Automatic Incident Detection Using Machine Learning

Daehyon Kim*

Machine Learning을 이용한 자동 돌발상황검지

김 대 현*

ABSTRACT : Incidents on the freeway disrupt traffic flow and the cost of delay caused by incidents is significant. To reduce the impact of an incident, a traffic management center needs to quickly detect and remove it from the freeway. Quick and efficient automatic incident detection has been a main goal of the transportation research for many years. Also many algorithms based on loop detector data have been developed and tested for the Automatic Incident Detection(AID). However, many of them have a limited success in their overall performance in terms of detection rate, false alarm rate, and the mean time to detect an incident. Until recently, the neural network models have been the one of the popular and efficient approach for real-time automatic incident detection and many researches have shown that the neural network models were much more efficient than various other previous models. The purpose of this research is to propose a more efficient and accurate model than the neural network model in the automatic incident detection problem. For this purpose, a machine learning model, Support Vector Machine (SVM) learning which is based on the statistical learning theory, has been used in this paper. The experiments have been done with real world freeway data, and the results show that the SVM could provide better performance in terms of DR(Detection Rate) and FAR(False Alarm Rate) than Backpropagation which is the most popular neural network model.

Key words : automatic incident detection(AID), machine learning, support vector machine(SVM), backpropagation

요약 : 단속류 및 연속류의 도로상에서 발생하는 돌발상황은 심각한 교통혼잡을 야기할 뿐만 아니라 2차적 교통사고로 이어질 수 있으며, 이로 인해 매우 큰 사회적비용을 초래한다. 따라서 교통관리센타에서는 예측 불가능한 돌발상황에 신속하고 효과적인 대응을 하기 위해서, 돌발상황에 대한 보다 신속하고 정확한 검지가 요구되어 왔다. 특히 연속류 고속도로상에 설치되어있는 검지기에서 수집되는 교통량, 속도, 점유율 등의 교통정보를 활용하여 돌발상황 발생을 자동으로 검지하는 알고리즘에 대한 연구는, 현재 세계적으로 관심이 고조되고 있는 지능형교통체계분야(ITS)의 주요한 연구분야로 인식되고 있으며, 지금까지 다양한 형태의 알고리즘들이 개발되어왔다. 그러나 지금까지 개발된 기존의 많은 알고리즘들은 낮은 검지율 및 높은 오검지율을 보여 왔으며, 따라서 보다

* Assistant Professor, Division of Transportation & Logistics System Engineering, Yosu National University(여수대학교, 교통물류시스템공학부, 교통공학전공 조교수)

신뢰성이 높은 돌발상황에 대한 검지알고리즘의 개발이 절실히 요구된다. 현재까지 개발된 돌발상황검지를 위한 알고리즘 중 가장 신뢰성이 높고 실시간 검지에 적합한 모형으로 신경망모형이 인식되고 있다. 본 연구에서는 기존의 신경망모형을 이용한 알고리즘보다 높은 검지율 및 낮은 오검지율을 갖는 보다 우수한 실시간 돌발상황검지 알고리즘을 제시하고자 한다. 이를 위해 최근에 패턴인식분야에서 몇몇 연구자들에 의해 전통적인 신경망 모형인 Backpropagation 모형보다 우수한 것으로 평가되고 있는 SVM(Support Vector Machine)을 사용하였다. 본 연구에서는 서울시 내부고속도로에서 수집한 교통상황 데이터 및 돌발상황에 대한 데이터를 이용하여 두 모형, 즉 Backpropagation 과 SVM에 대한 검지능력을 검지율 및 오검지율 측면에서 비교 검토하였으며, 연구결과 SVM 모형이 기존의 Back- propagation 모형보다 돌발상황검지에 더욱 우수한 것으로 나타났다.

