UC Berkeley
Working Papers

Title
Development Testing And Evaluation Of Advanced Techniques For Freeway Incident Detection

Permalink
https://escholarship.org/uc/item/5720z2mw

Authors
Ritchie, Stephen G.
Abdulhai, Baher

Publication Date
1997

Development Testing and Evaluation of Advanced Techniques for Freeway Incident Detection

Stephen G. Ritchie, Baher Abdulhai
University of California, Irvine

The contents of this report reflect the views of the authors who are
responsible for the facts and the accuracy of the data presented herein.
The contents do not necessarily reflect the official views or policies of
the State of California. This report does not constitute a standard,
specification, or regulation.

CALIFORNIA PARTNERS FOR ADVANCED TRANSIT AND HIGHWAYS

Development Testing and Evaluation of Advanced Techniques for Freeway Incident Detection

Stephen G. Ritchie
Baher Abdulhai

EXECUTIVE SUMMARY

Development of a universal freeway incident detection algorithm is a task that remains unfulfilled despite the promising approaches that have been recently explored. Incident detection researchers are now realizing that an operationally successful incident detection framework needs to fulfill all components of a set of recognized needs. In this research we introduce and define a universal incident detection framework on the basis of one such set of needs. Among other requirements, a freeway incident detection algorithm needs to be operationally accurate and automatically transferable. In this research we introduce an algorithm that has the potential to fulfill the defined universality requirements. The algorithm is a modified form of the Bayesian-based Probabilistic Neural Network (PNN) that utilizes the concept of statistical distance. The report is divided into three main sections. The first section is a detailed definition of the attributes and capabilities that a potentially universal freeway incident detection framework should possess. The second section discusses the training and testing of the PNN. In the third section, we evaluate the PNN relative to the universality template previously defined. In addition to a large set of simulated incidents, we utilize a fairly large real incident database from the I-880 freeway in California to comparatively evaluate the performance and transferability of different algorithms including the PNN. Experimental results indicate that the new PNN-based algorithm is competitive with the Multi Layer Feed Forward (MLF) architecture which was found in previous studies to yield superior incident detection performance. The PNN performance was competitive with the MLF in terms of Detection Rate (DR), False Alarm Rate (FAR), and average Time To Detection (TTD). In addition, results also point

to the possibility of utilizing the real-time learning capability of this new architecture to produce a transferable incident detection algorithm without the need for explicit off-line retraining in the new site. In this important respect, and unlike existing algorithms, the PNN has been found to markedly improve in performance with time in service. Moreover, the PNN-based framework possesses the remaining attributes that would make it potentially "universal".

TABLE OF CONTENTS

1. INTRODUCTION

Proper freeway incident detection and management are now well recognized key components of any potentially successful Advanced Traffic Management System (ATMS). Unfortunately, the performance of published conventional approaches to automatic freeway incident detection has proven inadequate for every day use at Traffic Management Centers (TMCs). Inadequacy stems from three main sources: *first* is the less-than-perfect performance at the original site that the algorithm was developed for, *second* is the lack of transferability to any new site, and *third* is the inability of the algorithm to consider important but often neglected issues such as the prior (predicted) probability of occurrence of incidents, posterior probability of an incident after an alarm, and the unequal costs of misclassifying a traffic pattern as an incident or non-incident. More advanced algorithms have been recently proposed and developed by different researchers using approaches such as neural networks (Cheu and Ritchie, 1996), filtering techniques (Stephanedes and Hourdakis, 1996) and catastrophe theory (Hall et al, 1991). In our view, the neural network approach has shown the greatest promise. However, all algorithms lack the universality that the incident detection community has been and still is waiting for. Very recent research *plans* have been made (Payne and Thompson 1996) to consolidate the benefits of the different promising incident detection approaches into a modular structure. However, to date there is no single stand alone algorithm that has shown sufficient promise to fulfill all the universality expectations simultaneously. In this research, we introduce a candidate universal approach that is expected to fulfill one such set of universality requirements, that we will define shortly. This new approach in based on a modified form of the Bayesian-based Probabilistic Neural Network (PNN) that utilizes the concept of statistical distance (Abdulhai and Ritchie, 1995).

2. UNIVERSALITY REQUIREMENTS

By combining TMC survey results, theoretical reasoning, intuition, and experience, we have concluded that for a freeway incident detection

algorithm to be universal, it needs to possess the following set of capabilities and attributes:

- **High Performance**: in terms of high Detection Rate (DR%), low False Alarm Rate (FAR%) and short mean Time To Detection (TTD).   High performance could only be achieved through careful selection of distinctive input features, a powerful pattern recognition logic, and training data that represents all possible scenarios. Acceptable limits on the above performance indicators or variables (DR, FAR, and TTD) can be defined by Traffic Management Centers as will be detailed later.

- **Fast Training/Calibration**: training of the algorithm should not be tedious and time consuming. In fact, there are no known limits on what a reasonable training time would be. However, the faster the training/retraining processes, the higher the potential for on-site real-time retraining, all other attributes being equal.

- **Reasonable TMC implementation requirements**: in terms of having no recalibration requirements of the TMC that would require skilled system developers.

- **Transferable logic**: the logic or theory on which the algorithm is built should not be constrained spatially, temporally or in other respects that would limit transferability to other locations.

- **Transferable training/calibration parameters**: an algorithm trained on one site should be usable in other new environments with as little performance deterioration as possible. A new environment could be defined as any freeway site that has significantly different statistics. Site statistics change, of course from site to site, and less evidently at the same site over time. The less the performance deterioration after transferability, the easier the subsequent process of on-site real-time updating of the knowledge content of the algorithm, in order to adapt to the new site.

- **Minimal initial training data requirements**: as is well known, real incident data are not only very sparse but also very difficult to obtain. It would be very time consuming, if not impractical to try to collect real

incident data that is diverse enough to represent all possible scenarios, and accurate enough to result in minimal calibration errors. It is easy to realize that a serious limitation of most existing incident detection algorithms is the quality of initial training due to the nature of the data used for calibration. They are often calibrated using either a very limited set of "real" incident data or a large set of artificially simulated data, with attendant simulation limitations. This may limit the quality of the resulting algorithms and perhaps also limit their transferability. Hence, a successful algorithm should be capable of being up and running using minimal incident data, and then *automatically* improve with time in service as more data become available.

- **Account for prior probabilities of incidents**: the algorithm should incorporate into an incident alarm decision the predicted prior probability of occurrence of an incident, based on such factors as weather condition, traffic conditions, road surface conditions, and geometric features of the freeway section.

- **Account for the unequal costs of misclassifying traffic patterns**: Although a false alarm and a missed incident are both misclassifications, their costs are not equal. One might be more expensive than the other depending on several factors such as the location of the freeway section, the time of the day, incident characteristics, resources deployed in responding to incidents, and TMC size, capabilities and preferences.

- **Capable of producing the posterior probability of an incident**: there usually are varying extents of uncertainty associated with algorithm-generated incident alarms. The certainty of an alarm depends among other things, on the quality of the algorithm itself, the prior probability of an incident, and the number of preceding alarms (which may be correlated to the elapsed time since the onset of the incident, and the extent of closure and queue formation on the freeway). The algorithm should be capable of producing a probabilistic estimate of the certainty associated with an alarm.

