Mathematical & Computational Applications, Vol. 6, No. 2, pp. 85-96, 2001

WAVELET-NEURAL NETWORK MODEL FOR AUTOMATIC TRAFFIC INCIDENT DETECTION

Mingyang Wu and Hojjat Adeli
Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210, U.S.A.

Abstract- The problem of traffic incident detection can be viewed as a pattern recognition problem. Neural networks are known to solve pattern recognition problems effectively, especially when there is no mathematical model. The computational complexity of neural network algorithms, however, increases exponentially with an increase in the size of the network. Furthermore, with an increase in the size of the network, the size of the training set has to be increased exponentially in order to achieve the same level of accuracy. To overcome this *double exponential complexity* a hybrid feature extraction algorithm and neural network architecture is created specifically for automatic detection of traffic incidents. The upsteam and downstream traffic data are first filtered by the discrete wavelet transform. Then, a linear discriminant network is used for feature extraction. Finally, the adaptive conjugate gradient learning algorithm of Adeli and Hung is used to train the network.

Keywords- Incident detection, intelligent transportation systems, ITS, neural networks, traffic engineering, wavelets

1. INTRODUCTION

Accurate and fast traffic incident detection is critical for minimizing traffic delays and increasing safety on freeways. With information obtained from the incident detection system, an Intelligent Transportation System (ITS) can use optimal control strategies [3, 26] to guide the traffic flow by routing incoming vehicles from the freeway upstream of the incident and communicating this information to travelers. As such, the development of an efficient, reliable, and robust incident detection algorithm is of paramount importance in creating an ITS.

Existing algorithms, however, are known to be unreliable for generating a high level of false alarms [28]. The process of automatic traffic incident detection consists of two tasks: data collection and incident detection. The data collection task is performed by placing sensors in the freeway. Consequently, the traffic incident detection algorithm depends on the type of sensor used to collect data.

Sensors can be classified by the type of data they provide into four categories [24]:

1. Point data sensors collect data such as occupancy, speed, and volume at a specific location on the freeway. Examples are ultrasonic traffic detectors [17].
2. Area data sensors collect data such as density over a segment of freeway.
3. Point-to-point sensors collect data such as travel time between two specific locations. Examples are vehicle-to-roadside communication sensors.
4. Video sensors collect traffic images. Using traffic scene analysis, the images can be used for automatic traffic incident detection [18]. The image sensors, however, have difficulty covering the entire freeway and their implementation cost is prohibitive.

According to surveys conducted by the Department of Transportation in various states within the United States and Canada [5], traffic information is typically collected from point data sensors (i.e., loop detectors) and includes the occupancy (percentage of road covered by vehicles) and volume (number of vehicles passing per minute) averaged at 20 to 30 second intervals, usually across all the lanes in one direction. On the average, two detectors are used in every mile of the roadway. Only a few existing systems can provide speed information (e.g., Ontario's Queen Elizabeth Way). The goal of this research is to develop a versatile traffic incident detection algorithm to be used widely. As such, the traffic incident detection algorithm presented in this paper uses occupancy and volume data only.

Traffic incident detection has been an active area of research in transportation engineering [10]. A good example of the existing approaches to traffic incident detection up to 1993 can be found in [27]. One of the earliest and the most widely used algorithm is the so-called *California Algorithm* [25]. This is a simple comparative algorithm which relies on the principle that a traffic incident most likely will increase the upstream occupancy significantly while at the same time significantly reduce downstream occupancy. When measured values of occupancy approach pre-selected thresholds, an incident is detected. Such an algorithm includes comparative tests to differentiate between a traffic incident and a bottleneck congestion, compression wave, and random traffic fluctuation.

Other traffic incident detection algorithms developed in the 1970's include the standard normal deviation algorithm, which employs a time-series model to predict short-term traffic [9], and the double exponential algorithm, which makes a double exponential smoothing of traffic occupancy to predict occupancy and identify incidents [7]. Recent work in this area includes the McMaster algorithm which is based on a two-dimensional analysis of the traffic data [13]. Due to the complexity of the traffic phenomena and the fact that no traditional mathematical model can capture the characteristics of various traffic patterns accurately, all of these algorithms suffer from low reliability resulting in a large number of false alarms.