주제어 : 자동돌발상황검지, backpropagation, support vector machine(SVM)

Ⅰ. Introduction

Incident is non-recurrent event that causes a severe reduction in the capacity or an abnormal increase in the demand of a transportation facility. The function of Automotive Incident Detection (AID) is to automatically identify the occurrence of unpredictable incidents that effect the capacity of freeways so that appropriate response and clearance procedures can be executed to minimize the effects of the incident on traffic operation. Since the 1970s, there has been growing interest in the incident detection, and a variety of algorithms have been developed(Dudek et al., 1974; Ahmed and Cook, 1982; Busch and Fellendorf, 1990; Chassiakos and Stephanedes, 1993). Unfortunately, the algorithms developed to date have met with only limited operational success, and it is clear that improved algorithms are needed to make loop data based incident detection technology operationally effective. Specifically, existing algorithms have largely unable to maintain the high degree of reliability required in practice(e.g., high detection rate and low false alarm rate). Because of the unreliability and high FAR of previously developed AID systems, many Traffic Information Centers (TICs) are not using them currently.

In order to achieve the better performance on the incident detection, some researches (Ritchie and Cheu, 1993; Payne and Thompson, 1997) have used Artificial Neural Networks which hold considerable potential for recognizing and classifying spatial and temporal patterns in traffic data. The findings of previous researches indicate that neural network models have the potential to achieve significantly better performance in terms of detection rate and false alarm rate, as well as operational improvements in real-time incident detection over more conventional algorithms such as a series of California algorithms and McMaster algorithm(Hall et al., 1993).

There are currently many different types of neural network models available, but the multilayer feedforward with backpropagation learning algorithm, usually called simply Backpropagation(Rumelhart et al., 1986) has been the most popular neural network for the incident detection.

However, the Support Vector Machine (SVM), which is generation learning systems based on advances in statistical learning theory, receives a great deal of attention recently with their remarkable performance. After the SVM was introduced by Vapnik (Vapnik, 1995), it has been successfully applied to numerous pattern recognition problems, including object detection(Blanz et al., 1996), handwritten character recognition (Cortes and Vapnik, 1995; Schölkopf et al., 1996), text categorization(Joachims, 1998), and face detection in images(Osuna et al., 1997). Moreover, SVM has been shown to provide higher performance than other algorithms, such as k-Nearest Neibor(kNN) and neural network model(Bazzani et al., 2001). Recently, Yuan and Cheu(2003) used SVM for incident detection. However, they haven't consider the normalization method for input vectors, even though the normalization method is very important to get the best prediction performance.

The purpose of this study is to propose SVM learning method with the best normalization method for automatic incident detection. The performance of proposed method will be compared with BPMP which is an advanced Backpropagation neural network model because the Backpropagation has been one of popular methods for automatic incident detection until recently. The system performance will be compared in terms of misclassification rate (MCR), detection rate (DR) and false alarm rate (FAR).

Ⅱ. Support Vector Machine (SVM)

1. Linear and separable classification problems

Let the training data $(\mathbf{x}_i, y_i)$, $i=1,, l$, $y_i \in \{\pm 1\}$ $\mathbf{x}_i \in \mathbf{R}^N$, then the support vector algorithm simply looks for the optimal hyperplane with largest margin. This can be formulated as follows:

$$\min \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 \qquad (1)$$

$$\text{s. t. } y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0, \quad i = 1,\dots, l$$

where w is a normal to the hyperplane and b is bias or offset. Using Lagrange multipliers, $\alpha_i \geq 0$, the primal form of the objective function can be:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i (y_i ((\mathbf{x}_i \cdot \mathbf{w}) + b) - 1) \quad (2)$$

The Lagrangian L has to be minimized with respect to the primal variables w and b and maximized with respect to the dual variables α_i. From the Karush-Kuhn-Tucker (KKT) conditions, the derivative of L with respect to the primal variables must be vanish, subject to constraints, $\alpha_i \geq 0$, i.e.,

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, b, \boldsymbol{\alpha}) = 0, \quad \frac{\partial}{\partial b} L(\mathbf{w}, b, \boldsymbol{\alpha}) = 0 \quad (3)$$

Eq. (3) leads to

$$\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i \text{ and } \sum_{i=1}^{l} \alpha_i y_i = 0 \quad (4)$$

Eq. (4) is equality constraints in the dual formulation, and following Eq. (5) which is the Wolfe dual of the optimization problem is given by substituting it into Eq. (2).

