol{x})$ is linear combination of radial basis function. When the learning error $\sum_i \xi_i$ is 0, the character of $f(\boldsymbol{x})$ can be controlled and the volume of $C(\boldsymbol{x})$ can be changed through adjusting $\sigma$.

**Step2** Obtain the data $y_1, y_2, \cdots, y_N$ of upstream and downstream current traffic parameters from each vehicle detector terminal in the detecting area;

**Step3** Obtain the average data $\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_N$ of each downstream traffic parameters from each vehicle detector terminal in the detecting area on the normal condition;

**Step4** According to normal data $\bar{y}_1, \bar{y}_2, \cdots, \bar{y}_N$ and experiential threshold, judge whether the obtained data from each vehicle detector terminal is reasonable. If the data are reasonable, save these data as effective data; or else, delete them as error data;

**Step5** For the effective data obtained from each vehicle detector terminal in step4, transform $N$ traffic parameters, and take $S$ transformed traffic variables $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_S$ as feature values of the eigenvector $\boldsymbol{x}$. Determine and pick up sample data;

**Step6** According to equations (8) and (9), sample at random; determine the weight $c$ in (2) and select the appropriate coefficient $\sigma$ in (8), train one-class classifier for each vehicle detector terminal;

**Step7** Judge if incidents happen on the freeway according to new sample data from each vehicle detector terminal. If incidents occur, then response and deal with the incidents; otherwise go to step 2.

## 4. Instance

It is showed from researches that the occurrence of traffic incident will affect the change of traffic parameters, among which upstream and downstream occupancy, traffic volume and average vehicle speed are the most obvious on changing trend[8]. When the incident happens, during a short time, the traffic jam will occur, and the traffic speed will turn slow, and the occupancy will ascend on the upstream. The traffic flow will evacuate, and the occupancy will descend on the downstream. According to the rationale, this paper selects three time-varying traffic parameters (such as traffic volume, occupancy, and average vehicle speed) as feature variable. For accommodating the data processing capability of the computer, we use 5-minute traffic volume, occupancy and average traffic speed in the traffic data from vehicle detector terminals. So, we may define following variables:

(1) The difference between upstream 5-minute traffic volume and downstream 5-minute traffic volume before $t$ time: $VOLDEV_t = VOL_{i-1,t} - VOL_{i,t}$

(2) The difference between upstream 5-minute occupancy and downstream 5-minute occupancy before $t$ time: $OCCDEV_t = OCC_{i-1,t} - OCC_{i,t}$

(3) The difference between upstream and downstream 5-minute average vehicle speed before $t$ time: $AOSDEV_t = AOS_{i-1,t} - AOS_{i,t}$

In this paper, 2140 groups of valid data on the spot gathered from the detection station of a section on the Kaiyang freeway are adopted to train samples for one-class classification. After data preprocessing according to the experiential threshold, we can obtain 2129 groups of data for no incident happening (show as table 1) and 11 groups of data for incident happening (show as table 2).

Based on preprocessed data, a model is established by using MATLAB for 3-dimensional eigenvector adopted by

Table 1: Preprocessed data for no incident happening

| | $VOL_{i-1,t}$ | $AOS_{i-1,t}$ | $OCC_{i-1,t}$ | $VOL_{i,t}$ | $AOS_{i,t}$ | $OCC_{i,t}$ | $VOLDEV_t$ | $OCCDEV_t$ | $AOSDEV_t$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 58 | 80.21 | 65.08 | 53 | 77.43 | 61.6 | 5 | 3.48 | 2.78 |
| 2 | 59 | 81.65 | 65.03 | 50 | 83.54 | 53.87 | 9 | 11.16 | 1.89 |
| 3 | 58 | 71.98 | 72.52 | 55 | 78.06 | 63.41 | 3 | 9.11 | 6.08 |
| 4 | 50 | 76.44 | 58.87 | 52 | 80.71 | 57.99 | 2 | 0.88 | 4.27 |
| … | … | … | … | … | … | … | … | … | … |
| 2126 | 60 | 78.21 | 69.04 | 54 | 81.38 | 59.72 | 6 | 9.32 | 3.17 |
| 2127 | 54 | 78.60 | 61.83 | 51 | 83.17 | 55.19 | 3 | 6.64 | 4.57 |
| 2128 | 51 | 79.25 | 57.92 | 52 | 80.44 | 58.18 | 1 | 0.26 | 1.19 |
| 2129 | 43 | 76.94 | 50.29 | 45 | 83.01 | 48.79 | 2 | 1.5 | 6.07 |