2. TRAFFIC INCIDENT DETECTION USING NEURAL NETWORKS

Neural network computing appears as a promising approach for solving the traffic incident detection problem for several reasons. First, automatic incident detection can be cast as a pattern recognition problem. Neural networks are known to solve pattern recognition problems effectively [2, 22, 23]. Second, neural networks are suitable for solving the problem when there is no mathematical model or explicit rules. Third, neural network classifiers are nonparametric and make no assumption on the shape of the underlying distribution and consequently are superior to statistical classifiers for solving complex multi-dimensional classification problems. Fourth, neural networks are able to handle problems with a large number of input attributes, which most model-based algorithms cannot handle effectively.

A few papers have been published recently on the use of the neural networks approach to solve the traffic incident detection problem [6, 15, 17, 28). These papers employ the occupancy and volume data of upstream and downstream sensing stations and the simple multi-layer perception or backpropagation learning rules [16]. Cheu and Ritchie [6] compare the multi-layer backpropagation algorithm with the self-organizing neural network model of Kohonen [19] and the Adaptive Resonance Theory (ART) classifier of Carpenter and Grossberg [4]. They conclude that the backpropagation algorithm provides better results compared with the other two approaches. However, shortcomings of the backpropagation algorithm have been documented by Adeli and Hung and Haykin (1994), among others. The backpropagation algorithm is known for its simplicity but it suffers from two major and fundamental shortcomings; its convergence rate is very slow, usually requiring thousands of interactions. The convergence rate depends heavily on the learning and momentum ratios that can be chosen only by trial and error (Adeli and Hung, 1994). In addition, the convergence is highly sensitive to the architecture chosen for the network, such as the number of nodes in the hidden layer (Stephanedes and Liu, 1995).

Another problem with existing incident detection attempts is that they do not take into account the large difference between the high incidence probability of the incident training data set and the low incidence probability of the normal situation. This can result in performance degradation. Further, instead of using generic neural network models, substantial improvement in performance can be achieved by carefully analyzing the traffic incident detection problem and then creating a neural network model specifically for it, which is the subject of this paper.

3. THE MAJOR CHALLENGE

As mentioned before, the problem of traffic incident detection can be viewed as a pattern recognition problem. There are two classes to be classified: the incident pattern and the incident-free pattern. The development of an incident detection algorithm is equivalent to the design of a certain pattern recognition classifier. The selection of input attributes (or features) is a key issue for pattern recognition problems which has strong influence on classifier design and system performance. There is a complicated relationship between the number of features selected as inputs of the class and the complexity of the classifier, training efficiency, and system performance. When very few input attributes are selected, the classifier usually is easy to design and the training of the classifier is relatively easy and numerically stable. For example, the *California Algorithm* only selects a few occupancy values of upstream and downstream stations as the input of the incident detection algorithm. As a result, the *California Algorithm* is simple and so is the training process. However, the drawback of selecting too few input attributes is obvious, most of the useful information is lost in the incident detection process resulting in less than satisfactory performance. Some algorithms, such as the standard normal deviation algorithm and the double exponential algorithm, use only the occupancy data. Two-dimensional algorithms, such as the McMaster algorithm, use both the occupancy and volume data but only a small portion of the historical data.

An obvious advantage of neural networks algorithms is that occupancy and volume data are fully employed. However, the large number of input attributes also increases the complexity of network structure and training difficulties, thus affecting the overall performance. The performance and efficiency of neural networks algorithms depend on two main factors: the size of the network and the size of the training set. Computational complexity increases exponentially with an increase in the size of the network. Furthermore, with an increase in the size of the network, the size of the training set has to be increased exponentially in order to achieve the same level of accuracy. This *double exponential complexity* is a major and fundamental challenge for application of neural network to complicated real-life pattern recognition problems in terms of both accuracy and efficiency (CPU time), which is the subject of our research.