The hyperplane decision function can thus be given as

$$f(x) = \operatorname{sgn}((\mathbf{x} \cdot \mathbf{w}) + b) = \operatorname{sgn}(\sum_{i=1}^{l} y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b) \quad (5)$$

2. Soft Margin Hyperplane

In practice, real-world data sets are linearly inseparable in input space since a high noise level causes a large overlap of classes. This means that the constraints need to be relaxed somewhat to allow for the minimum amount of misclassification. The points that subsequently fall on the wrong side of the margin are considered to be errors, and slack variables have been introduced in order to deal with these errors(Cortes and Vapnik, 1995; Vapnik, 1995).

By adding slack variables $\xi_i \geq 0$($i = 1, \ldots l$) to Eq. (1) and (2), the objective function and relaxed constraints are:

$$\min \tau(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} \zeta_i \quad (6)$$

$$\text{s. t. } y_i (\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i, \quad i = 1, \ldots, l$$

where C is upper bound for the Lagrange multiplier, λ_i, i.e., $0 \leq \lambda_i \leq C$. It is the trade-off between maximum margin and classification error and a higher C value will give a large penalty for classification error. The only difference from the separable problem is the upper bound C on the Lagrange multiplier (Schölkopf et al., 1996).

3. Nonlinear classification problems

To handle nonlinear problems, the optimal hyperplane algorithm needs to be augmented. The basic idea is to transform the data from the origin space into another dot product space called the *Feature Space*. This can be achieved by a mapping function Φ:

$$\Phi : \Re^N \rightarrow F$$

In this high dimension space, the data can be linearly separable, hence above linear algorithm can be applied for the problem. Then the training algorithm would only depend on the data through dot products in *F*, i.e., on functions of the form $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$. However, if F is high-dimensional, the dot product will be very expensive to compute. The dot product in the high dimension space can be replace by a kernel function:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$$

By the use of a kernel function, it is possible to compute the separating hyperplane without explicitly carrying out the map into the feature space. A widely used kernels satisfying Mercer's condition(Vapnik, 1995) is as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \qquad (7)$$

Eq. (7) is a Gaussian kernel, usually called *Radial Basis Function*(RBF) kernel and it has been used in this research because it is the most popular and has nonlinear and effective features.

4. Multi-class support vector machines (Multi-class classifiers)

Support vector machines(SVMs) are primarily designed for two-class classification problem. However, the two-class classifier described above can be expanded to the multi-class classifier. There are two main approaches for this, i.e., one-against-all and one-against-one(Weston and Watkins, 1998; Salomon, 2001; Lee et al., 2001; Hsu and Lin, 2002).

The one-against-all approach construct k SVM models where k is the number of classes. The ith classifier is trained with all of the examples in the ith class with positive labels, and all other examples with negative labels. However, the one-against-one approach, which is a popular and simple technique, creates a SVM for all possible combinations of classes. This approach simply constructs all possible two-class classifiers from a training set of k classes. Each classifier is trained on only two out of k classes. Thus, there will be $k(k-1)/2$ classifiers. The one-against-one approach has been used in many researches (Friedman, 1996; Weston and Watkins, 1998; Platt et al., 2000) and shown that its performance is better than the one-against-all approach(Hsu and Lin, 2002). In this research, the one-against-one approach has also been used as many other researches.

