Technical Correspondence

Traffic-Incident Detection-Algorithm Based on Nonparametric Regression

Shuming Tang and Haijun Gao

Abstract—**This paper proposes an improved nonparametric regression (INPR) algorithm for forecasting traffic flows and its application in automatic detection of traffic incidents. The INPRA is constructed based on the searching method of nearest neighbors for a traffic-state vector and its main advantage lies in forecasting through possible trends of traffic flows, instead of just current traffic states, as commonly used in previous forecasting algorithms. Various simulation results have indicated the viability and effectiveness of the proposed new algorithm. Several performance tests have been conducted using actual traffic data sets and results demonstrate that INPRs average absolute forecast errors, average relative forecast errors, and average computing times are the smallest comparing with other forecasting algorithms.**

Index Terms—**Automatic incident detection, forecast, nonparametric regression algorithms, state vectors, traffic incidents.**

I. Introduction

With the rapid increase in metropolitan population and other urbanization activities, a huge demand has been imposed on metropolitan transportation systems. Traffic congestion has become a serious problem, not only in developed countries, but also in developing countries. As traffic operation conditions deteriorate, the frequency of traffic incidents increases significantly.

Traffic incidents are defined as being events that reduce the capacity of network links to carry traffic, such as accidents, disabled vehicles, spilled loads, temporary maintenance and construction activities, and other unusual events [1], [2]. Accurate and early detection of traffic incidents is vital for the restoration of smooth traffic flow. Therefore, we must adopt an integrated approach and utilize the concept of intelligent transportation systems (ITS), particularly, advanced traffic-management systems (ATMS) that use new communication, control, sensing, and information technology to solve traffic-congestion problems [3]–[6].

Besides congestions, traffic incidents have caused heavy life and economic losses. For example, according to the recent report of China's Ministry of Public Safety, 667 507 traffic accidents were reported in 2003, resulting in 104 372 deaths, 494 174 injured people, and a direct economic cost of $3.37 billion U.S. [7]. In the United States, more than half of congestion on freeways is caused by incidents and almost all congestion on rural freeways is involved with incidents, road reconstruction, and maintenance [8]. According to an investigation conducted in Los Angeles, if an incident lasted one more minute, traffic delay time or congestion would amount to $4 \sim 5$ min in a nonrush hour. If an incident happened in a rush hour, its delay and congested time would be much longer and the corresponding cost would be larger [8], [9]. This clearly shows that the loss caused by incidents is proportional to their durations, so it is critical to detect them as earlier as possible in order to take remedial actions before traffic situations get worsen.

Manuscript received December 15, 2003; revised August 25, 2005. This work was funded in part by the Outstanding Young Scientist Research Fund under Grant 60125310, by the Key Project on Networked Systems, National Natural Science Foundation, under Grant 60334020, and a Shandong 863 Project, Shandong Provincial Government, under Grant 030335. The Associate Editor for this paper is N. Zheng.

S. Tang is with the Institute of Automation, Shandong Academy of Sciences, Jinan, Shandong 250014, China (e-mail: sharron@ieee.org).

H. Gao is with the Key Laboratory of Complex Systems and Intelligence Science, Chinese Academy of Sciences, Beijing 100080, China (e-mail: zkyghj@vip.sina.com).

Digital Object Identifier 10.1109/TITS.2004.843112

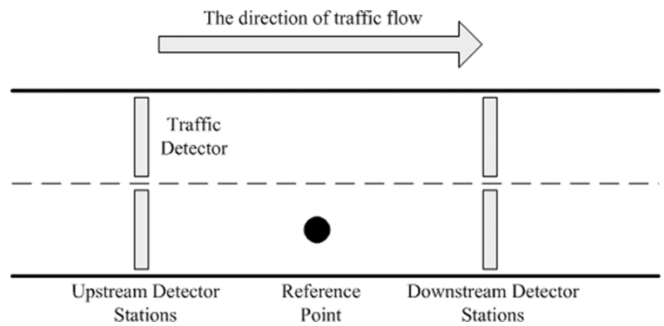

Fig. 1. Traffic-data collection from traffic detectors.