The performance (quality and efficiency) of a neural network pattern classifier depends on

a) The quality and quantity of available data. If the traffic data includes the features needed for incident detection, the incident classification is more accurate. In reality, the data is quite noisy. Further, at present there is no consensus on which traffic parameters best describe the traffic dynamics and indicate incident information. Current implementations of sensors provide only a few traffic characteristics such as volume and lane occupancy data. These data may not represent traffic incidents effectively.

b) The size of the network. A large network can provide more reliable results, only if a large amount of data is available and at a very large and often impractical computational cost.

In the papers mentioned in the previous section, typically two volume and two occupancy data were collected at upstream and downstream sensing stations every 30 seconds. The typical input pattern is data collected over a 5-minute period, which amounts to 40 input nodes. In other words, the input space is a 40-dimensional space. In the training set, only a very small fraction of the data points are traffic incidents, the rest are incident-free data. For example, the training data set used by Stephanedes and Liu [28] includes only 31 traffic incidents, where 89 incident patterns were used for training. Unfortunately (fortunately for the traveler!), this very small rate of traffic incidents creates a theoretical and computational challenge for accurate incident detection by the neural network approach, which is being addressed by this research. The formation of an accurate decision boundary in a 40-dimensional input space requires a tremendous training data set, which is impossible to obtain. In other words, the size of the existing traffic incident detection training set precludes accurate generalization. This major shortcoming has to be overcome for a neurocomputing traffic incident detection algorithm to move from an academic exercise to real-life acceptance. To overcome this shortcoming, we present a new hybrid neurocomputing model in the following section.

4. A NEW HYBRID NEURAL NETWORK MODEL FOR TRAFFIC INCIDENT DETECTION

To reduce the dimension of the input space without any significant loss of information, we employ a robust *feature extraction* approach. Feature extraction is generally viewed as a process of mapping the original attributes into more effective features. If a few features can be extracted from the original data, thereby showing significant differences between the normal and the incident situation, a more effective classifier can be designed with better performance and, more importantly, with a lot fewer examples required for training. In other words, feature extraction is used to overcome difficulties caused by the large number of input attributes. The advantages would include substantial reduction in (a) computation complexity (that is, the network becomes easier to train), (b) the size of the network, (c) the required number of training examples (more accurate generalization), and (d) the influence of the noise due to the random nature of traffic.

Generally speaking, there are two main approaches for feature extraction. The first approach employs prior knowledge to separate features that are more important than others for classification; less important features are discarded in the selection process. An example is the double exponential algorithm that filters out the random fluctuation of traffic patterns. Double exponential filtering [7] can be viewed as an approach of feature extraction based on prior knowledge. In our research, discrete wavelet transform [20] is first applied to the raw data and the finest-resolution coefficients (which represent random fluctuations of the traffic patterns, and therefore are not important for incidents detection) are discarded.

The second feature extraction approach is called the *blind* feature extraction approach, which does not employ any prior knowledge. It uses generic criteria independent of the data set at hand. In our research, after filtering by discrete wavelet transform and discarding unnecessary or less important coefficients, a feature extraction network, called linear discriminant analysis network [21], is employed as the second stage feature extractor.

A hybrid feature extraction algorithm and neural network model has been developed in this research using a combination of discrete wavelet transform, linear discriminant analysis network, and the adaptive conjugate gradient learning algorithm of [1]. The upstream and downstream traffic data are first filtered by the discrete wavelet transform. Then, a linear discriminant network is used for feature extraction. Finally, the adaptive conjugate gradient learning algorithm of Adeli and Hung [1] is used to train the network.

5. DISCRETE WAVELET TRANSFORM

Over the past decade or so the wavelet transform has been studied extensively and formalized into a rigorous mathematical framework. Compared with the traditional orthogonal transforms, such as the Fourier transform, the wavelet transform has attractive properties such as time-frequency localization, multi-rate filtering, and scale-space analysis. Consequently, the wavelet transform has been used in a variety of areas including signal-image analysis, nonlinear dynamics, process control, and geophysics [11]. We find the wavelet transform to be effective for extracting important features of traffic patterns. Traffic patterns are characterized by local properties in time. Traffic incidents, as well as other traffic patterns such as traffic pulse, bottlenecks, and compression wave, have different time local properties. The wavelet transform can effectively extract features from different time scales and different resolutions effectively. Most of the unreliability of previous incident detection algorithms is due to their inability to distinguish traffic incidents from other traffic patterns, especially the compression wave. A wavelet transform is used to overcome this insidious problem.

