Data Fusion for Traffic Incident Detection Using D-S Evidence Theory with Probabilistic SVMs

Dehuai Zeng[1,2] Jianmin Xu[1]
1. School of Civil Engineering and Transportation, South China University of China, Guangzhou, China
Email: dhzeng@scut.edu.cn

Gang Xu[2,3]
2. Institute of Intelligent Technology, Shenzhen University, Shenzhen, China
3. Shenzhen Key laboratory of mould advanced manufacture, Shenzhen, China
Email: xugang@szu.edu.cn

Abstract—**Accurate Incident detection is one of the important components in Intelligent Transportation Systems. It identifies traffic abnormality based on input signals obtained from different type traffic flow sensors. To date, the development of Intelligent Transportation Systems has urged the researchers in incident detection area to explore new techniques with high adaptability to changing site traffic characteristics. From the viewpoint of evidence theory, information obtained from each sensor can be considered as a piece of evidence, and as such, multi-sensor based traffic incident detector can be viewed as a problem of evidence fusion. This paper proposes a new technique for traffic incident detection, which combines multiple multi-class probability support vector machines (MPSVM) using D-S evidence theory. We present a preliminary review of evidence theory and explain how the multi-sensor traffic incident detector problem can be framed in the context of this theory, in terms of incidents frame of discernment, mass functions is designed by mapping the outputs of standard support vector machines into a posterior probability using a learned sigmoid function. The experiment results suggest that MPSVM is a better adaptive classifier for incident detection problem with a changing site traffic environment.**

Index Terms—**traffic incident detector, evidence theory, support vector machine, data fusion, pattern recognition**

I. INTRODUCTION

Traffic incidents have become a serious problem not only in developed countries, but also in developing countries. Traffic congestions or accidents caused by incidents have cost the world billions of dollars a year in the lost productivity, property damage, and personal injuries. It is critical to detect and handle traffic incidents as promptly as possible so as to reduce the adverse effects of incident. Provide comprehensive and accurate traffic data for traffic incident recognition is one of important issues in Intelligent Transportation Systems (ITS), since ITS aim to provide road users and traffic managers with accurate and reliable real-time traffic information [1]. As the principal component of ITS data, the massive urban road traffic flow data provide the data support for real-time traffic control, traffic management, transportation planning, and so on. There is a particular need for highly accurate traffic data, measured by accurate and reliable sensors, yielding a high degree of acceptance and credibility concerning the significance of the measured traffic parameters. Today, traffic information comes from a variety of sources such as inductive loops, video observation and floating car data (FCD) [2]. However, to get comprehensive information of nearly all lanes and turning movements based on measurement equipment is not realistic. Such supply coverage would be too expensive for most public budgets. Furthermore, for many reasons such as transmission equipment failures and changes in environmental factors, the collected traffic flow data always have certain types of quality problems.

Raw traffic data obtained from detectors may contain a lot of corrupted or missing data items [3]. Real-time traffic data from loop detectors are inevitably corrupted by unexpected missing values or appear to be giving nonsensical or erroneous data due to detector faults or transmission distortion. In most traffic data sets, corrupted input failures or missing values occurred at some time-spots of the traffic data time-series (temporal errors), due to temporary power or communication failures in the traffic surveillance system. On the other hand, missing data and input failures inevitably occurred in a whole series of detector data (spatial errors), due to damages to detectors or roadside equipments. It is imperative that a fusion mechanism be devised so as to minimize such imprecision and uncertainty. The

This work is supported by National High-Tech Research and Development (863) program with Grant No. 2006AA11Z211, Natural Science Foundation of China with Grant No. 50578064, and China Postdoctoral Foundation with Grant No. 20070410827.

effectiveness of such mechanism depends to a large extent on how redundant and complementary are the information cues obtained from the sensors.

The multi-source ITS data fusion has become one of the most important ITS data quality control techniques to resolve the data quality problems in the field of ITS. Therefore, the focus is to improve traffic demand estimation using sparse detector data including FCD as input combined with algorithms based on integrated data fusion techniques. These techniques contribute additional information on turning movements, flow, delay and routes and in this way enhance data coverage and data quality. The benefit of such a system will be the improved accuracy of the estimation by using the existing detection infrastructure.