- **Estimate incident severity**: a key requirement that helps prioritize the allocation of TMC resources in response to several simultaneous incident alarms.

- **Capture incident duration**: key information required in order indicate the restoration of normal traffic flow.

- **Statistically and theoretically well established with minimal site-specific heuristics**: the clarity and soundness of the theoretical basis of the algorithm are key factors that underlie its acceptability. The fewer the number of site-specific heuristics involved, the more general the algorithm should be.

   Additionally, there are four attributes that are not directly related to the theory of the pattern recognition algorithm itself but rather are related to the extracted features and formulation of the pattern space on which the algorithm is to operate. These attributes are listed below for the sake of completeness of the universality template. However, due to both their extreme importance, and space limitations in this report, they will be addressed in detail in a future report:

- **Flexibility in working with different surveillance system designs and technologies**: successful field implementation of the algorithm should not be limited to specialized surveillance system designs, including sensor technology, spacing and placement. For instance, some algorithms require a special loop detector placement configuration within a funneling freeway section (Lin, 1995). Such a requirement would limit the applicability of the algorithm in the absence of the required configuration.

- **Immunity to minor traffic fluctuation effects**: the algorithm should not be affected by minor, and very short term traffic fluctuations which tend to cause false alarms.

- **Immunity to bottleneck effects**: the algorithm should be capable of distinguishing between true incident patterns and incident-like patterns due to physical bottlenecks that cause queuing patterns similar to an incident. It should also be immune to virtual bottleneck effects caused by abrupt geometry changes or demand changes at major on and off ramp

locations. We call the latter scenario a virtual bottleneck because of the associated sudden change in the traffic flow variables from the upstream station to the downstream station of the freeway section, which is similar to the effect of an incident and the effect of physical bottlenecks, despite the absence of both.

- **Immunity to consistent detector malfunctions or biases**: the algorithm should not be affected by consistent biases such as an upstream station giving a consistently higher occupancy reading than the downstream station for no obvious reasons. Such a situation has been observed in field loop data.

The above template of requirements and attributes, if fulfilled, could yield a complete, operationally successful incident detection algorithm.   The presence or absence of these attributes will be used to evaluate our proposed framework as well as any other algorithm that might be used for comparative purposes.

3. THE PROPOSED PROBABILISTIC-NEURAL-NET-BASED ALGORITHM

The Probabilistic Neural Network (PNN) is a neural network implementation of the well established multivariate Bayesian classifier, using Parzen estimators to construct the probability density functions of the different classes.

3.1 Multivariate Bayesean Discrimination and the PNN

The objectives are to: [1] separate classes of objects, i.e. define the boundaries between the existing classes, and [2] classify new objects to one of the existing classes. An object is defined by a vector in a p-dimensional input space, where p is the number of features or variables. In the following sections the mathematics will be explained for the case of 2-classes in 2-dimensions. Extension to higher cases can be done in a straight forward manner without loss of generality.

Let $f_1(x)$ and $f_2(x)$ be the probability density functions (PDFs) associated with the p-dimensional input vector X for the populations π_1 and π_2 respectively. A reasonable classification rule that minimizes the Expected Cost of Misclassification (ECM) is to assign a new vector to either class π_1 or class π_2 based on the density ratio, the misclassification cost ratio, and the prior probability ratio as follows :

X belongs to:

π_1 if : $f_1(x) / f_2(x) \geq \{ [C(1|2) / C(2|1)] * [P_2 / P_1] \}$

π_2 otherwise.

where :

- $C(i|j)$ is the cost of misclassifying an object as belonging to population

π_i while it belongs to population π_j.

- P_i is the prior probability of occurrence of population π_i

The key to using the above classification is the ability to estimate the PDFs based on training patterns. Typically, the a priori probability can be estimated, and the cost ratio requires subjective evaluation.

The accuracy of the decision boundaries and the subsequent classification depends on the accuracy with which the underlying PDFs are estimated. A good feature of this approach and the related PNN implementation is estimation consistency. Consistency implies that the error in estimating the PDF from a limited sample gets smaller as the sample size increases. The estimated PDF (the class estimator) collapses on the unknown true PDF as more patterns in the sample become available.

An example of the Parzen estimation of the PDFs is given below for the special case that the multivariate kernel is a product of the univariate kernels. In the case of the Gaussian kernel, the multivariate estimates can be expressed as:

$$f_k(X) = \frac{1}{(2\pi)^{p/2} \sigma^p} \frac{1}{m} \sum_{i=1}^{m} \exp\left[-\frac{(X - X_{ki})^T (X - X_{ki})}{2\sigma^2} \right]$$

where :

k   =   class or category
i   =   pattern number
m   =   total number of training patterns
X_{ki} =   ith training pattern from category or population π_k
σ =   smoothing parameter
p   =   dimensionality of feature (input) space.

Note that the estimated PDF for a given class, say $f_1(x)$ is the sum of small multivariate Gaussian distributions centered at each training sample. However, the sum is not limited to being Gaussian.   It can in fact approximate any smooth density function. The smoothing factor σ can alter the resulting PDF. Larger σ causes a vector X to have about the same probability of occurrence as the nearest training vector. The optimal σ can be easily found experimentally.

An interesting feature of the PNN approach is that the estimated PDFs can be used not only for classification but also to estimate the posterior probability that a vector X belongs to class, say π_1. If the classes are mutually exclusive, we have from Bayes theorem :

$$P[\pi_1 | X] = \frac{P_1 f_1(X)}{P_1 f_1(X) + P_2 f_2(X)}$$

Also the maximum of $f_1(x)$ and $f_2(x)$ is a measure of the density of the training samples in the vicinity of X which can be used to indicate the reliability of the classification.

For more details on the theory the reader is referred to Specht (1996), NeuralWare (1993), and Johnson and Wichern (1992).