By transforming the original data sequences into different resolutions, prior knowledge can be used to retain the most effective features and discard less effective ones (including noise due to traffic fluctuation and other components irrelevant to the incident detection problem), thereby making the detection algorithm more reliable. We can reduce the influence of the random traffic fluctuation (noise) by using a discrete wavelet transform and ignoring its finest resolution components.

The discrete wavelet transform is applied to four traffic data series: volume and occupancy of both upstream and downstream stations. Any one of these data can be represented by a series x[k] where $k \in Z$ and Z is the set of integers (in this paper square brackets are used to identify series). For any sequence of real numbers $\alpha_{i \in Z}$, the vector space of square-summable sequences is defined as

$$\ell^2(Z) = \left\{ (\alpha_i)_{i \in Z} : \sum_{i=-\infty}^{+\infty} |\alpha_i|^2 < \infty \right\} \tag{1}$$

Denoting the orthonormal wavelet bases of $\ell^2(Z)$ by $\{\phi_{I,\ell}\}_{\ell \in Z}$ and $\{\phi_{j,\ell}\}_{\ell \in Z}$, where j=1, 2, ..., I, I is a positive integer, and the brackets { } denote a set of series. The output of the discrete wavelet transform is the coordinates $s_{(I)}[\ell]$ and $d_{(j)}(\ell)$ of the orthonormal wavelet bases defined as

$$s_{(I)}[\ell] = \langle x(k), \phi_{I,\ell}(k) \rangle \tag{2}$$

and

$$d_{(j)}[\ell] = \langle x(k), \phi_{j,\ell}(k) \rangle \qquad j = 1, 2, \ldots, I. \tag{3}$$

where $\langle , \rangle$ denotes the inner product of two sequences in $\ell^2(Z)$.

The first step to actually perform a discrete wavelet transform and compute the coordinates of the wavelet bases is to construct the orthonormal wavelet bases using the quadrature mirror filter. Next, the coordinates of the wavelet bases are computed using signal convolution and downsampling [29]. Three sequences $d_1[k], d_2[k]$, and $s_2[k]$, which correspond to the wavelet coefficients in Eqs. (2)-(3), are obtained from the discrete wavelet transform. The finest resolution information is denoted by $d_1[k]$; $d_2[k]$

represents the medium resolution information; and $s_2[k]$ is the lowest resolution information.

An 8-minute pattern length is selected in this research. An 8-minute time period is sufficiently long to allow the incident and non-incident features to be differentiated. A smaller time period may not be sufficient for such a differentiation. A larger time period creates an undesirable delay between the actual time of the incident and the time the incident is detected. The time-interval is chosen on the basis of trade-off between reducing the false alarms and decreasing the delay in incident detection. Because volume and occupancy data are collected every 30 seconds, each sample consists of 16 volume and 16 occupancy measurements for each of the two detection stations. Discrete wavelet transform is performed on each of the 16 data measurement sequences. The result of the transform is composed of 8 finest resolution coefficients $d_1[k]$, 4 medium resolution coefficients $d_2[k]$, and 4 lowest resolution coefficients $s_2[k]$.

Now, an eliminating process is applied to discard less important wavelet coefficients. The effect of random traffic fluctuations has been noted for a long time in transportation research. Many smoothing techniques have been used to reduce the influence of random fluctuations. For example, Cook and Cleveland [7] use the double exponential smoothing method. Chassiakos and Stephanedes [5] investigate the performance of several smoothing methods such as moving average, median filter, and exponential smoothing. All of this research shows the importance of smoothing measurement sequences and the reduction of the influence of random traffic fluctuations in the traffic incident detection research. In this research, 8 finest resolution features, $d_1[k]$, are thrown out in order to minimize the influence of the random traffic fluctuations (noise).

This step basically reduces the number of input nodes (traffic features) to 6x4=24. However, it is still too large for the available number of training examples. Therefore, further feature extraction is desirable. In our model the discrete wavelet transform is followed by a linear discriminant analysis network in order to reduce the number of final features for incident detection (input nodes) to 4.