5. Normalization for Input Vectors

In Neural Networks, the input vectors should be normalized before using them when the input vectors are large values. Otherwise they can not be categorized properly because of the properties of activation function(Kim, 1999). For the SVM model, normalization of input vectors are also required as in the Neural Networks(Kim, 2004). A loop-based incident detection system typically uses volume, occupancy and speed measured at upstream and downstream detector stations, to detect any incident in between the two stations. In the automatic incident detection system using SVMs, the input vectors should be the normalized values of the above parameters for some time periods. In this research, following normalization method which is detector-by-detector normalization and the formulation for each detector has been used for incident detection problem.

$$\tilde{a}_{pi}^{d} = \frac{a_{pi}^{d} - a_{p\min}^{d}}{a_{p\max}^{d} - a_{p\min}^{d}} \qquad (8)$$

where $a_{p\max}^{d} = \max(a_{pi}^{d}; d = 1, 2 \text{ and } p = 1,\dots,n)$, $a_{p\min}^{d} = \min(a_{pi}^{d}; d = 1, 2 \text{ and } p = 1,\dots,n)$, d is the detector station and $p(p=1,\dots,n)$ are the input patterns. In the equation, $\tilde{a}_{pi}^{d}$ denotes the normalized value of the unit i of input vector. In this research, two detector stations, upstream station and downstream station, have been used and the input vectors for neural networks are combined the normalized three parameters, volume, speed and occupancy on each lane of the two detector stations.

Ⅲ. Experiments and Results

1. Experimental Data Sets

The traffic data were collected from video image processors placed on the Naebu Expressway in Seoul, Korea. The video image processor uses the detection zone approach to emulate loop detectors and provides three traffic parameters, traffic volume, speed and occupancy on every lane of the Expressway. The detectors are spaced at distances about 500 meters for basic section and about 250 meters for tunnel.

The Traffic Information Center (TIC) for the north part of expressways reported 40 incidents for the period between 5 February 2003 and 11 February 2003. For the simulated data, the incident start and end times can be known precisely. However, when the real-world data are being used, the incident start and end times are rarely, if ever, known precisely. The incident times will be reported as the times when the operator detects (or confirms) the occurrence of incidents, and not the times when the incidents actually occur. Therefore, it is necessary to determine the specific 30-second interval that represents the start of an incident.

The 40 incidents recorded by the operators in the log were examined individually and only 31 incidents which are relatively precise for the incident start time were used in this research. For proper evaluation, any test data set should not be included in the training sets. In order to satisfy this requirement and improve the reliability of the experimental results, and thereby overcome shortage of the incident data, the following method was used. The total 31 incident and 200 incident-free data sets were split randomly into two subsets, Data set I and Data set Ⅱ - one for training and the other for testing. In the first experiment, learning is performed on Data set Ⅰ, and network performance is evaluated on the Data set Ⅱ. In the second experiment, learning is performed on Data set Ⅱ, and model performance is evaluated on Data set Ⅰ(see 〈Table 1〉).

Data collected include traffic volume, speed and occupancy on each lane of two stations in 30-second intervals. In addition, the input vectors use each parameter data of every lane during four time intervals, i.e., the current interval t and three previous intervals, (t-1), (t-2), and (t-3). Therefore, input vectors x consisted of 72 units for 3 lanes, 2 stations, 3 traffic parameters, and 4 times intervals.

〈Table 1〉 Number of data sets for training and testing

Data	Training data	Test data	Total
Data Set Ⅰ	16	100	116
Data Set Ⅱ	15	100	115
Total	31	200	231

2. Implementation and Results

For the neural network model, the Backpropagation with Momentum & Prime-offset (BPMP) model, which is one of advanced Backpropagation models, was used since it has been shown to possess better performance than other Backpropagation models such as the standard Backpropagation and Backpropagation with momentum (Kim, 2002). The implementation of this model was for sequential mode learning with a learning rate of 0.01, a momentum of 0.95, and a prime-offset of 0.1. In this research, three-layer network with 144 hidden neurons, 72-144-1 was used and all networks were fully interconnected. Also, the sigmoid function was used for activation function. Training of the network has been stopped when 100% recognition accuracy was achieved on the training set.