The original neural network implementation of the above theory (Specht 1996) is shown in Figure 1 for a 2-class problem. The input units are merely distribution units that supply the same input values to all of the pattern
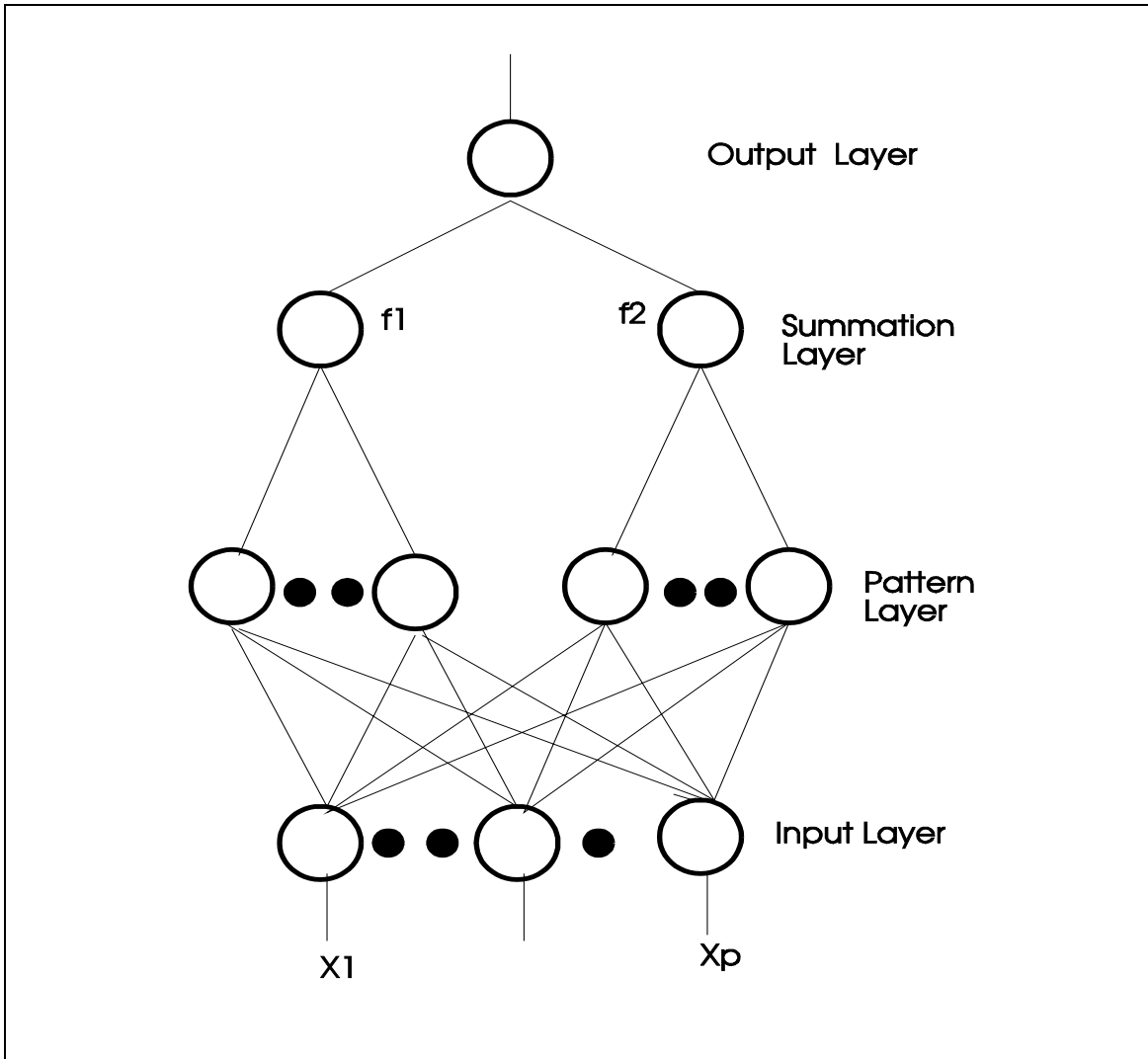
Figure 1. The Original Probabilistic Neural Network (PNN)

units. Each pattern unit forms a dot product of the input pattern vector X with the weight vector W_i such that $Z_i = X.W_i$, and then performs a nonlinear operation on Z_i before outputting its activation level to the summation unit. Instead of the sigmoid activation function commonly used for the MLF the nonlinear operation used here is $\exp[(Z_i - 1) / \sigma^2]$. Assuming that both X and W_i are normalized to a unit length, this is equivalent to using : $\exp[- (W_i - X)^t (W_i - X) /2 \sigma^2]$.

Each summation unit simply sum the outputs from the pattern units that correspond to one of the classes. The output, or decision units are two-input neurons that produce binary outputs. They have only a single variable weight C_k :

$$C_k = \{ \ [\ C(1|2) / C(2|1) \ ] \ * \ [P_2 / P_1] \ * \ [n_1 / n_2] \}$$

where :

$n_1 =$ number of training patterns for category π_1

$n_2 =$ number of training patterns for category π_2

The network is trained by assigning a pattern unit for every training pattern, and setting the W_i weight vectors in each one of the pattern units equal to the corresponding X pattern in the training set, and then connecting the pattern unit output to the appropriate summation unit.

3.2 Statistical Distance and The Proposed Modified PNN2

The PNN uses Euclidean distance as a measure of nearness of the different patterns. Euclidean distance is unsatisfactory for some statistical purposes because it doesn't account for differences in variations along the axes nor the presence of correlation among the variables constituting the pattern vector. The need to consider statistical rather than Euclidean distance is illustrated heuristically in Figure 2 where the Euclidean distance from a point such as P to the center of the cluster Q is larger than the Euclidean distance from O to Q. However, P appears to belong to the cluster
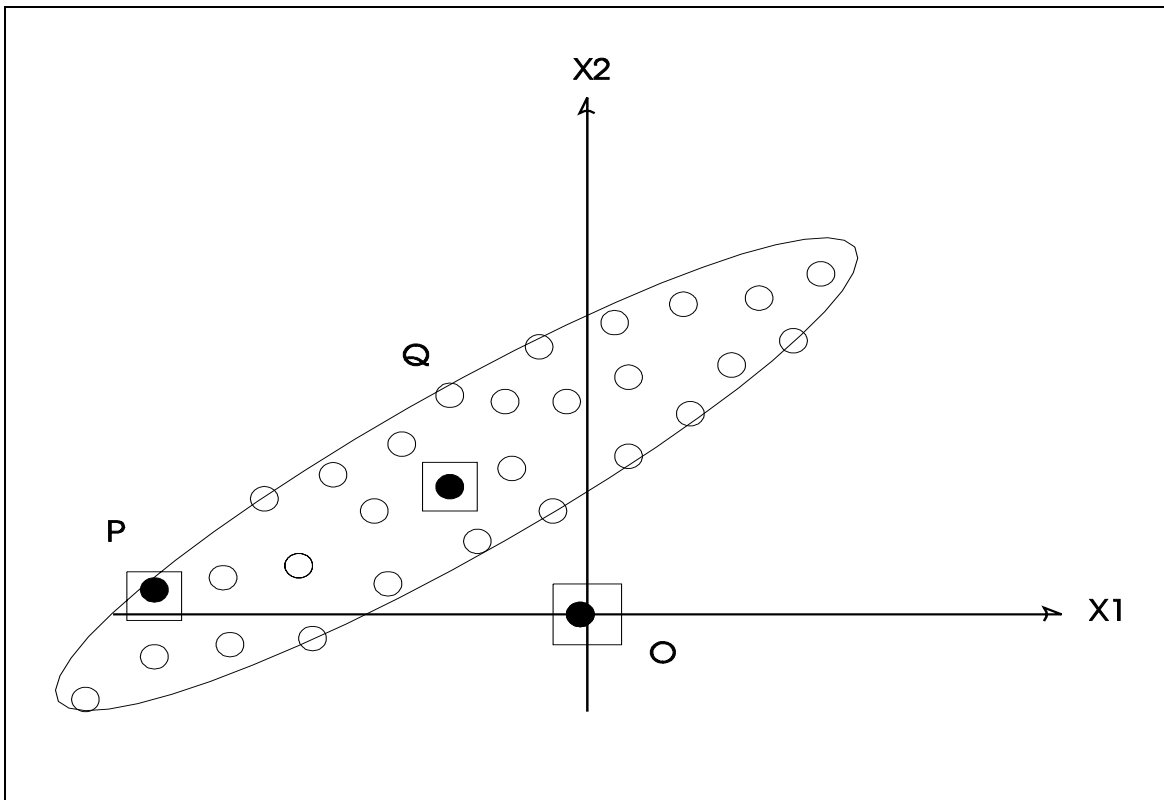
Figure 2. Euclidean versus Statistical Distance

more than O does. If we follow the procedure described above and use the statistical distance instead, then Q will be closer to P than O (Johnson and Wichern, 1992).