Information fusion technique was introduced into the problems of traffic state estimation, expecting to get more accurate and integrated detected data to reflect real traffic state. Using multi-sensor information comprehensively and effectively, it can dominate the information of the same observed object reasonably and combine multi-sensor information in time and space under certain criterion, in order to obtain synthetically optimal estimation. Information fusion technology not only can wipe off redundancy but also obtain more accurate and integrated estimation than from any single source. Although information fusion technology has recently been applied in transportation, majority of existed traffic management systems have employed the technology now. In recent years, the rapid ITS development has further promoted the application of the information fusion technology in transportation [4, 5]. The characteristics of freeway incident management system determine that the information fusion technology is especially suitable for freeway detection domain. However, the present situation that most existed incidents detection systems are built based on single-source decides the necessity and urgency of the research on information fusion technology in incident detection.

At present, Conventional techniques for incident detection include decision trees [6], time series analysis [7], Kalman filters [8] and Artificial Neural Network (ANN), and so on. Of the existing methods, traditional Kalman Filter is sensitive to the abnormal data when processing the massive amount of data and the calculation of its filter value is unstable. Furthermore, the ANN technique presents such disadvantages as overfitting, dimensional disaster and the inherent limitation of local minimum instead of global minimum. Support Vector Machine (SVM) is a new approach of machine learning proposed by Vapnik et al. [9, 10], which is developed based on the statistical learning theory and the principle of the structure risk minimization. Because of its excellent learning characteristics SVM has become a fast evolving research topic in the area of machine learning.

In decision level fusion, a good many technology of decision level fusion can be used in ITS, such as, Bayesian inference, D-S evidence theory and fuzzy logic etc. D-S evidence theory is suitable to taking into account the disparity of knowledge types due to the fact that it is able to provide a federative framework, and combine cumulative evidences for changing prior opinions in the light of new evidences [11, 12]. Therefore, the study and application of D-S evidence theory for information fusion attract researchers' interests [13~16]. We still face some challenges during using D-S evidence theory to reason multi-sensor information, for example, how to determine the Basic Probability Assignment (BPA) from evidence. However, there hasn't any uniform method of determining the BPA up to now, due to the complexity of multi-sensor information representation. In this paper, we extend the SVM to yield an output in the frame of D-S theory. The output of this kind of SVM can directly be combined by using the combination rule of evidences.

Using the geometrical interpretation of the classifying hyperplane and the distance of the pattern from the hyperplane, one can calculate the posterior probability in binary classification case. Many researchers proposed to solve this problem. Vapnik suggested a method for mapping the output of SVMs to probability by decomposing the feature space. Hastie and Tibshirani [21] fitted probabilities to the output of an SVM by using Gaussians to the class-conditional densities $p(y=1|f)$ and $p(y=-1|f)$. Another method proposed by Platt trains the parameters of an additional sigmoid function to map the SVM outputs into probabilities. The results were promising, but they did not extend their method to multi-class phase. Therefore, the D-S theory based multi-class SVM is constructed by designing the BPA according to the multi-class probabilities SVM in this paper.

This work focuses on the development of Automatic Incident Detection (AID) technique using D-S evidence theory data fusion based on probabilistic output of multi-class SVM. The remainder of this paper will be organized as follows. Section 2 presents an introduction to D-S evidence theory, where some basic concepts of decision fusion are introduced. Section 3 is devoted to the probabilistic output of multi-class SVM classifiers. Section 4 constructs the traffic incident detecting structure based on D-S evidence theory combing MPSVM. Section 5 details the experiments on a real traffic data and check their performance against rough sets theory and multilayer neural network. Finally, conclusions are drawn in section 6.

II. Brief Review of D-S Evidence Theory

D-S evidence theory, a statistical-based data fusion classification algorithm, is used when the sensors (or more generally, the information sources) contributing information cannot associate a 100 percent probability of certainty to their output decisions. D-S evidence theory applies belief function as measurement, which allows one to quantify the confidence that a particular event could be the one observed. Then, while new information arrives, the identification system integrates it using conditioning rules to provide a representation of the obviousness of the situation [17~19]. In the following, terminology of theory of evidence and the notation used in this paper are defined.