Under such background, automatic incident-detection (AID) algorithms have become an interesting and active research topic and received considerable attention over the last few decades [5]. However, very few AID algorithms have been used successfully in actual traffic-management systems, mainly due to their poor detection accuracy and high demand for real-time and reliable traffic information. Recently, several advanced AID models using computational intelligence, especially neural networks, fuzzy logic, and image-based processing, have been proposed and have shown great promises [1], [10], [11], but the corresponding complexity and sophistication required in their implementation prevent their fast application in actual ATMS.

This paper aims to construct a simple, accurate, and reliable AID model using conventional but effective methods so that it can be implemented and tested easily by practicing traffic engineers. To this end, nonparametric regression and standard deviation algorithms are used. This paper is organized into four parts. A brief review of four types of AID algorithms is given in Section II. The proposed INPR algorithm and its performance tests are described in Section III, followed by a traffic-simulation case study in Section IV. Finally, Section V concludes this paper with a brief summary and discussion on the future direction.

II. Overview on AID Algorithms and Nonparametric Regression

A. Overview of AID Algorithms

Over the past decades, many AID algorithms have been developed to detect the occurrence of incidents by detecting the change of traffic flow parameters measured at upstream and downstream detector stations (see Fig. 1) [1], [3], [12], [13]. In general, AID algorithms can be classified into four classes [14], [15]: 1) prediction algorithms; 2) mode identification algorithms; 3) methods based on a traffic flow model or theory; and 4) incident detection using computational intelligence, such as neural networks, fuzzy logic, or image-based processing algorithms. A brief summary of those algorithms is given as follows.

1) Prediction Algorithms: A prediction algorithm is a statistical procedure to forecast traffic flows. Its implementation can be divided into two steps. First, it predicts values of traffic-flow parameters according to historical traffic data. Second, it compares the predicted values with the current data to obtain their differences and determines whether the differences are beyond a predefined threshold [9], [14]. There are three major predication algorithms in the literature: standard deviation [14], time series [15] and filtering [16]. As a simple example, the simple implementation procedure of standard-deviation-based prediction algorithms is illustrated here, since it is used here.

Standard Deviation Algorithms (SND): In an SND algorithm, the average of sampling historical traffic data (volume or density) is used as the value of a query traffic parameter. Denote the prediction value as $\hat{x}(t)$, the deviation of these sampling traffic data as S, and the current value as $x(t)$. Then, the standard deviation can be calculated as

$$\mathrm{SND}(t) = \frac{\hat{x}(t) - x(t)}{S}.$$

For a given traffic flow, the algorithm will compute and check the value of its SND: if its SND is larger than a predefined threshold, the algorithm will send out an alarming signal for a possible incident.

The most distinguished feature of this prediction algorithm is its simplicity. However, the evolving trend of an incident is difficult to included by this method and, generally, the corresponding rate of false alarms is high [9], [14].

2) Traffic-Pattern-Recognition Algorithms: The function of a traffic-pattern-recognition algorithm is to identify and distinguish different traffic patterns according to data from detector stations. One of the simplest methods for recognizing accident patterns is to use increases in upstream occupancy and decreases in downstream occupancy around loop detectors on a freeway. In this case, a warning message for a possible traffic incident and its location will be issued when the upstream occupancy increase and the downstream occupancy decrease pass their respective predefined values. The most commonly and widely used algorithms in this area are the California and filtering algorithms and their variants [14], [15].

3) Model-Based Detection Algorithms: Model-based detection algorithms use sophisticated traffic-flow theories to model traffic flow and estimate traffic states [14]. One of such algorithms is based on the catastrophic theory, actually a modified McMaster algorithm [17].

Theoretically, it is difficult to construct models to describe traffic flow because it is an infinite dimension, nonlinear, stochastic, time variant, and complicated dynamic system. Even if a traffic-flow model can be attained, it impossibly covers all characteristics of traffic flow. Thus, it is uncommon to use model-based detection algorithms to detect incidents.

4) Computational Intelligence Incident Detection Algorithms:

Fuzzy Algorithms: Fuzzy algorithms use fuzzy logic, the concept of fuzzy boundary, and the changing tendency of occupancy or speed-density relationships among two adjacent detector stations to detect traffic incidents. When traffic data are difficult to be collected or there are no enough traffic data, it is effective and useful to apply fuzzy algorithms. Normally, fuzzy algorithms have high robustness and can overcome the boundary condition problem inherited in conventional threshold-based methods [10].