Two questions need to be answered in order to complete the discussion. First, how to replace the employed Euclidean distance with the better statistical distance alternative? Second, how to modify the PNN to account for this distance measure ?.

The answer to the first question is to use principal components rather than the original variables. Algebraically, principal components are particular linear combinations of the original set of random variables. Geometrically, these linear combinations represent a new coordinate system of axes by rotating the original ones. The new rotated linear combinations are uncorrelated and their variances are maximized. The following result gives the principal components given the original variables and their covariance matrix (Johnson and Wichern, 1992):

- Let $\sum$ be the covariance matrix associated with the random vector **X**;

- Let $\sum$ have the eigenvalue - eigenvector pairs:

$$(\lambda_1, e_1), (\lambda_2, e_2), \ldots \ldots (\lambda_p, e_p)$$

where

$$\lambda_1 \geq \lambda_2 \geq \ldots \ldots \ldots \lambda_p \geq 0;$$

- The ith principal component is given by

$$Y_i = e_i'X = e_{1i}X_1 + e_{2i}X_2 + \ldots \ldots + e_{pi}X_p \qquad i = 1, 2, \ldots \ldots p$$

and

$$Var(Y_i) = e_i'\Sigma e_i = \lambda_i \qquad i = 1, 2, \ldots \ldots p$$

$$Cov(Y_i, Y_k) = e_i'\Sigma e_k = 0$$

The original input vector **X** is transformed into the rotated vector **Y** using the above relations.   It should be emphasized that the original input

vector is unchanged, but the coordinate system used for describing the vector is changed. The component variables of the vector in terms of the rotated axes are then divided by their standard deviations $\sqrt{\lambda_i}$ to equalize the variances. By this we obtained a new set of inputs free of the effects of correlation and widely varying variances.

Figure 3 shows our modified version of the PNN (referred to as PNN2) that takes the above transformations into account.

The previous input layer of the PNN is replaced by two layers: an input layer and a transformation layer. The weights between the input layer and the transformation layer are the eigenvectors of the sample covariance matrix. The transfer function in the units of the transformation layer simply divides the weighted input to the unit by the standard deviations $\sqrt{\lambda_i}$. Beyond this layer everything is identical to the original PNN described before, only using the principal components instead of the original variables.

4. SEMI-CONDUCTOR PNN CHIP, AND ROAD SIDE INCIDENT DETECTION

The PNN family has several advantages (Specht 1996) that makes it most suited for incident detection (Abdulhai and Ritchie 1995). These advantages will be evident when we evaluate the PNN relative to the universality template later in this report. However, an additional feature is the possible chip implementation of a PNN-based AID framework. Unlike many other neural networks such as the MLF, the PNN operates entirely in parallel, without the need for feedback from individual neurons to the preceding layer of neurons. For systems involving thousands of neurons that can not fit in a single semiconductor chip, such feedback paths would quickly exceed the number of pins available on the chip. With the PNN, any number of chips could be connected in parallel to the same inputs, a design that has already been implemented (Specht 1996). An important implication of this is the possibility of incorporating incident detection chips in a road side cabinet. All the necessary processing could hence be done on site and only the resulting detection output transferred back to a TMC for further action.
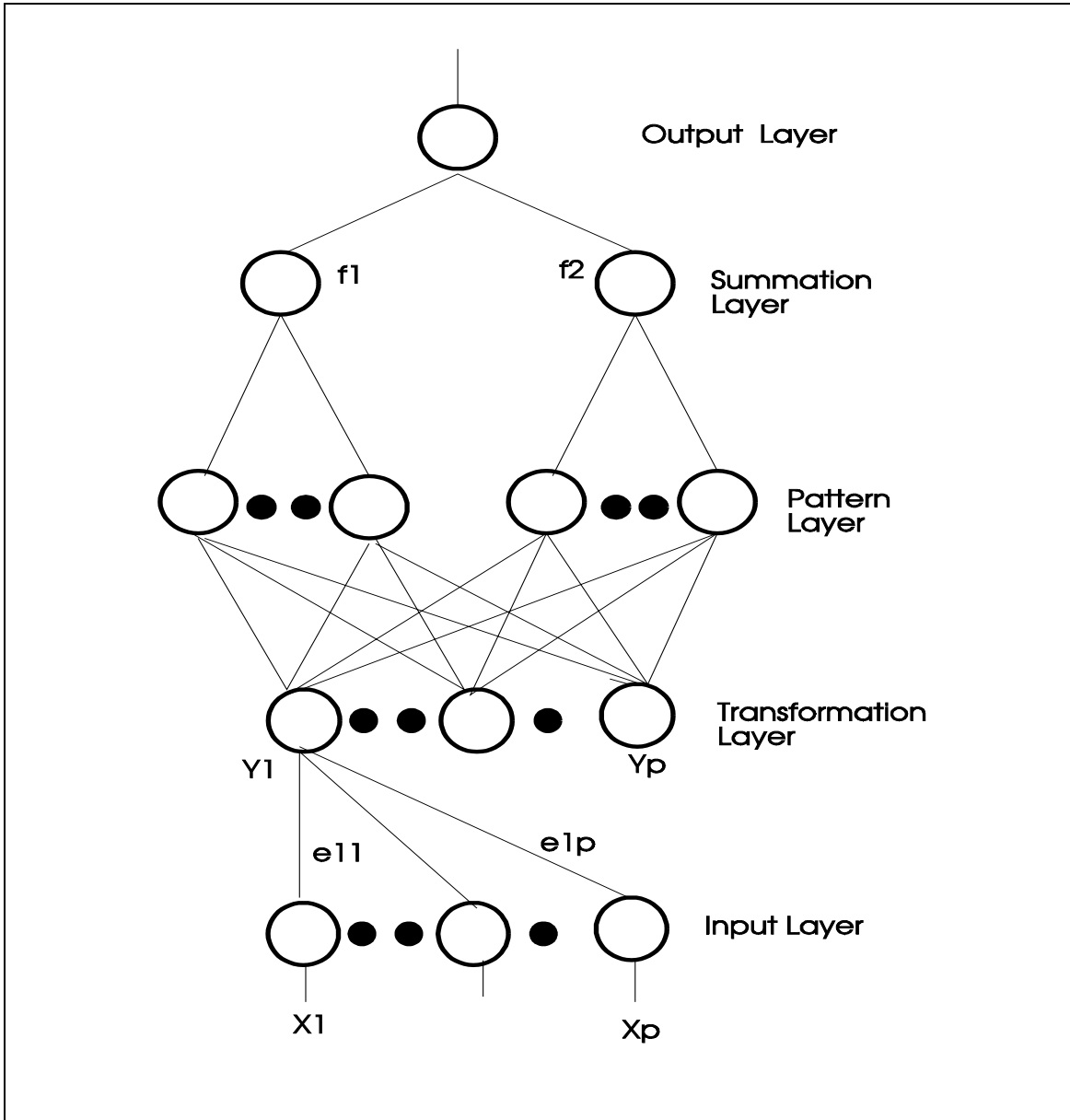
Figure 3. The Modified Probabilistic Neural Network

This adds important flexibility to the incorporation of incident detection into different Intelligent Transportation Systems (ITS) architectures.