(1) *Frame of discernment*

If Θ denotes the set of $\theta_N (\theta_N \in \Theta)$ corresponding to N identifiable objects, let $\Theta = \{\theta_1, \theta_2, \cdots, \theta_N\}$ be a frame of discernment. It is composed of N mutually exhaustive and exclusive hypotheses. The power set of Θ is the set containing the all the 2^N possible subsets of Θ, represented by $P(\Theta)$:

$$P(\Theta) = \{\Phi, \{\theta_1\}, \{\theta_2\}, \cdots, \{\theta_N\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \cdots, \Theta\}$$

Where Φ denotes the null set. The $\{\theta_N\}$ subsets containing only one element are called singletons.

(2) *Basic probability assignment (BPA) function*

When the frame of discernment is determined, BPA function m mapping of the power set $P(\Theta)$ to [0,1] is defined by $m: P(\Theta) \to [0,1]$ and which satisfies the following conditions:

$$\sum_{A \in P(\Theta)} m(A) = 1, \ m(\Phi) = 0 \tag{1}$$

m(A) expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Θ belongs to the set A but to no particular subset of A. The elements of $P(\Theta)$ that have none-zone mass are called focal elements. A body of evidence is the set of all focal elements elements. And the union of all the focal elements is called a kernel of mass function m in Θ.

(3) *Belief and plausibility functions*

Given a BPA m, a belief function *Bel* is defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{2}$$

and a plausibility function *Pl* is defined as:

$$Pl(A) = 1 - Bel(\overline{A}) = \sum_{A \cap B \neq \Phi} m(B) \tag{3}$$

The belief function *Bel*(A) measures the total amount of probability that must be distributed among the elements of A; it reflects inevitability and signifies the total degree of belief of A and constitutes a lower limit function on the probability of A. While *Pl*(A) denotes the extent to which we fail to disbelieve A. The belief interval [*Bel*(A), *Pl*(A)] reflects uncertainty. The interval-span *Pl*(A)–*Bel*(A) represents the ignorance in hypothesis A.

(4) *Combination rule of evidence*

In the case of imperfect data, fusion is an interesting solution to obtain more relevant information. Evidence theory offers appropriate aggregation tools. Suppose m_1 and m_2 are two mass functions formed based on information obtained from two different information sources in the same frame of discernment; according to Dempster's orthogonal rule [10] we have

$$m(C) = (m_1 \oplus m_2)(C) = \begin{cases} 0, & C = \Phi \\ \dfrac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - K}, & C \neq \Phi \end{cases} \tag{4}$$

Where K represents a basic probability mass associated with conflicts among the sources of evidence. The conflict K is defined as:

$$K = \sum_{A \cap B = \Phi} m_1(A) m_2(B) < 1 \tag{5}$$

K is measures the degree f of the conflict between m_1 and m_2. The denominator 1-K in Eq. (5) is a normalization factor. K=0 corresponds to the absence of conflict between m_1 and m_2, whereas K=1 implies complete contradiction between m_1 and m_2. The produced function m is also a mass function in the same frame of discernment. Note that $m = m_1 \oplus m_2$, which represents the combination of m_1 and m_2 and carries the joint information from the two sources.

Dempster's combination rule can be generalized to more than two hypotheses, the belief function resulting from the combination of J information sources S_J is defined as:

$$m(C) = m_1(S_1) \oplus m_2(S_2) \oplus \cdots \oplus m_J(S_J) = \frac{\sum_{\cap_{i=1}^{J} S_J = C} \left(\prod_{i=1}^{J} m_i(S_J) \right)}{1 - \sum_{\cap_{i=1}^{J} S_J = \Phi} \left(\prod_{i=1}^{J} m_i(S_J) \right)} \tag{6}$$

Where S_1, S_2, …, S_J are focal elements.

A key problem in Dempster–Shafer evidence reasoning is the calculation of the mass function based on the information provided by the sources of information (e.g., sensors), once the frame of discernment is established. In order to apply Dempster-Shafer theory for the evaluation of the traffic detection sensor multi-classifier, it is necessary to derive BPA S_J from the outputs of the individual classifiers. In general, a belief function can be transformed into a posterior probability produced by a classifier is convenient for post processing. Platt [20] proposed a probability SVM (PSVM) method for fitting a sigmoid function that maps SVM outputs to posterior probabilities.