Neural-Network Algorithms: Neural networks are trained using historical traffic data to recognize the pattern of the traffic flow and identify incident or incident-free states [1], [12], [18]. Compared with model-based detection algorithms, neural network algorithms are easier to use and better for real-time implementation. However, these algorithms still have many drawbacks. First, the rate of convergence of a neural-network model can be very slow. Second, it is difficult to understand the meanings of neural-network operations, since it is a black-box approach. Third, the implementation of neural networks requires large traffic historical data sets and the state range covered by those data must be wide and large enough. Otherwise, their detection performance will not be sufficient, even no better than that of traditional algorithms.

Image-Based Processing Algorithms: Imaged-based processing algorithms use computer vision and image-based processing technology to extract the information of traffic parameters from the video sequences taken by video cameras, then detect and verify the occurrence of traffic incidents.

With the wide application of video-traffic surveillance in current traffic-management systems, this kind of methods have become increasingly popular and important [2], [11], [19]. Generally, image-based processing AID algorithms have a high detection rate, a low false-alarm rate, and a short time to detection [20]. Therefore, it could be the key for future AID technology as advanced video traffic surveillance and image processing become available.

B. Overview on Nonparametric Regressive (NPR) Algorithms

NPR is a forecasting technique based on the nearest neighbor searching in which forecasts are generated from past observations that are similar to the current conditions [21].

In 1993, Fine and Yuan suggested a method to use NPR to decompose the training process of a neural network and reduce its training time [22]. In 1994, Schaal and Atkeson presented an approach for robot learning based on the NPR technique [23]. In 1995, Smith used NPR to predict short-time traffic flow of a single traffic state [24]. Recently, many researchers have proposed many modified methods to forecast traffic flows, for instance, using k-nearest neighbor graph or approximate nearest neighbor searching, etc. [25], [26].

III. INPRA-Based Traffic-Incident-Detection Algorithm

In this section, a new traffic-incident-detection algorithm based on the combination of an improved nonparametric regression (INPR) and standard deviation is proposed. Basically, the INPR is constructed based on the searching method of nearest neighbors for a traffic state vector.

A. INPR

In INPR, traffic forecasts are derived from past observation vectors that are similar to the current state vector. It can be described simply and briefly as follows: Given a set of n vectors $\vec{P} = \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n\}$ in a vector space V, preprocess $\vec{P}$ so that a vector in $\vec{P}$ closest to a query vector $\vec{q} \in V$ can be found efficiently.

The main advantage of INPR lies in forecasting through possible trends of traffic flow, instead of just current possible traffic states, as commonly used in the previous nonparametric regression forecasting algorithms.

The construction of INPR consisting of the following five steps: 1) choice of state vectors; 2) historical observation data; 3) data filtering; 4) k-nearest neighbor searching; and 5) traffic predication. Here, those issues will be addressed based on two traffic volume (q) and speed (v).

1) Choice of a State Vectors: The choice of a state vector depends on the changing trend of a traffic flow. Basically, more elements involved in a chosen state vector, more information of traffic flow will be included and more accurate the forecast will be. Nevertheless, more elements in a chosen state vector will result in a long computing time. Taking into the relativity among neighboring traffic data and the coming traffic data [27], a state vector consisting of six traffic instances is selected here

$$X(t) = [q(t-5), q(t-4), q(t-3), q(t-2), q(t-1), q(t)] \quad (1)$$

where $q(t-i), i = 0, 1, \ldots, 5$ is the volume value at previous i time relative to the current time t of a traffic query point.