Table 2: Preprocessed data for incident happening

| | $VOL_{i-1,t}$ | $AOS_{i-1,t}$ | $OCC_{i-1,t}$ | $VOL_{i,t}$ | $AOS_{i,t}$ | $OCC_{i,t}$ | $VOLDEV_t$ | $OCCDEV_t$ | $AOSDEV_t$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 60 | 71.24 | 75.80 | 53 | 85.64 | 55.70 | 7 | 20.10 | 14.4 |
| 2 | 63 | 74.68 | 75.92 | 47 | 89.17 | 47.44 | 16 | 28.49 | 14.49 |
| 3 | 67 | 70.51 | 85.52 | 41 | 80.28 | 45.96 | 26 | 39.56 | 9.77 |
| 4 | 69 | 73.35 | 84.66 | 49 | 83.51 | 52.81 | 20 | 31.85 | 10.16 |
| … | … | … | … | … | … | … | … | … | … |
| 8 | 71 | 70.12 | 91.13 | 50 | 83.57 | 53.85 | 21 | 37.28 | 13.45 |
| 9 | 64 | 72.47 | 79.48 | 59 | 80.94 | 65.61 | 5 | 13.88 | 8.47 |
| 10 | 58 | 68.18 | 76.56 | 47 | 81.03 | 52.20 | 11 | 24.36 | 12.85 |
| 11 | 70 | 74.7 | 84.34 | 54 | 85.42 | 56.90 | 16 | 27.44 | 10.72 |

training. In order to select the appropriate parameters, and considering the computing time, this paper draws out 70 samples randomly from 2129 groups of data for no incidents happening, hold $c$ =1, let $\sigma$ = 0.5 to 1 increased by 0.01 for samples learning. When $\sigma$ = 0.84, the decision region can fit better the distributing region of samples. The projection maps of learning outcome show as Fig. 1 and Fig. 2. When unknown samples from this vehicle detector terminal fall into the inner of decision region, it will be judged as no incident happened, or else incidents happened in this detection area.