It is also noted that Backpropagation is sensitive to the initial value of the weights. In this research, 30 trials were implemented on the same network with the same learning parameters, but with different initial weights that are initialized to random values between +0.5 and -0.5, in order to avoid the effect of initial value of the weights and compare the performance of input vectors more precisely.

For SVMs using RBF kernel, two parameters, C and γ, should be determined beforehand. The parameter of C which is a positive regularization parameter that controls the tradeoff between complexity of the machine and the allowed classification error, and γ is the parameter of the Gaussian kernel determining the width of the kernel function. In this research, a cross-validation method has been used to determine parameters, C and γ. Experiments have been done with $C=1$ and $\gamma=0.03$.

〈Table 2〉 shows the performance comparison of two models, Backpropagation and SVM (Support Vector Machine). With total 31 incident data of Data Set Ⅰ and Data Set Ⅱ, Backpropagation gave average 6.9 errors and SVM gave 5 errors. More importantly, even though Backpropagation gave 0.43 errors with total 200 incident-free data sets, there was no prediction error from SVM.

The experimental results show that the SVM could provide better performance in terms of DR(Detection Rate) and FAR(False Alarm Rate) than Backpropagation which is the most popular neural network model. In the experiments with real world freeway traffic data including incident and incident-free data, Backpropagation gave 79.13 % DR (Detection Rate) and 2.15 % FAR(False Alarm Rate). However, SVM produced 83.87 % DR and zero percent FAR. Even though video image processing technique appears to be very promising for meeting future data collection and surveillance needs, it may be less accurate than the Inductive Loop Detector which is by far the most common form of detector. This implies that the prediction accuracy could be increased if more accurate data obtained using inductive loop detectors have been used in this study.

〈Table 2〉 Performance on the two models

Performance	BP	SVM
Errors on Incident-free data	0.43(0.46)	0
Errors on Incident data	6.47(0.46)	5
Total error	6.9(0.51)	5
Detection Rate(%)	79.13	83.87
False Alarm Rate(%)	2.15	0

Note : The value of ( ) implies variance.

Ⅳ. Conclusion

In this research, Support Vector Machine (SVM) learning which is based on the statistical learning theory has been explored for incident detection. The experiments have been done with real world freeway data, and the results show that the SVM could provide better performance in terms of DR (Detection Rate) and FAR(False Alarm Rate) than Backpropagation which is the most popular neural network model. The SVM approach produced higher DR and lower FAR. The SVM could be more efficient than any other methods for incident detection.

In a practical application, more efficient input vectors should be determined in order to achieve the best performance in automatic incident detection. The input vector could be one of three traffic parameters, volume, speed, and occupancy, or combination of two more parameters. In addition, the proper number of time intervals should also be determined to provide low FAR and high DR. In order to increase the performance of automatic incident detection, a large, accurate data sets on incident and non-incident data should be obtained and Traffic Information Center (TIC) should install more reliable detectors and maintain them continuously to get more accurate data sets.

The SVM learning method may also applicable to other areas of traffic engineering and produce a very good performance. Even though results show that the proposed method in this research could be efficient than others, we may need further proof of validity in order to be useful in various real world fields.

References

Ahmed, S. R. and Cook, A. R., 1982. “Application of time-series analysis techniques to freeway incident detection”. *Transportation Research Record* 841, 19~21.

Bazzani, A., Bevilacqua, A., Bollini, D., Brancaccio, R., Campanini, R., Lanconelli, N., Riccardi, A. and Romani, D., 2001, “A SVM classifier to separate false signals from microcalcifications in digital mammograms”, *Physics in Medicine and Biology*, 46 (6), 1651~1663.

Blanz, V., Schölkopf, B., Bülthoff, H., Burges, C., Vapnik, V. and Vetter, T., 1996, “Comparison of view-based object recognition algorithms using realistic 3D models”, In: von der Malsburg, C., von Seelen, W., Vorbrggen, J. C. and Sendhoff, B. (Eds.) *Artificial Neural Networks ICANN'96*, Berlin, Springer Lecture Notes in Computer Science, Vol. 1112, 251~256.