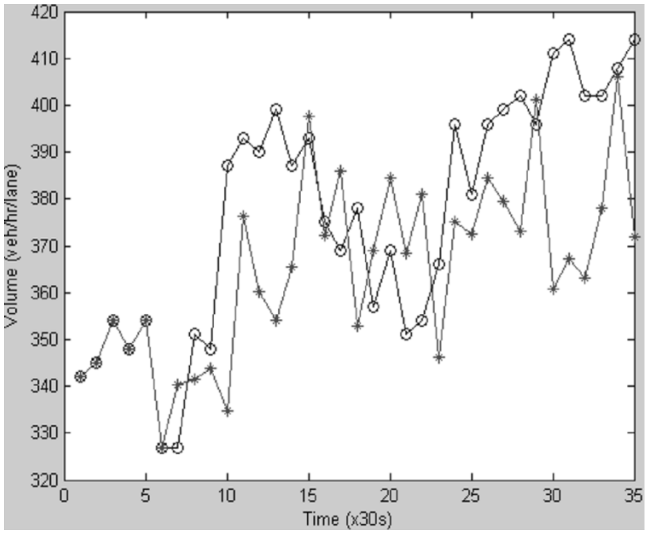

Fig. 2. Actual data versus forecasted data under incident-free condition.

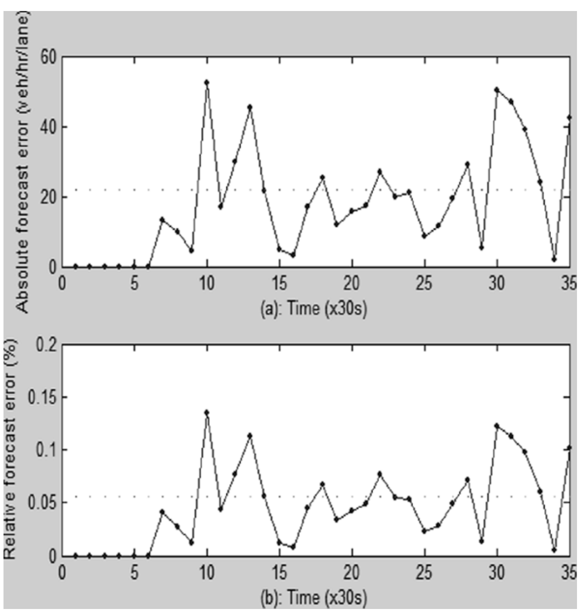

Fig. 3. Absolute and relative forecast errors.

TABLE I
PERFORMANCE COMPARISON OF AID ALGORITHMS

Index	AAFE	ARFE（%）	ACT（s）
INPR Algorithm	21.80	5.92	0.07
Traditional NPR Algorithm	178	8.13	2.7
Neural Network Algorithm（BP）	201	11.35	2.5
Kalman Filtering Algorithm	192	10.62	0.5

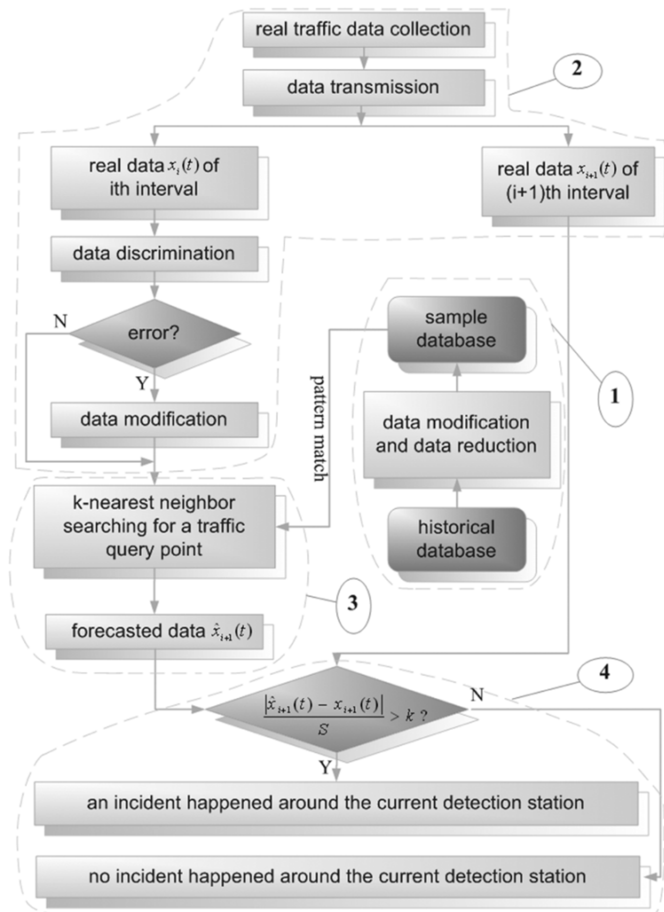

Fig. 4. Procedure for INPR–SND traffic incident detection.