The essential principle of D-S evidence theory's application in traffic state recognition is as follows: we observe the pattern to be recognized through multi-feature and multi-classifier, then, combine the results from the observation of multi-class SVM classifier by applying D-S evidence theory. At last, we can reach a decision on classification and recognition according to fusion result.

III SVM FOR MULTI-CLASS CLASSIFICATION PROBLEM

SVMs discriminate two classes by fitting an optimal linear separating hyperplane (OSH) to the training samples of two classes in a multidimensional feature space. The optimization problem being solved is based on structural risk minimization and aims to maximize the

margins between the OSH and the closest training samples—the so-called support vectors. For linearly inseparable cases, it maps the sample space into a high-dimensional or infinite feature space through some specific non-linear mapping, and seeks an optimal separating hyperplane in feature space which is regarded as classifier decision surface.

A. Support vector machines for binary classification

Given a training data set of $(x_i, y_i)(i=1,2,\ldots,l)$ where $x_i \in R^n$ and $y_i \in \{+1,-1\}$. In our case, vectors x_i are the sensor responses during one cycle or vectors composed by the features extracted from those responses, whereas y_i are the labels associated to each pattern indicating one of the classes. Thus, a two-class problem basically consists of finding the optimal hyperplane that separates the samples label −1 from those label +1, with a given margin between one set and the other. Such a hyperplane is found when the margin is maximal. SVMs optimize the classification boundary by separating the data with the maximal margin hyperplane.

The hyperplane $f(x)$ is defined by the normal vector $w \in R^n$ and the bias $b \in R$, where $|b|/\|w\|$ is the distance between the hyperplane and the origin, with $\|w\|$ as the Euclidean norm form w

$$f(x) = w \cdot x + b \tag{7}$$

The support vectors lie on two hyperplanes $w \cdot x + b = \pm 1$, which are parallel to the OSH. Maximization of the margin (the distance between the hyper plane and the nearest point) leads to the following optimization problem, minimize:

$$\begin{aligned} &\varphi(w,\xi) = \frac{1}{2}(w \cdot w) + C\sum_{i=1}^{l}\xi_i \\ &s.t.\ \ y_i((w \cdot x_i) + b) \geq 1 - \xi_i \\ &\quad \xi_i \geq 0, i = 1,2,\ldots,l \end{aligned} \tag{8}$$

where the slack variables ξ_i and the regularization parameter C are introduced to deal with misclassified samples in a inseparable case. The constant C is added as a penalty for cases which are located on the wrong side of the hyperplane. Using so-called kernel methods, the above linear SVM approach is extended for nonlinear separable cases. Based on a nonlinear mapping of the data into a higher dimensional feature space, e.g., an OSH can be fit to a more complex class distribution, which is inseparable in the original feature space. The input sample x can be described by in the new high dimensional space. The computationally extensive mapping of in a high dimensional space is reduced by using a positive definite kernel k, which satisfies the Mercer conditions [22]. Thus, we can find that is easier to solve the dual problem:

$$\begin{aligned} &\min W(\alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \\ &s.t.\ \ 0 \leq \alpha_i \leq C, \\ &\quad \sum_{i=1}^{l}\alpha_i y_i = 0, i = 1,2,\cdots,l \end{aligned} \tag{9}$$

where $K(x_i, x_j)$ is the kernel function. In this work, we have used a radial function basis (RBF) kernel for the binary classifiers. Consequently, the final hyperplane decision function can be defined as

$$f(x) = sign[(\sum_{i=1}^{N_s}\alpha_i y_i K(x, x_i)) + b] \tag{10}$$

where α_i are Lagrange multipliers, N_s is the number of support vectors found as a result of the optimization problem, x_i is the support vectors and b is a threshold parameter updated in the training phase. The classification rule depends on the sign of $f(x)$.

B. Probailistic output of SVMs

The support vector machine (SVM) works well in two-class classification, but the output $y_i \in \{+1,-1\}$ of SVM classifier is only qualitative, which can not be used to obtain the posterior probability. In order to apply Dempster-Shafer theory for the decision fusion of the SVM classifier output, it is necessary to derive basic probability assignments from the output of the individual classifier. Our implementation of the Platt's probabilistic SVM outputs [20] includes the modifications suggested by Lin et al [23] for numerical stability.