Busch, F. and Fellendorf, M., 1990, “Automatic Incident Detection on Motorways”, *Traffic Engineering and Control* 31(4), 221~227.

Chassiakos, A. P. and Stephanedes, Y. J., 1993, “Smoothing Algorithms for Incident Detection”, *Transportation Research Record* 1394, 8~16.

Cortes, C. and Vapnik, V., 1995, “Support Vector Networks”, *Machine Learning,* 20, 273~297.

Dudek, C. L., Messer, C. J. and Nuckles, N. B., 1974, “Incident Detection on Urban Freeways”. *Transportation Research Record* 495, 12~24.

Friedman, J., "Another approach to polychotomous classification", Technical report, Department of Statistics, Stanford University

Hall, F. L., Shi, Y. and Atala, G., 1993, "On-Line Testing of the McMaster Incident Detection Algorithm Under Recurrent Congestion". *Transportation Research Record* 1394, 1~7.

Hsu, C.-W. and Lin, C.-J., 2002, "A comparison of methods for multi-class support vector machines", *IEEE Transactions on Neural Networks*, 13, 415~425.

Joachims, T., 1998, "Text Categorization with Support Vector Machine: Learning with Many Relevant Features", *Proceedings of ECML-98*, 10th European Conference on Machine Learning, 137~142.

Kim, D., 1999, "Normalization Methods for Input and Output Vectors in Backpropagation Neural Networks", *International Journal of Computer Mathematics*, 71(2). 161~171.

Kim, D., 2002, "Standard and Advanced Backpropagation Models for Image Processing Application in Traffic Engineering", *ITS Journal*, Vol.7(3-4), 199~211.

Kim, D., 2004, "Prediction Performance of Support Vector Machines on Input Vector Normalization Methods", *International Journal of Computer Mathematics*, 81(5). 547~554.

Lee, Y., Lin, Y. and Wahba, G., 2001, "Multicategory support vector machines", Technical Report, University of Wisconsin.

Osuna, E, Freund, R. and Girosi, F., 1997, "Training Support Vector Machines: an Application to Face Detection", *IEEE Conference on Computer Vision and Pattern Recognition*, 130~136.

Payne, H. J. and Thompson, S. M., 1997, "Development and Testing of Operational Incident Detection Algorithms", Technical Report R-009-97, US Department of Transportation, Washington, DC.

Platt, J., Cristianini, N. and Shawe-Taylor, J., 2000, "Large margin DAGs for multiclass classification", *Advances in Neural Information Processing Systems*. 12, 547~553

Ritchie, S. G. and Cheu, R. L., 1993, "Simulation of Freeway Incident Detection using Artificial Neural Networks", *Transportation Research C* 1(3), 203~217.

Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (Eds), 1986, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. (Vol. 1 & 2), Cambridge, Massachusetts: MIT Press.

Salomon, J., 2001, "Support vector machines for phoneme classification", M.Sc. Dissertation. School of Artificial Intelligence, University of Edinburgh.

Schölkopf, B., Burges, C. and Vapnik, V., 1996, "Incorporating Invariances in Support Vector Learning Machines", in C. von der Malsburg, W. von Seelen, J. C. Vorbrggen, and B. Sendhoff (Eds.), *Artificial Neural Networks ICANN'96*, Berlin, Springer Lecture Notes in Computer Science, Vol. 1112, 47~52.

Yuan, F. and Cheu, R. L., 2003, "Incident detection using support vector machines", *Transportation Research* C 11, 309~328.

Vapnik, V., *The Nature of Statistical Learning Theory*, Springer Verlag, New York, 1995.

Weston, J. and Watkins, C., 1998, "Multi-class support vector machines", Technical Report *CSD-TR-98-04*, Department of Computer Science, Royal Holloway, University of London, UK.

원 고 접 수 일 : 2004년 11월 1일
1차심사완료일 : 2004년 11월 11일
최종원고채택일 : 2004년 11월 18일