2) Nearest Neighbor Searching Rule: The k-nearest neighbor method is adopted to search k neighbors of a traffic query point and estimate the next traffic data according to these k-nearest neighbors. A review of existing literature shows that as k start to increase from 0, forecasting accuracy improves initially. However, as k continues to increase, forecasting accuracy begins to deteriorate [21], [28]. This indicates that too few nearest neighbors would poorly represent the current traffic conditions while too many neighbors could also reduce forecasting accuracy, since they would eventually approach the arithmetic mean of the whole data set. Based on numerical and empirical testing, $k = 5$ is used here for forecasting state vector $X(t)$ [29].

3) Similarity Rule: A weighted Euclidean distance is used to determine the distance between two traffic vectors. Assumed that two traffic vectors can be designated as in (1) and

$$X'(t) = [q'(t-5), q'(t-4), q'(t-3), q'(t-2), q'(t-1), q'(t)]. \quad (2)$$

Then, the weighted Euclidean distance between them is

$$d = \sqrt{\frac{S_{qq}}{W}} \quad (3)$$

where $S_{qq} = \sum_{i=0}^{5} r_{qq}^{i+1}[q(t-i) - q'(t-i)]^2$, $W = \sum_{i=0}^{5} r_{qq}^{i+1}$, $r_{qq} = 0.96$ are the weighted coefficients [27].

4) Prediction Algorithm: In the proposed prediction algorithm, the weighted average of the reciprocals of Euclidean distances is used as weighing factors, specifically

$$q(t+1) = \sum_{i=1}^{k} \frac{\frac{1}{d_i}}{d} q_i(t+1) \text{ where } d = \sum_{i=1}^{k} \frac{1}{d_i}.$$

B. Performance Test of INPR Algorithm

Part of incident-free data collected at the I-880 Freeway in the San Francisco Bay area, CA, is used to test performances of the proposed

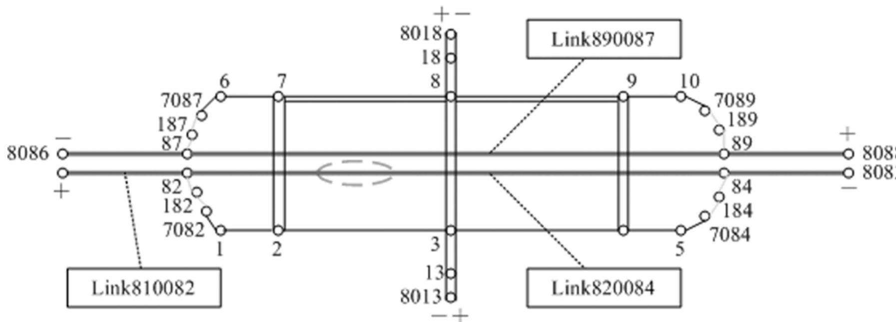

Fig. 5. Simulated traffic network.

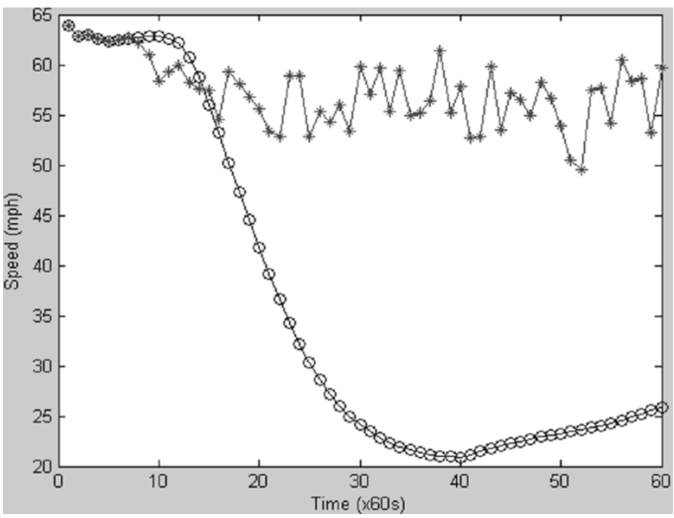

Fig. 6. Actual data versus forecasted data in a traffic incident (AAFE $=$ 24.4150 mi/h; ARFE $=$ 93.82%).