Analytic geometry can be used to provide an explanation of the meaning of the outputs of the SVM classifiers. The classification of the SVM is given by Eq.(7), The relation is invariant under a positive rescaling of the argument inside Eq.(7), thus a canonical hyperplane is defined so that $\|f(x)\| = 1$ for the closet point. It is clear that for a given hyperplane $f(x) = 0$, and for a vector x that does not belong to the hyperplane, we have:

$$f(x) = \pm d\|w\|$$

where, d is the distance from the point x to the given hyperplane. The different signs determine the side of the hyperplane for the vector x. So we can see that the output $f(x)$ of the SVM is actually the multiplication of the norm of the vector w and the distance from the chosen hyperplane, and analogously

$$d_x = \frac{f(x)}{\|w\|} \tag{11}$$

And the margin between the canonical hyperplane and the closet points is:

$$d_{\mathrm{sv}} = \frac{1}{\|w\|} \tag{12}$$

Clearly, the ration of d_x and d_{sv} is $f(x)$. Using the rate of the distance, we can convert the output of the SVM to the posterior probability. Platt [20] proposed to use a parametric model to adjust this a posteriori probability by means of a sigmoidal function. Instead of directly estimating the a posteriori probability, we will estimate the parameters A and B of a sigmoid:

$$P(y=1 \mid f) = \frac{1}{1+e^{Af+B}} \tag{13}$$

As long as A<0, the monotonicity of Eq.(13) is assured.

Parameters A and B can be found using any regularization method. In our case, we followed Platt's criteria by means of maximizing information entropy estimation. The entropy of probability distribution $P = (p_1, p_2, \ldots, p_n)$ is defined as:

$$H(P) = E(-\log P) = \sum_{i=1}^{n} p_i \log(p_i) \tag{14}$$

Given a training data set of (f_i, y_i) $(i=1,2,\ldots,l)$ where $y_i \in \{1,-1\}$, N_+ is the number of positive training patterns and N_- is the negative training points. A positive example will use a target value of t_i=1- ε_+, and the negative example will use a target value of t_i= ε_-. The probability of correct label can be derived using Bayes' rule, thus, $\varepsilon_+ = 1/(N_+ + 2)$ and $\varepsilon_- = 1/(N_- + 2)$. Therefore the maximum a posteriori probability (MAP) for the target probability is expressed as:

$$t_i = \begin{cases} \dfrac{N_+ + 1}{N_+ + 2}, & y_i = 1 \\ \dfrac{1}{N_- + 2}, & y_i = -1 \end{cases} \tag{15}$$

To estimate the optimal values of parameters A and B, we can minimize the cross-entropy error function

$$\min_{z=(A,B)} -\sum_{i=1}^{l} t_i \log(p_i) + (1-t_i)\log(1-p_i) \tag{16}$$

Where

$$p_i = P(y=1 \mid f_i) = \frac{1}{1+e^{Af_i+B}}$$

The advantage of using this criterion to adjust the sigmoid parameters is that the Hessian $H(z) = \square^2 F(z)$ is positive definitive and so the problem can be solved using Newton methods without any risks of finding local minima. Parameters A and B can be solved by the following Eq.(17) iteration [23].

$$H(z) + \sigma I = -\nabla F(z) \tag{17}$$

where $\nabla F(z)$ is the gradient of $F(z)$.

$$\nabla F(z) = \begin{bmatrix} \sum_{i=1}^{l} f_i (t_i - p_i) \\ \sum_{i=1}^{l} (t_i - p_i) \end{bmatrix} \tag{18}$$

$$H(z) = \begin{bmatrix} \sum_{i=1}^{l} f_i^2 p_i (1-p_i) & \sum_{i=1}^{l} f_i p_i (1-p_i) \\ \sum_{i=1}^{l} f_i p_i (1-p_i) & \sum_{i=1}^{l} p_i (1-p_i) \end{bmatrix} \tag{19}$$

Thus, the probabilistic output of SVM is

$$P(x) = \frac{1}{1+\exp(A(\sum_{i=1}^{N_s} \alpha_i \gamma_i K(x, x_i) + b) + B)} \tag{20}$$

Probabilistic SVM for multi-class classification

As previously mentioned, SVMs have originally been developed for binary classification problems, which normally do not exist in the context of urban traffic state recognition applications. In the literature, several approaches have been introduced to solve multiclass problems. In general, the n-class problem is split into several binary problems and the individual binary classifiers are combined in a classifier ensemble. Two main approaches exist: the one-against-one (OAO) strategy and the one-against-all (OAA) strategy.