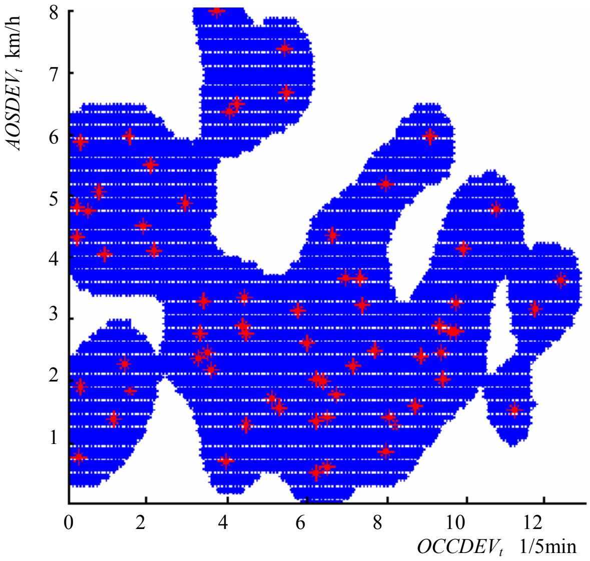


Fig. 1 Maps of relationship between $AOSDEV_t$ and $OCCDEV_t$

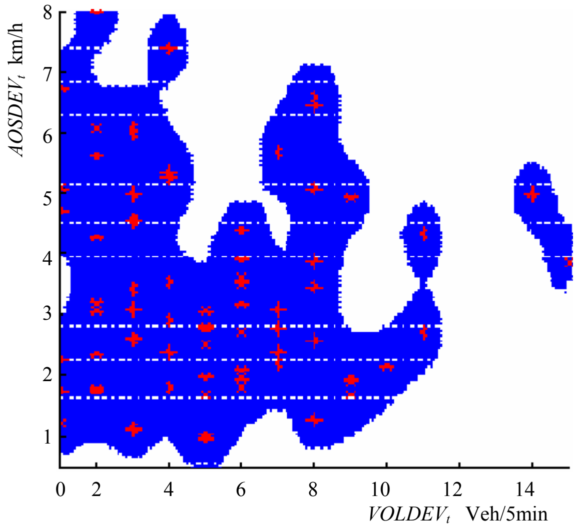


Fig. 2 Maps of relationship between $AOSDEV_t$ and $VOLDEV_t$

In general, performance index for evaluating AID algorithms includes identification rate (IR), false identification rate (FIR) and mean time to identification (MTTI). However, there are a contradiction between IR and FIR. The incident detection with high IR must need high sensitivity, but this will lead to the increasing of misinformation times. In the incident detection on the freeway, the number of detected data is quite big, so we will meet the real-time problem for data processing. Considering the practicability of algorithm, this paper gives the tradeoff between these contradictions, that is, on the precondition of ensuring high IR, reduces FIR of incident detecting, enhances the capacity of real-time processing.

This paper draws out 136 samples randomly from 2129 groups of data for no incidents happening, and appends 11 groups of data for incidents happening to compose 147 test samples. By using the AID algorithm based on one-class classification, we can obtain: IR=90.09%, FIR=1.47%, MTTI=0.29 min. Comparing with other AID algorithms, the new algorithm can get higher IR and lower FIR, MTTI is controlled well. The comparison is showed as table 3.

Table 3 The outcome compared new algorithm proposed in this paper with other algorithms

| Algorithms | Performance index | | |
|---|---|---|---|
| | IR | FIR | MTTI |
| California algorithm 7 | 78 | 6.33 | 1.12 |
| Dual-section Mc Master algorithm | 87.0 | 8.17 | 0.68 |
| Exponential smoothing algorithm | 92.0 | 1.87 | 0.7 |
| Neural networks algorithm | 91.0 | 5.41 | 0.84 |
| new algorithm proposed in this paper | 90.09 | 1.47 | 0.29 |

## 5. Conclusion

Having studied the characteristic of each traffic parameter changing when incidents happen on the freeway, this paper introduces one-class classification based incidents detection method, and adopts the transform of three traffic parameters which are the highest sensitivity to incident as the eigenvalue to train the samples according to data on the spot, and calibrates the decision module. A tradeoff method is given in the paper to deal with the contradiction between IR and FIR. The experiment results show that the new algorithm can obtain a higher IR, and limit FIR effectively, so it is a new and potential method for AID.

## References

[1] JIANG Zifeng, LIU Xiaokun. Artificial Neural Network (ANN) Algorithm for Traffic Incidents Detection[J]. Journal of Xi'an Highway University, 2000,20(7):67-69

[2] ZHOU Wei, LUO Shigui. Incident Detection Algorithm Based on Fuzzy Set Theory [J]. Journal of Xi'an Highway University, 2001, 21(4): 70-73

[3] DavidM.J.Tax , Robert P.W.Duin. Data Domain Description using Support Vector[C]. Proceedings of

European Symposium on Artificial Neural Network, Brugge, 1999: 251-256

[4] FAN Keqing, Ni Yiqing, GAO Zanming. Bridge Condition Monitoring Approach using SVM Based Novelty Detection Algorithm. Journal of highway and transportation research and development. 1999:251-256

[5] R. Fletcher. Practical Methods of Optimization. John Wiley and Sons, New York, 1987

[6] Campbell, C. and Bennett, K.P. A linear programming approach to novelty detection. Adv. Neural Inf. Process. Syst., 2001, 14: 395-401

[7] A. Schölkopf, R. Williamson, A. Smola, J.S. Taylor, J. Platt, Support vector method for novelty detection, in: S.A. Solla, T.K. Leen, K.R. Müller (Eds.), Neural Information Processing Systems, Elsevier, New York, 2000, pp. 582-588

[8] LIU Zhiyong. Intelligent Traffic Control Theory And Applications. Beijing: Science Press, 2003

[9] David M. J. Tax, Robert P. W. Duin. Uniform Object Generation for Optimizing One-class Classifiers. Journal of Machine Learning Research,2001(2), pp. 155-173

**Zhiyong Liu** received the B.S., M.S. and Ph.D. degrees in Systems Engineering from Xi'an Jiaotong University in 1983, 1986 and 1989, respectively. His research fields are: intelligent control, intelligent transportation systems, patterns recognition and intelligent systems. He now is Professor in Informatics School, Wuyi University.