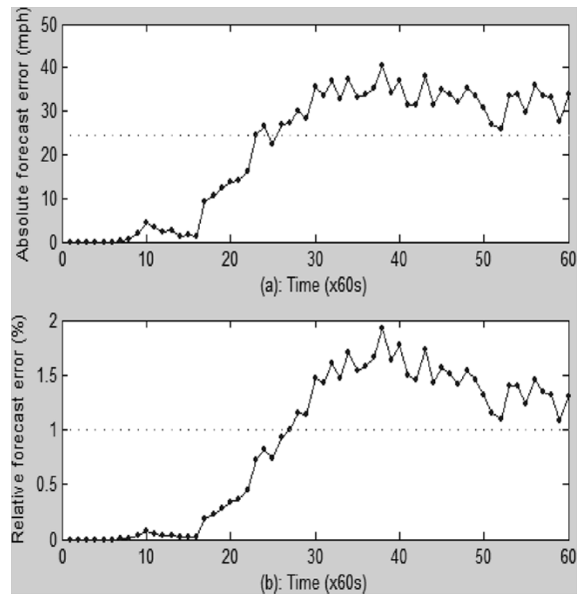

Fig. 7. Absolute and relative forecast errors.

INPR algorithm. The incident-free data are in the form of lane specific volumes (in vehicles/h/lane) collected at 30-s intervals.

Fig. 2 shows the predicted and actual data and Fig. 3 gives the corresponding forecast error for incident-free cases. Based on Figs. 2 and 3, the average absolute forecast error (AAFE), relative forecast error (ARFE), and computing time (ACT) of INPR algorithm are compared with those of traditional NPR algorithm, neural-network algorithm, and Kalman filtering algorithm [27]. The comparative results are shown in Table I.

Comparing the performance indexes in Table I, it clearly indicates the advantages of the INPR algorithm: its AAFE, ARFE, and ACT are the smallest. However, it should be noted that those results are obtained with a historical data set including 1318 instances and the nearest neighbors searching of 29 observations (see Fig. 2). The average computing time will be long if the size of historical data is large.

C. *Traffic-Incident Detection Based on INPR and SND*

Based on the previous discussion, the INPR and SND algorithms are combined into an INPR–SND algorithm. This algorithm consists of four parts: 1) constructing a historical data set; 2) collecting and filtering real-time traffic data for fast computation; 3) forecasting traffic flow based on INPR; and 4) detecting traffic incidents by SND algorithm. Fig. 4 shows the procedure of the INPR–SND algorithm.

IV. Case Study With Simulations

A. *Traffic Simulated Network*

To further validate the performance of the proposed algorithm for traffic incidents detection, traffic simulations with CORSIM software from the Traffic Software Integrated Systems (TSIS) are conducted. Fig. 5 is a simulated traffic network used for this purpose.

In this simulation, a traffic incident has occurred on Link820 084 and its position is shown in the dashed-line area in Fig. 5. There is no any other traffic incident happened on Link 890 087 and Link 810 082. Meanwhile, there are four OD nodes, 8081, 8088, 8018, and 8013 in the simulated network. The simulated traffic incident occurs at $t = 10$ min, lasts 30 min, and, thus, is removed at $t = 40$ min.

In order to compare simulation results, the same flow rate 4700 vehicles/h has been used for both incident and incident-free conditions in the simulated traffic network. The traffic data with incident-free condition are classified as traffic historical data and the data with incident condition are classified as the actual data to be forecast. Each simulation experiment is lasting for 1 h and traffic speeds (in mi/h) are average in a 60-s period.

B. *Simulation Results*

The simulation results are given in Figs. 6 and 7. Fig. 6 shows the actual (dotted-solid line) and forecasted (star-solid line) data. Fig. 7 presents absolute and relative forecast errors. The dotted-dash line in Fig. 7(a) is the average absolute forecast errors while the dotted-dash-line in Fig. 7(b) is the average relative forecast errors. Fig. 8 is the SND values calculated according to the INPR–SND procedure specified in Fig. 4.