6. LINEAR DISCRIMINANT ANALYSIS

The motivation for feature extraction is to avoid the "*dimensionality curse*," thus improving the generalization ability of the pattern classifier and reducing its computational processing requirements [12]. Feature extraction can be formulated as a mapping from a d-dimensional input space to an m-dimensional feature space.

$$\Psi : R^d \to R^m, \qquad m<d \tag{4}$$

The linear discriminant analysis extracts features by linearly mapping the input space to the feature space and maximizing the between-class scatter while holding the within-class scatter constant. Let $\xi_i^{(\ell)} = (\xi_{i1}^{(\ell)}, \xi_{i2}^{(\ell)}, \ldots, \xi_{id}^{(\ell)})^T$ denote the ith training sample outputted by the discrete wavelet transform in class ℓ, i=1,2,... , n_ℓ; ℓ=1, 2; ℓ=1 denotes the incident-free samples and ℓ=2 denotes the incident samples. The number of training samples is denoted by n. The within-class covariance matrix, Σ_W, is defined as

$$\Sigma_W = \frac{1}{n}\sum_{\ell=1}^{2}\sum_{i=1}^{n_\ell}(\xi_i^{(\ell)} - m^{(\ell)})(\xi_i^{(l)} - m^{(l)})^T \tag{5}$$

where $m^{(\ell)}$ is the mean vector of class ℓ, ℓ=1,2. The between-class covariance matrix, Σ_B, is defined as

$$\Sigma_B = \frac{1}{n}\sum_{\ell=1}^{2} n_\ell(m^{(\ell)} - m)(m^{(\ell)} - m)^T \tag{6}$$

where m is the mean vector of all of the data. The total scatter matrix is

$$\Sigma_T = \Sigma_W + \Sigma_B \tag{7}$$

The goal of linear discriminant analysis is to find a dxm transform Ψ such that $\frac{\left|\Psi^T \Sigma_T \Psi\right|}{\left|\Psi^T \Sigma_W \Psi\right|}$ is maximized, where the superscript T denotes the transpose of the matrix. The function Ψ is a d by m rectangular transformation matrix that transforms a d by 1 vector to another m by 1 vector. The goal is to maximize the between-class scatter without increasing the within-class scatter. At the same time the dimension of the original input is reduced. The maximization of the aforementioned matrix yields the transformation matrix Ψ. The dimension m of the matrix is assumed in the process. As such, we know the dimension of the vector we want to obtain in advance (for example, 4).

In the final step of the hybrid algorithm for traffic incident detection, the adaptive conjugate gradient neural network (ACGNN) learning algorithm of Adeli and Hung [1] is used for final classification. Input of the ACGNN algorithm is the foremost features obtained from the LDA network.

7. TRAFFIC DATA COLLECTION AND PROCESSING

For automatic incident detection, we are basically interested in 30- to 60-second incremental occupancy and volume data in both upstream and downstream stations. Data should contain normal traffic data as well as infrequent traffic incident data. The collection of such data requires the installation of a complete sensor system in a freeway system. Only a few cities within the United States currently have such a system; Minneapolis, Minnesota is one such city.

Raw data and maps were gathered from the Traffic Management Center (TMC) within the Minnesota Department of Transportation. Loop detectors were installed in Minneapolis, specifically along I-35. Data are automatically collected every day from detector stations. Several important factors concerning this gathered data are considered as follows.

a) <u>Time Period for the Traffic Incident Detection Study</u> The study is confined to the 3:00PM-7:00PM afternoon peak period because of the importance of incident detection under moderate to heavy conditions for an advanced Freeway Management System (FMS).
b) <u>Freeway Segment for the Study</u> After a thorough study of the map and the data and equipment available in Minneapolis, a 10-mile (16 km) segment of freeway along I-35

(north) from 90th Street to 26th Street is selected for this research. This segment includes 22 sensor stations. This segment includes most types of geometric configurations such as bottlenecks (reduction of the number of lanes), ramp entrances, and so on. This information allows us to test the new incident detection algorithm in various geometric configuration situations.