5. TMC SURVEYS AND THE IDENTIFICATION OF PERFORMANCE CONSTRAINTS

As part of this research, operational constraints and criteria were elicited from freeway operations personnel in order to assist in evaluating the algorithms' performance. A short survey was prepared and sent to 15 TMCs all over the United States that we thought might be interested in automated freeway incident detection. The actual survey form is shown in Figure 4. The three questions in the survey targeted the identification of what might be acceptable limits of DR and FAR. Seven replies have been received to date.

Analysis of the different replies indicated that, on the average, TMC personnel put more emphasis on a high DR, which is consistent with the prime objective of incident detection. Out of the seven responses received, four TMCs (57%) stated their preference for a high DR in response to question #1 in the survey form. In response to question #2, four TMCs chose system #1 which has a 100% DR despite a relatively high FAR of 5%. One TMC picked system #2, one TMC picked system #3, and one TMC picked alternative #4, rejecting the three systems. The response to the third question, which involved an explicit statement of the acceptable boundaries on DR and FAR, varied from one TMC to another with an average requirement of the DR to be at least 88.3% and of the FAR to be at most 1.8%. The actual responses are listed in table 1.

								Average	Extreme
DR% at least:	80	95	max.[a]	70	90	95	100	88.3%	100
FAR% at most:	1	5	0.25	2	0.5	2	min.[b]	1.8%	0.25

[a] max attainable.
[b] min attainable.

Table 1. DR and FAR Limits Extracted from TMCs' Responses.

UNIVERSITY OF CALIFORNIA, IRVINE

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO   SANTA BARBARA • SANTA CRUZ

INSTITUTE OF TRANSPORTATION STUDIES

IRVINE, CALIFORNIA 92717-3600
(714) 824-5989
FAX (714) 824-8385

    Despite research attempts to develop the 'perfect' incident detection algorithm that would yield 100% Detection Rate (DR) and 0% False Alarm Rate (FAR), this ultimate performance has not been reached yet. There exists a trade-off between a low FAR and a high DR. We define DR as the ratio of the number of incidents successfully detected to the total number of incidents that have occurred. Also the FAR is defined as the ratio of the number of times the algorithm falsely indicates an incident to the total number of times the algorithm is applied (usually once every 30 sec.). For instance, if 10 incidents occur, and the algorithm detects only 8 of them, then the DR is 80%. The FAR computation is a little bit more tricky. For instance, in one hour the algorithm is applied 120 times at a given section. If the algorithm indicates an incident 3 times during this hour while there are no incidents in the field, the FAR is 3/120 = 2.5%.

Please answer the following three questions related to DR and FAR:

- Which is more important to you? please do not answer 'equally important'.
  Low FAR   ____
  High DR   ____

- If you have access to different incident detection algorithms with the following performances, which algorithm would you be most likely to implement?
  #1____   #2____   #3____   none of them ____

	Algo #1	Algo #2	Algo #3
DR%	100%	70%	50%
FAR %	5%	1%	0.25%
Average number of false alarms per hour per mile*	12	2.4	0.6

based on loop stations .5 mile apart and 30 sec apart applications

- Please give percentages for the boundaries of DR and FAR that you believe are acceptable. (For instance DR should be at least 90% and FAR should be at most 1%)
  DR should be *at least*:   ____%
  FAR should be *at most*:   ____%

Traffic Management Center:_____
Contact Person Name & Title:_____
Signature:_____   Date:_____

Figure 4. Survey Form Mailed to TMCs

The above results indicate that a reasonable set of limits on DR and FAR would be 88% and 1.8% respectively. A more stringent set of limits could be obtained from using the extreme value limits of 100% and 0.25% respectively as shown in Table 1. All incident detection algorithms under consideration in this research will be evaluated against both the above two sets of limits.

6. INCIDENT DETECTION USING THE MODIFIED PNN2

6.1 The Study Sites and The Data Sets

To be consistent with previous research (Cheu and Ritchie 1996) and to enable comparison with the incident detection results from the MLF network developed in that research, the original study site, the associated data sets, and the performance measures were kept exactly the same. Also, the prior probability and cost of misclassification ratios of the PNN were set to unity as non of the other algorithms under investigation utilize them. To test the transferability of the algorithms and the potential performance enhancement from real-time on-line retraining, a new site from the I-880 freeway with associated new data sets, were prepared and used.

The original study site is a 5 mile section of the SR-91 Riverside Freeway in Orange County, California, between the SR-57 and the Interstate 5 Freeways. Two data sets associated with the site were used. The first is a large simulated data set, while the second is a limited real data set. The simulated data set was prepared using the well known INTRAS (Integrated Traffic Simulation) model, with which 1780 incidents were simulated. In each case, 27 minutes of loop detector outputs were used which included an incident that lasted for 10 minutes.   For the real data from this site, nine days of field data were used for testing purposes only without retraining. The same data set was described and used by Cheu (1994) to test a collection of algorithms, including the MLF. A total of 63 hours of traffic data were available after removing the periods with missing data due to loop malfunctioning. Each day of data contained one incident as recorded by the

police or in the TMC incident log. Of the nine incidents, one was a minor shoulder incident, one had no details, five involved one lane blocked, and two resulted in multiple lane blockage.

A unique new loop and incident database from California was utilized in this research for transferability evaluation, and for examining the real-time on-line retraining after transferring the PNN2. This database is the result of a comprehensive data collection effort on a fully instrumented section of the I-880 (Nimitz Freeway) in Oakland. The selected section of the I-880 freeway is 49700 feet in length between the Marina exit and Wipple exits. The freeway section is instrumented with inductive loops in the pavement. The instrumentation, however, is confined between the Lewelling and Industrial exits. There are a total of 18 loop stations covering all lanes of the freeway and selected on and off ramps. The spacing between the different loop stations ranges from approximately 1000 feet to 3300 feet. On the main line lanes the detectors are placed in pairs, but on the on and off ramps they are single detectors. The data collected are vehicle counts, occupancy, speeds, and loop on-times. Software specially developed as part of the data collection effort was used for reading the binary loop data and converting it into readable ASCII format. The same program was also used for running several checks in order to fix any abnormalities in the loop database. Analysis of the I-880 incident data revealed that it contains about 45 usable lane blocking incidents. None of these lane blocking incidents, however, produced average upstream occupancies beyond the 40-50% range. Since all the incident detection algorithms in this research were trained on data from a completely different site, testing them on this new data set would give one indication of their transferability.