Let $H = \{h_i\}_{i=1}^{N}$ be a set of n possible class labels. The OAO strategy trains $N(N-1)/2$ individual binary SVMs, one for each possible pairwise classification problem h_i and h_j $(h_i \neq h_j)$. The sign of the distance to the hyperplane is used for the OAO voting scheme. For the final decision, the score function S_i is computed for each class h_i, which sums all positive and negative votes for the specific class. In case of the OAA approach, a set of n binary classifiers is trained to separate each class h_i from the remaining $H-h_i$. Instead of using the simple sign of the decision function, the maximum decision value (i.e., the distance to the hyperplane) determines the final class label.

Investigation [24] indicated that the OAA approach is more suitable for practical use. The multiclass extension to SVM can also be modified to include posterior probability estimates instead of hard labels. The optimal constraint condition tries to place each example on the correct side of each hyperplane with at least two-units distance. In the canonical formulation, this comes from the fact that the true class label is +1 and the wrong class label is -1. Therefore, we extend the OAA multi-class SVM to multi-class probability SVM (MPSVM). Given the training data set $\Omega = \{x_i, y_i\}_{i=1}^{l}$ where $x_i \in R^m$ and $y_i \in \{1,2,\ldots,N\}$. The objective is to correctly discriminate these classes from each other. Based on OAA approach and probability SVM [25,26], the algorithm of MPSVM is designed as follows:

Step1: Construct N binary SVM classifiers where $f_n(x)$ separates the training examples belonging to class n from the other training examples. The training set for the nth binary SVM is $\Omega = \{x_i, y_i'\}_{i=1}^{l}$ ($y_i' = 1$, if y_i=n; $y_i' = -1$ otherwise)

Step2: Train the corresponding sigmoid using the modified training set $\{f_i, t_i\}_{i=1}^{l}$, N binary PSVM classifier

with output $p_n(x)$, n=1,2,…,N according to previously mentioned Eq.(13).

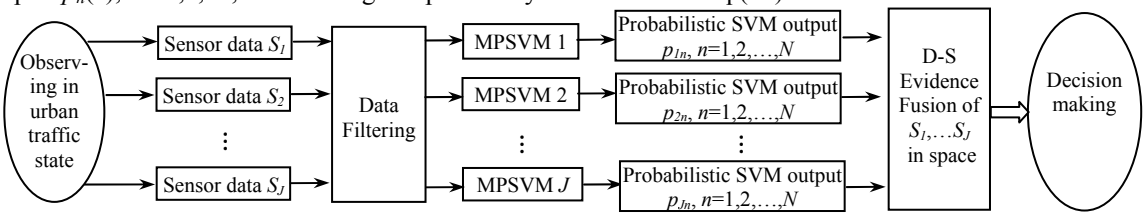

Fig.1 System structure of DS-SVM classifier

Step3: Construct the N-class MPSVM classifier by choosing the class corresponding to PSVM with the maximal value among $p_n(x)$, n=1,2,…,N. Hence, the decision function is

$$d(x) = \arg\max\{p_1(x), p_2(x), \cdots, p_N(x)\} \quad (21)$$

IV DS-MPSVM CLASSIFICATION STRUCTURE

From the above discussion, to determine the BPA of all features and classifiers in the whole frame of discernment is the key point in the application of D-S theory in traffic state recognition. In this paper, we determine the BPA from the view of various mono-sensor based on the outputs of multi-class SVM classifiers. The structure of the system is shown as Fig. 1.