c) Acquiring 30-second Volume and Occupancy Data for the Stations and Time Intervals Under Consideration The station data consists of 1-minute volume and occupancy data updated every 30 seconds and averaged over all lanes. Unlike the 5-minute interval data (which are widely used in many traffic/transportation studies and are therefore well sorted and documented) the 30-second data has to be manually extracted from original binary data files. It took months to extract the appropriate traffic data from the original file due to the large amount of data used in this research.

d) Incident Data These data are stored in several database files separate from the raw traffic data binary files. The incident log file typically includes time and location of the incident occurrence, incident type, duration, severity, impact on traffic, roadway conditions, and so on. The incidents within the freeway segment and the time interval under investigation are extracted from the files.

e) Data Used in this Research Seven months of traffic data were gathered from the Traffic Management Center of the Minnesota Department of Transportation. Twenty-one incidents have been identified in this 7-month period. Most incidents in the data set include blocking one lane or the shoulder. There were no incidents of blocking two or more lanes; detecting these types of incidents is more challenging than one-lane blocking.

8. TRAINING AND NEURAL NETWORK

The raw data consists of a time-series of 1-minute volume and lane occupancy data obtained from every station in the highway section included. The occurrence of incidents is recorded separately with information about the time and location of each incidence. There are 21 incidents in the training data. For the purpose of training, 8-minute segments of data are extracted from the raw data. A total of 83 incident patterns plus 500 incident-free patterns were used. In other words, there are eighty three 8-minute segments which include the 21 incidents at various times in the 8-minute periods. The training of the network is done in three steps. In the first step, the discrete wavelet transformation is applied to each 8-minute pattern consisting of volume and occupancy data from upstream and downstream stations (a total of 2X2X2X8=64 points). Each one of the four data elements of the 8-minute pattern (volume and occupancy data from upstream and downstream stations) has 16 data points. After eliminating all 8 coefficients of the highest (finest) resolution, one coefficient of the medium resolution, and one coefficient of the lowest resolution, the 16 data points for each data element are reduced to 6 data points. Therefore, every 8-minute data pattern with 64 input points is reduced to a new pattern of 6X4=24 points. In other words, through the wavelet transformation, the 64-point input is transformed to a 24-point output which itself becomes the input to the LDA network.

In the second step, the new data set along with the knowledge of incident and incident-free training sets are used to train the LDA network. The input of the LDA network consists of 24 nodes. After the linear discriminant analysis, the output of the LDA network becomes only 4 nodes. A new data set is formed after the linear discriminant analysis is performed and the length of every pattern is decreased to 4.

In the third step, the adaptive conjugate gradient neural networks learning algorithm [1] is used for the final training. The network of this step consists of a 4-node input layer, a 4-node hidden layer, and a 1-node output layer.

9. CONCLUSIONS

A new hybrid neural networks computational model and algorithm has been developed for automatic traffic incident detection using the wavelet transform, a linear discriminant network, and the adaptive conjugate gradient neural network algorithm of Adeli and Hung (1994). Testing of the new incident detection algorithm on the limited data obtained from the Traffic Management Center of the Minnesota Department of Transportation indicates a promising and powerful algorithm for creating a new generation of traffic incident detection system.

Acknowledgment-This research was funded by the Ohio State Center for Intelligent Transportation Research. The cooperation of the personnel at the Traffic Management Center of the Minnesota Department of Transportation is greatly appreciated.

REFERENCES

1. Adeli, H. and Hung, S.L., An Adaptive Conjugate Gradient Algorithm for Efficient Training of Neural Networks, *Applied Mathematics and Computation*, **62**, 81-102, 1994.
2. Adeli, H. and Hung, S.L., *Machine Learning-Neural Networks, Genetic Algorithms, and Fuzzy Systems*, John Wiley & Sons, NY, 1995.
3. Adeli, H. and Saleh, A., Optimal Control Adaptive/Smart Bridge Structures, *Journal of Structural Engineering*, ASCE **123**, 218-226, 1997.
4. Carpenter, G. and Grossberg, S., The Art of Adaptive Pattern Recognition by a Self-Organizing Neural Network, *Computer* **21**, 77-88, 1987.
5. Chassiakos, A.P. and Stephanedes, Y.J., Smoothing Algorithms for Incident Detection, *Transportation Research Record* 1394, TRB, National Research Council, Washington, DC, 8-16, 1993.
6. Cheu, RL. and Ritchie, S.G., Automated Detection of Lane-Blocking Freeway Incidents Using Artificial Neural Networks, *Transportation Research-Part C: Emerging Technologies* **3**, 371-388, 1995.
7. Cook, A.R. and Cleveland, D.E., Detection of Freeway Capacity-Reducing Incidents by Traffic-Stream Measurements, *Transportation Research Record* 495, TRB, National Research Council, Washington, DC, 1-11, 1974.
8. Daubechies, I, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