6.2 Direct Performance Measures

The performance measures used were the Detection Rate (DR), False Alarm Rate (FAR) and Time to Detection (TTD). The DR is defined as

$$DR = \frac{\text{no. of detected incidents}}{\text{total no. of incidents in the data set}} \times 100\%$$

If the algorithm issues an incident warning for a freeway location at a particular time, but in the absence of an actual incident in the field, a false alarm is said to have occurred. The FAR is generally defined as

$$FAR = \frac{\text{no. of false alarms}}{\text{total no. of applications of the algorithms}} \text{ x } 100\%$$

The TTD of an incident is the time in seconds between the actual occurrence of the incident and the time it is detected by the algorithm. For multiple incidents in the data set, the average TTD of the detected incidents is always used. For further details the reader is referred to (Cheu and Ritchie, 1996; Cheu, 1994).

6.3 Implementation and Results Using Simulation Data

Both the PNN and the modified PNN2 were trained using the simulation data to compare their performance with the results from the MLF network previously developed. Several other conventional algorithms were previously calibrated and tested on the same data set, but their performances were reported to be inferior to that of the MLF (Cheu and Ritchie, 1996). Hence the performance of the MLF is used here as an upper-end bench mark.

Each network was trained using a portion of the simulation data set, titled set 1. Performance during training was monitored using another subset of the simulation data, set 2. Finally, all the networks were tested using another subset of the data, titled data set 3. The DR, FAR, and average TTD were computed separately for single lane blocking incidents and for multilane blocking incidents. Persistence tests of up to three intervals were used to further reduce the FAR due to random fluctuations of traffic. A persistence test of n-intervals means that an incident alarm is declared only if the incident condition is found to persist for (n+1) consecutive applications of the algorithm. Table 2 categorizes the performance measures for one lane and multi-lane blocking incidents on data set 3. The measures computed from these two categories were combined using weights of 0.9 and 0.1 respectively to produce a final weighted DR, FAR, and average TTD. This

_____

was based on a previous study in Los Angeles which found that for lane blocking incidents, 90% were one lane blocking and 10% were multilane blocking (see Guiliano 1989, Cheu 1994). The performance envelopes (DR versus FAR using different persistence tests) are shown in Figure 5 for the PNN, the modified PNN2 and the MLF networks. The results in Table 2 and Figure 5 show that the performance of the original PNN is lower than that of the MLF.   However, using the modified PNN2 which incorporates the statistical distance metric, the performance improves and becomes competitive with the MLF.

Network	P.I.[1]	One-Lane Incidents			Multi-Lane Incidents			Combined		
		DR	FAR	TTD	DR	FAR	TTD	DR	FAR	TTD
PNN	0	97.0	14.00	35	100.0	3.00	70	97.3	12.9	39
	1	72.5	2.40	147	97.6	0.96	105	75.0	2.26	143
	2	67.0	1.60	176	96.3	0.34	137	70.0	1.47	172
	3	55.5	0.84	222	95.9	0.14	173	59.5	0.77	217
PNN2	0	100.0	4.80	73	100.0	4.50	44	100.0	4.77	70
	1	95.1	2.68	172	99.8	1.79	88	95.6	2.59	163
	2	83.7	1.30	229	98.4	0.88	121	85.1	1.26	218
	3	68.4	0.66	260	98.2	0.50	153	71.4	0.64	249.3
MLF	0	78.0	1.50	206	97.0	1.06	87	79.7	1.46	194
	1	65.0	0.44	252	96.0	0.20	114	68.1	0.42	238
	2	56.0	0.22	287	95.0	0.16	146	59.9	0.21	273
	3	46.0	0.18	311	93.0	0.10	175	50.7	0.17	297

1. P.I. = Persistence Interval

Table 2. MLF, PNN, and PNN2 Testing on Data Set 3.

6.4 Evaluation Using Field Data

The real loop data previously described from the SR-91 freeway were used in this testing phase. The loop data during and immediately around each incident were used to compute both the DR and TTD, and the rest of the incident-free data were used to compute the FAR. The detection performance of the PNN2 and the MLF with up to three intervals of persistence is
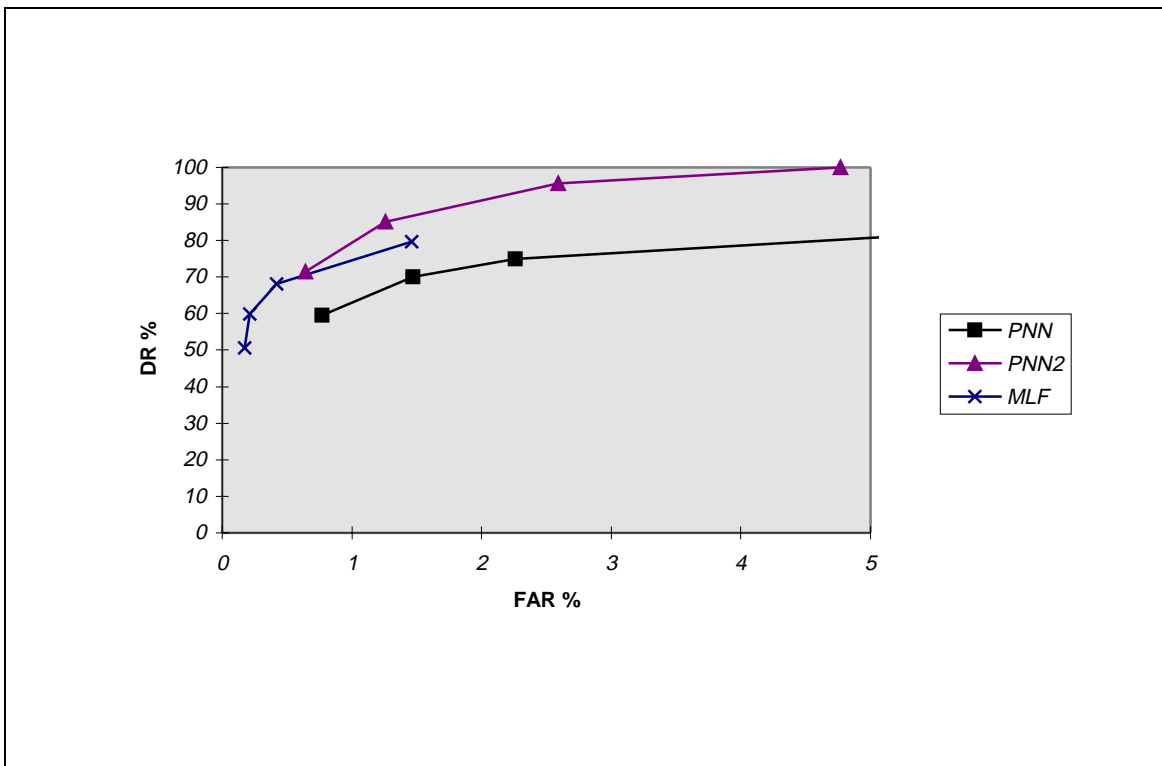
Figure 5. MLF, PNN, and PNN2 Performance

summarized in Table 3. The PNN2 detected the nine incidents in the case of no persistence test. With one or more persistence intervals, the shoulder incident was missed due to its minor effect on the loop data. The PNN2 gave relatively higher FAR in the absence of persistence intervals. However, the FAR dropped significantly by using one or more persistence intervals. Based on this limited data set, both the PNN2 and MLF have similar performance in terms of DR, FAR, and TTD especially with 2 or more intervals of persistence.