From Fig. 8, the SND value increases over the entire simulation period; thus, the traffic incident will definitely be detected by the proposed INPR–SND algorithm. However, different thresholds defined in

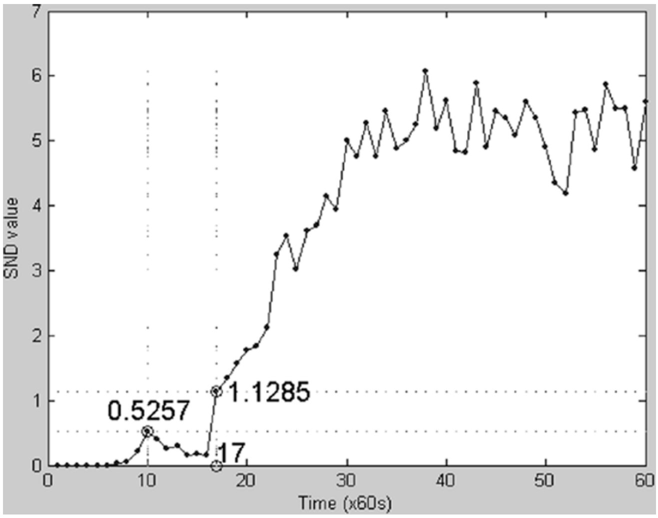

Fig. 8. History of SND values.

the SND algorithm will result in different detection performances, especially different times of traffic-accident occurrence. In our simulation, the traffic incident will be instantly detected if the threshold for an incident in the INPR–SND algorithm is 0.52, while if the threshold is set as 1, then the detected incident time will be delayed by 7 min, as one can see from Fig. 8.

V. Conclusion

This paper presents an improved nonparametric regression method to forecast traffic flow. A performance test shows it has low average absolute, relative forecast errors, and short average computing time. Based on this forecast method, a new traffic-incident-detection algorithm, called INPR–SND, is proposed by combing with the simple standard deviation method. A case study with simulation has demonstrated the effectiveness of the proposed AID algorithm.

The potential disadvantages of the proposed method could be as followings: 1) the average computing time for traffic forecasting will increase if the historical data is large and 2) the selection of SND thresholds will affect its performance, especially the alarm trigging times, significantly, and no analytical procedure is available to determine appropriate thresholds. Therefore, further research will be needed to increase the usability and reliability of the proposed INPR–SND algorithm for traffic-incident detection.

Acknowledgment

The authors would like to thank Prof. F.-Y. Wang of the University of Arizona, Prof. R. L. Cheu of the National University of Singapore, and Dr. X. Gong and Mr. K. Wang of the Chinese Academy of Sciences for their discussions. Feedback from the reviewers is also greatly appreciated.

References

[1] D. Srinivasan, X. Jin, and R. L. Cheu, "Evaluation of adaptive neural network models for freeway incident detection," *IEEE Trans. Intell. Transport. Syst.*, to be published.

[2] H. Ikeda, T. Matsuo, Y. Kaneko, and K. Tsuji, "Abnormal incident detection system employing image processing technology," in *Proc. IEEE Int. Conf. Inteeligent Transportation Systems (ITS'99)*, Oct. 1999, pp. 748–752.

[3] P. B. Farradyne, *Traffic Incident Management Handbook*: Office of Travel Management, Federal Highway Administration, 2000.

[4] R. E. Turochy and B. L. Smith, "A New procedure for detector data screening in traffic management systems,", TRB Paper 00–0842.

[5] F.-Y. Wang, S. Tang, Y. Sui, and X. Wang, "Toward intelligent transportation systems for the 2008 Olympics," *IEEE Intell. Syst.*, vol. 18, no. 6, pp. 8–11, Nov./Dec. 2003.

[6] F.-Y. Wang and S. Tang, "Artificial societies for integrated and sustainable development of metropolitan systems," *IEEE Intell. Syst.*, vol. 19, no. 4, pp. 82–88, Jul./Aug. 2004.

[7] N. Zheng, S. Tang, Q. Li, G. P. Lai, and F. Y. Wang, "Toward intelligent driver-assistance and safety warning systems," *IEEE Intell. Syst.*, vol. 19, no. 2, pp. 8–11, Mar./Apr. 2004.

[8] W. Liu and Z. Yang, *Control Methods for High Freeway Systems*. Beijing, China: Renmin Jiaotong, 1998.