9. Dudek, C.L. and Messer, C.L., Incident Detection on Urban Freeways, *Transportation Research Record* 459, TRB, National Research Council, Washington, DC, 12-24, 1974.
10. Forbes, G.J., Identifying Incident Congestion, *Institute of Transportation Engineering*, June, 17-22, 1992.
11. Foufaoula-Georgiou, E. and Kumar, P. Eds., *Wavelet in Geophysics*, Academic Press, Inc., 1994.
12. Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Second Edition, Academic Press, NY, 1990.
13. Gall, A.I. and Hall, F.L., Distinguishing Between Incident Congestion and Recurrent Congestion: A Proposed Logic, *Transportation Research Record* 1232, TRB, National Research Council, Washington, DC, 1-8, 1990.
14. Haykin, S., *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Co., Inc., NY, 1994.
15. Hsiao, C., Lin, C. and Cassidy, M., Application of Fuzzy Logic and Neural networks to Automatically Detect Freeway Traffic Incidents, *Journal of Transportation Engineering* **120**, 753-772, 1994.
16. Hung, S.L. and Adeli, H., Parallel Backpropagation Linearing Algorithms on Cray Y-MP8/864 Supercomputer, *Neurocomputing* **5**, 287-302, 1993.
17. Ito, K., Namai, T. and Kojima, M., Applying Neural Networks to Automatic Incident Detection, *Traffic Technology International*, Aug-Sept, 32-35, 1996.
18. Kimachi, M., Kanayama, K. and Teramoto, K., Incident Prediction by Fuzzy Sequence Analysis, Proc, Vehicle Navigation and Information Systems Conference, 51-56, 1994.
19. Kohonen, T., *Self-Organization and Associative Memory*, Springer-Verlag, Berlin, 1989.
20. Mallat, S.G., A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **11**, 674-693, 1989.
21. Mao, J. and Jain, A.K., Artificial Neural Networks for Feature Extraction and Multivariate Data Projection, *IEEE Trans on Neural Networks* **6**, 296-317, 1995.
22. Mehrotra, K., Mohan, C.K., and Ranka, S., *Elements of Artificial Neural Networks*, MIT Press, Cambridge, MA, 1997.
23. Pao, Y.H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA, 1989.
24. Parkany, E. and Bernstein, D., Design of Incident Detection Algorithms Using Vehicle-to-Roadside Communication Sensors, *Transportation Research Record* 1494, TRB, National Research Council, Washington, DC, 67-76, 1995.
25. Payne, H.J. and Tignor, S.C., Freeway Incident Detection Algorithms Based on Decision Tree with States, *Transportation Research Record 682,* TRB, National Research Council, Washington, DC, 378-382, 1978.
26. Saleh, A. and Adeli, H., Robust Parallel Algorithms for Solution of the Ricatti Equation, *Journal of Aerospace Engineering* ASCE **10**, 126-133, 1997.

27. Stephanedes, Y.J., Chassiakos, A.P. and Michalopoulos, P.G., Comparative Performance Evaluation of Incident Detection Algorithms, *Transportation Research Record* 1360, TRB, National Research Council, Washington, DC, 50-57, 1992.
28. Stephanedes, Y.J. and Liu, X., Artificial Neural Networks for Freeway Incident Decision, *Transportation Research Record 1494*, TRB, National Research Council, Washington, DC, 91-97, 1995.
29. Vetterli, M. and Kovacevic, J., *Wavelets and Subband Coding*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1995.