Network	Persistence	Combined		
		DR	FAR	TTD
PNN2	0	100	0.39	22
	1	89	0.09	78
	2	89	0.01	108
	3	89	0.00	198
MLF	0	89	0.02	30
	1	89	0.01	86
	2	89	0.01	116
	3	89	0.00	146

Table 3. MLF and PNN2 Testing on Field Data.

7. TRANSFERABILITY TESTING AND REAL-TIME PERFORMANCE IMPROVEMENT

7.1 Transferability Testing and Evaluation

The real loop data and reported incidents from the I-880 freeway in Oakland California were used to test the developed algorithms which were trained on simulation data from the SR-91 freeway in Orange County. The loop data corresponding to the lane blocking incidents were extracted and formatted for testing the algorithms. One hundred hours of incident-free loop data were also randomly extracted to use for FAR testing.

Both the PNN2 and the MLF were tested on the prepared I-880 incident and normal data sets. Table 4 shows the testing results. The DR and TTD

were obtained from testing on the incident data set and the FAR was obtained from testing on the incident-free data set. Up to three intervals of persistence were used.

Network	Persistence	DR	FAR	TTD
PNN2	0	31	0.40	1080
	1	29	0.02	1188
	2	29	0	1218
	3	27	0	902
MLF	0	35	0	306
	1	35	0	336
	2	35	0	366
	3	35	0	396

Table 4. Testing Results on the I-880 Data Set.

The above results indicate the following:

1. Based on the performance limits established earlier, neither of the algorithms showed acceptable performance.

2. The nature of the incidents in the I-880 database probably caused this obviously low performance. This is due to the fact mentioned earlier that the upstream occupancies were consistently low and did not exceed the 40-50% range. On the other hand, the SR-91 data used for training showed more pronounced effect of the incidents, as the upstream occupancies were as high as 80-100%.   The large difference between the characteristics of the training SR-91 data and the testing I-880 data is the prime cause of the common low performance amongst all algorithms.

3. Since the statistics of the traffic in the new testing site are significantly different from the training site as explained in "2" above, recalibration or retraining on data from the new site is necessary for the algorithms to show acceptable performance after they are transferred. However, retraining or recalibrating an incident detection algorithm for each and every new site is usually not a trivial task.   This problem motivated us to utilize the

instantaneous learning capability of the PNN class of networks to counter balance the severe drop in performance due to transferability.

7.2 On-Site Real-Time Performance Improvement

As was briefly discussed earlier, the PNN family of neural networks uses the training patterns as connection weights to represent the knowledge content of the network. Since the performance of the network deteriorates after transferability, it will detect less than 100% of the incidents that occur in the field. How much less than 100% the DR will be depends on how different is the new site and how far or close are the estimated PDFs during initial training relative to the actual unknown PDFs from the new site. Nevertheless, the algorithm will detect some incidents. If the traffic patterns for detected incidents in the new location are used to update the knowledge content of the network by overwriting the older patterns in real time, the performance would gradually improve as more and more incidents are detected with time in service. Updating the non-incident patterns is not a problem as such traffic data are always available.

To verify this theoretical hypothesis, the detected incidents from the I-880 incident database were isolated and used to update the weights of the PNN2 (about one third of the available lane blocking incident data). The same was repeated for the correctly classified non-incident patterns. After the network was *instantly* retrained with the I-880 data that it managed to correctly recognize in the first round of testing, a second round of testing was performed on the entire database in order to assess the improvements. As shown in Table 5 below, the retraining or updating process resulted in significant improvement in the performance of the network. The DR improved from only 30% to 98% and the FAR dropped to 0%. Similar significant improvement in the TTD was also evident.

	PNN2					
	Before Patterns Update			After Patterns Update		
Persistence	DR	FAR	TTD	DR	FAR	TTD
0	31	0.40	1080	100	0.5	15
1	29	0.02	1188	98	0	79
2	29	0	1218	98	0	112
3	27	0	902	98	0	142

Table 5. Performance Improvements After Real-Time Updating of the Training Patterns.

7.3 Discussion

The following several points should be noted:

1. The amount of incident data needed from the new I-880 site is less than one third of the actual number of incidents in a two month period. Thus, had the updating process been implemented on line, the network would have taken only two months to improve its performance from a totally unacceptable level of performance to almost an ideal performance of 100% DR and 0% FAR.

2. Testing the PNN2 algorithm on the I-880 data set was a fairly harsh test since the effects of the incidents on traffic conditions were not severe, and hence were less detectable. In an average case, we would expect the drop in performance due to transferability to be less than that for the I-880 test. The on-line time required for the PNN2 to achieve ideal performance could then also be less than 2 months, assuming a similar rate of incident occurrence.

3. No other algorithm or network that we know of can achieve a similar "*on-line*" improvement in performance. This is due to the nature of the training process.   The MLF, for instance, uses the training patterns to develop connection weights using an error minimization technique.   When new patterns become available at the new site, the network has to be retrained off-line, presumably by a user TMC, to reach a new set of weights that minimize the error given the new conditions. In fact, the MLF, when trained off-line on the same data, achieved similar optimal performance to the PNN2.

However, the intensive process of off-line re-training or re-calibration at each and every new site is what we are trying to avoid, so long as there is not a significant difference in the ultimate performance.

4. The "real-time" learning capability of the PNN should not be confused with clustering algoritms and related unsupervised ANN's, which might create new classes during learning in real time. The PNN improves in classifying the given classes (e.g. incident versus non-incident traffic) without creating or even proposing new classes.

5. As false alarms are possible, one might correctly argue that the detected incidents should not be used directly to update the training patterns without incident verification. If incident verification is a necessary step then the updating process could only take place in "pseudo-real time", delayed by the amount of time required to verify the truth of an incident alarm, and perhaps conditioned upon an "update permission" by a TMC operator. However, it might be possible to avoid conditioning the update process on incident verification. This could be achieved by utilizing the "cost of misclassification ratio" discussed earlier which could be set in such a way that a 0% FAR is favored over a high DR. This way, the network would have a 0% FAR right after transferability at the expense of an initially low DR. As the network started to detect a few _true_ incidents and update its weights, the DR would start to improve gradually with time in service.

8. POTENTIAL UNIVERSALITY OF THE PNN FOR INCIDENT DETECTION

In the following section, we evaluate the proposed framework relative to the universality template defined earlier on the basis of both the theoretical aspects of the PNN and the obtained results. Only the algorithm-related universality attributes will be discussed and not the four feature-extraction-related attributes, the discussion of which is beyond the scope of the present report. The attributes will be discussed in the same order as in the universality template.

- **High Performance**: Results from testing the PNN2 on both the real SR-91 data and the real I-880 data indicate that the performance of the network satisfies the TMC requirements established from TMC surveys.

- **Fast Training / Calibration**: Specht (1990) reported that, in one particular application, the PNN paradigm was 200,000 times faster than the MLF. In our particular incident detection application, the PNN learned 35,000 training vectors in less than one minute of CPU time on a SUN SPARCstation IPX. The MLF on the other hand took approximately 13 hours of CPU time on the same machine. The training of the PNN, in this particular application, was approximately 1000 times faster than the MLF.