[9] Y. Wu, Y. Wang, and J. Cao, "The investigation on automatic incident detection algorithm of traffic accident," *Manuf. Automat.*, vol. 22, no. 12, pp. 42–45, 2000.

[10] S. Lee, J. Yen, and R. A. Krammes, "Application of fuzzy logic for detecting incidents at signalized highway intersections," in *Proc. 5th IEEE Int. Conf. Fuzzy Systems*, New Orleans, LA, Sep. 1996, pp. 867–872.

[11] W. Hu, X. Xiao, D. Xie, T. Tan, and S. Maybank, "Traffic accident prediction using 3-D model-based vehicle tracking," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 677–694, May 2004.

[12] R. L. Cheu, D. Srinivasan, and E. T. Teh, "Support vector machine models for freeway incident detection," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems (ITS'03)*, Shanghai, China, Oct. 2003, pp. 238–243.

[13] K. Thomas and H. Dia, "Incident detection by fractal dimension analysis of loop detector data," in *Proc. 22nd Conf. Austral. Institute of Transport Research (CAITR;00)*, Dec. 2000, pp. 1–8.

[14] H. Huang, "The Investigation of traffic incident detection algorithm," Ph.D. dissertation, China, 2001.

[15] A. Guin, "An Incident detection algorithm based on a discrete state propagation model of traffic flow," M.S. thesis, Georgia Inst. Technol., Atlanta, 2004.

[16] Y. Zhang and Y. Liu, *A Filtering Method of Traffic Incident Detection*: Yunnan Univ. Technol., 1998, vol. 14, pp. 17–23.

[17] S. Cohen, "Comparative assessment of conventional and new incident detection algorithms," in *Proc. Inst. Elect. Eng. Conf. Road Traffic Monitoring and Control*, Apr. 1994, pp. 156–159.

[18] A. Samant and H. Adeli, "Enhancing neural network traffic incident-detection algorithms using wavelets," *Comp.-Aided Civil Infrastructure Eng.*, vol. 16, no. 4, pp. 239–259, Jul. 2001.

[19] R. Woesler, "Fast extraction of traffic parameters and reidentification of vehicles from video data," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems (ITS'03)*, Oct. 2003, pp. 774–778.

[20] J. Versavel, "Road safety through video detection," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems (ITS'99)*, Oct 1999, pp. 753–757.

[21] B. L. Smith and R. K. Oswald, "Effects of parameter selection on forecast accuracy and execution time in nonparametric regression," in *Proc. IEEE Intelligent Transportation Systems Conf. (ITS'00)*, Dearborn, MI, Oct. 2000, pp. 252–257.

[22] T. L. Fine and J.-L. Yuan, "Nonparametric regression-based method for neural network training," in *Proc. IEEE Int. Symp. Information Theory*, Jan. 1993, p. 351.

[23] S. Schaal and C. G. Atkeson, "Robot learning by nonparametric regression," in *Proc. IEEE/RSJ/GI Int. Conf. Intelligent Robots and Systems*, vol. 1, Sep. 1994, pp. 478–485.

[24] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transport. Eng.*, vol. 123, no. 4, pp. 261–266, 1997.

[25] I. Piotr, "On approximate nearest neighbors in noneuclidean spaces," in *Proc. 39th Annu. Symp. Foundations of Computer Science*, Nov. 1998, pp. 148–155.

[26] F. Pasi, V. Olli, and H. Ville, "Fast PNN-based clustering using K-nearest neighbor graph," in *Proc. IEEE 3rd Int. Conf. Data Mining*, Nov. 2003, pp. 525–528.

[27] S. Tang, X. Gong, and F.-Y. Wang, "Traffic incident detection algorithm based on nonparameter regression," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems (ITS'02)*, Singapore, Sep. 2002, pp. 1032–1036.

[28] X. Gong, "Short-term traffic flow forecasting and routing based on data mining," Ph.D. dissertation, Chinese Acad. Sci., 2003.

[29] X. Gong and S. Tang, "Integrated traffic flow forecasting and traffic incident detection algorithm based on nonparametric regression," *China J. Highway Transport.*, vol. 16, no. 1, pp. 82–86, 2003.