- **Reasonable TMC implementation requirements**: As was hopefully obvious from the discussion of the real-time on-line retraining of the PNN and the instantaneous nature of the training process, the additional demands placed on TMC personnel to participate in training of the algorithm are minimal or non-existent.

- **Transferable logic**: The theory of the PNN relies on the estimation of the generic PDFs of the different pattern classes, and hence it is not limited to particular distributions. Whatever the actual probabilistic distributions are, the PNN generates statistically *consistent* estimates for them.

- **Transferable training/calibration parameters**: As was found in the above results, and as expected, all algorithms including the PNN can suffer significant performance deterioration after transferability. However, the on-line retraining results of the PNN show its great potential to solve the transferability problem.

- **Minimal initial training data requirements**: Consistency of the estimated PDFs allows the PNN to perform as soon as one pattern representing each category has been observed. The network can begin to generalize to new patterns. As additional incident and incident-free patterns are observed and stored in the network, the generalization will

consistently improve and the decision boundary can become more complex.

- **Account for prior probabilities of incidents**: the PNN uses the prior probabilities to alter the decision accordingly. Although the utilization of the prior probabilities is an integrated component of the PNN, the actual computation of the prior probabilities of incidents is beyond the algorithm. Existing incident prediction algorithms could be readily "plugged in" to the PNN framework to supply the required prior probabilities (see for instance Madanat, 1995).

- **Account for the cost of misclassifying traffic patterns**: the PNN uses the relative costs of misclassification to alter the decision accordingly. The actual relative cost values are in fact TMC-dependent, and the PNN allows these to be set by the user TMC.

- **Capable of producing the posterior probability of an incident**: the actual output of the PNN takes the form of probabilistic estimates of an incident or incident-free situation. i.e. the algorithm does not require any modification to produce probabilistic outputs. On the contrary, if the user needs a binary incident/incident-free output then the actual probabilistic output has to be converted to a binary decision.

- **Estimate incident severity**: the probabilistic output of the PNN can be interpreted as an estimate of the severity of the detected incident. Severe incidents cause more pronounced freeway blockage and the PNN would consequently produce higher probabilities.

- **Capture incident duration**: the duration of the detected incident could be estimated directly from the time-profile of the probabilistic output of the PNN. The duration of the incident would be from the time the incident probabilities start rising to the time the incident probabilities start falling. The concept could be extended in a straight forward manner to capture the return of traffic conditions to normal.

- **Statistically and theoretically well established with minimal heuristics**: the PNN is based on the well known and accepted Bayes theorem, and also the well known Parzen windows concept. Both Bayes

theory and Parzen windows have been used for decades and are treated in length in a variety of statistical text books.

9. CONCLUSIONS

In this research we introduced a new potentially universal freeway incident detection framework. A number of attributes for an incident detection algorithm to be universal were defined, based on TMC requirements and our own experience. The core algorithm is based on the Probabilistic Neural Network modified to utilize the concept of statistical distance. The algorithm has several peripheral attributes that complement the potential universal applicability of the overall framework. Results using both simulation and real incident data demonstrated its competitiveness with the best neural network incident detection algorithm developed previously, namely the MLF algorithm. The PNN was competitive with the MLF in initial classification accuracy in terms of DR, FAR, and TTD. In addition, the theory of the PNN shows its capability to learn in real time by overwriting the older patterns stored in the connections weights by newer patterns reflecting the updated situation. The updated situation could reflect changes in the site or a completely new site. The correctness of this theoretical hypothesis was demonstrated using transferability testing on a real incident data set from the I-880 freeway in Hayward, California. The obtained results indicated a very promising and relatively effortless improvement in performance. Thus, the PNN is transferable to new sites without the need for *explicit off-line retraining,* as its performance improves with time in service. Moreover, the algorithm was shown to fulfill the entire set of universality conditions previously defined. Continued research is geared toward more extensive testing of the framework on real data sets from other locations.

REFERENCES

· Abdulhai Baher and Ritchie S. G. (1995), "Performance of Artificial Neural Networks for Incident Detection in ITS", ASCE Transportation Congress, pp. 227-238.

· Cheu, R. L., and Ritchie, S. G. (1996), "Automated Detection of Lane-Blocking Freeway Incidents using Artificial Neural Networks", accepted for publication, Transportation Research-C.

· Cheu, R. L. (1994), "Neural Network Models for Automated Detection of Lane-Blocking Incidents on Freeways.", Ph.D. Dissertation, University of California Irvine.

· Gall, A. I., and Hall, F. L. (1989), "Distinguishing between Incident Congestion and Recurrent Congestion: A Proposed Logic", Transportation Research Record 1232, pp. 1-8.

· Giuliano, G. (1989), "Incident Characteristics, Frequency, and Duration on a High Volume Urban Freeway", Transportation Research Vol. 23A, No. 5, pp. 387-396.

· Hall, A. L., Hall, F. L., Shi, Y., and Lyall, B. (1991), "A Catastrophe Theory Approach to Freeway Incident Detection", Proceedings of the Second International Conference of Applications of Advanced Technologies in Transportation Engineering, Edited by Stephanedes, Y. J. and Sinha, K. C., American Society of Civil Engineers, pp. 373-377.

· Johnson, R. A., and Wichern, D. W. (1992), "Applied Mutivariate Statistical Analysis", Prentice Hall, Inc.

· Lin, W. (1995), "Incident Detection with Data from Loop Surveillance Systems: The Role of Wave Analysis", Ph.D. Dissertation, UCB-ITS-DS-95-4, University of California Berkeley.

· Madanat, S. and Teng, H.L. (1995), "A Sequential Hypothesis Testing Based Decision Making System for Freeway Incident Response", Final Report for the IDEA Program, Transportation Research Board, National Research Council, Washington D.C.

· NeuralWare (1993), "NeuralWorks Professional II/Plus Neural Networks Software Manuals.

· Payne, H. J., and Thompson. S. M. (1996), " Development and Testing of Incident Detection Algorithms: A Status Report on the Ongoing FHWA-Sponsored Project", Transportation Research Board 75th Annual meeting, Preprint #961358.

· Payne, H. J., Helfenbein, E. D. and Knobel, H. C. (1976), Development and Testing of Incident Detection Algorithms, Vol. 2: Research Methodology and Results, Report No. FHWA-RD-76-20, Federal Highway Administration.

· Ritchie, S. G,. and Cheu, R. L. (1993), "Neural Network Models for Automated Detection of Non-Recurring Congestion.", PATH Research Report, UCB-ITS-PRR-93-5, University of California Irvine.

· Specht, D. F. (1996), "Probabilistic Neural Networks and General Regression Neural Networks", in Fuzzy Logic and Neural Network Handbook, C.H. Chen, Editor.

· Stephanedes, Y. J. and Hourdakis J. (1996), "Transferability of Freeway Incident Detection Algorithms", Transportation Research Board 75th Annual meeting, Preprint #961214.

_____

·   Stephanedes, Y. J. and Chassiakos, A. P. (1993), "Application of Filtering Techniques for Incident Detection", Journal of Transportation Engineering, ASCE, Vol. 119, No. 1, pp. 13-26.