The principle of the system is as follows: first, we observe the traffic incident pattern to be recognized from independent single sensors, and extract the different features of the pattern, and then we use different multi-class SVM classifiers to classify the pattern. The probabilistic outputs of the MSVM then are used to determine the BPA of each class and the uncertainty of classification in each classifier, which are transformed and then combined by Dempster's rule. Thus we may make the finale decision of the fused recognition.

Suppose there are J data sources and exist exhaustive J patterns in there data sources. J MPSVM classifiers can be constructed corresponding to the J data sources. Let the focal elements $p_{jn}(x)$ be the probabilistic output of the nth traffic flow pattern in the jth MPSVM, and the combined mass output be $m_{jn}(x)$. To normalize the probability $p_{jn}(x)$ using the following formula [26~28]

$$\overline{p}_{jn} = \frac{p_{jn}(x)}{\sum_{n=1}^{N} p_{jn}(x)} \quad (22)$$

The BPA function $m_{jn}(x)$ is defined as

$$m_{jn}(x) = \overline{p}_{jn},\ j = 1,2,...,J, n = 1,2,...,N \quad (23)$$

The combined mass function $m_1 \oplus m_2 \oplus \cdots \oplus m_J : 2^{\Theta} \to [0,1]$ is given by

$$m_n(C) = m_1(S_1) \oplus m_2(S_2) \oplus \cdots \oplus m_J(S_J) = \frac{\sum_{\cap_{i=1}^{J} S_J = C} \left(\prod_{i=1}^{J} \overline{p}_{jn} \right)}{1 - \sum_{\cap_{i=1}^{J} S_J = \Phi} \left(\prod_{i=1}^{J} \overline{p}_{jn} \right)} \quad (24)$$

The decision function based on the maximal belief rule is

$$d(x) = \arg\max_{n=1,2,...,N} \{Bel(\{1\}), Bel(\{2\}), \cdots, Bel(\{N\})\} \quad (25)$$

V. CASE STUDY

This paper has implemented the proposed traffic incident patterns recognition based on D-S theory based MPSVM by developing a program for the three-source traffic flow data fusion using the SVM toolbox in Matlab.

A. Description of traffic data source

The fusion data used in this case study are traffic flow data detected from three detectors located at Dongpu on-ramp of Guangyuan Highway in Guanghzou, China. Detecting time was from 06:00:00 to 20:59:59 on Sep. 12, 2006. The data came from three-source traffic flow data such as loop inductive detector, AVI observation and floating car data (FCD). The traffic data including speed, volume and occupancy under incident-free and incident situations was collected every 4 minutes. In the practical operation, real traffic flow data could be obtained by the following method: record the number of vehicles at a specific cross-section by a digital video, and then retrieve the data from the video. After pretreatment, all data from three detectors were organized in a format with records once every five minutes, which resulted in a total of 225 data records from each detector in 15 hours, including 45 incident state examples and 180 incident free state examples. Data from the three detectors are quite different. In order to study the performance of traffic flow under the condition of incident and incident free traffic patterns, loop detectors are placed at the site of 150 m, 300 m, 400 m on upstream side and the site of 150 m, 300 m, 400 m on downstream side of the detection area in the main lanes.

Four states are researched: smooth traffic flow, stable traffic flow, congested traffic flow, and jammed traffic flow. Among these four states, three incident types are simulated in the urban highway. The method of extracting features for traffic incident detection using wavelet transform and linear discriminant analysis is in detail described by [29]. Thus, each example in the dataset is composed of 12 condition attributes (four features from each sampling point) and one class attributes (four states). In the distributed schemes, the four features from each sampling point, adding the class attribute, form an individual dataset. Thus, three datasets corresponding to the three sampling points are constructed.

B. Evaluation indexes

The performance of the incident detection model is mainly evaluated using three indexes: detection rate

(DR), false alarm rate (FAR), mean time-to-detect (MTTD) and classification rate (CR), which are defined as

$$DR = \frac{Num\ of\ \det ected\ incident\ cases}{Total\ num.\ of\ incident\ cases} \times 100\%$$

$$FAR = \frac{Num\ of\ false\ alarm\ cases}{Total\ num.\ of\ imput\ patterns} \times 100\%$$

$$MTTD = \frac{1}{n}\sum_{i=1}^{n}(t_{\text{detected}} - t_{onset})$$

where $t_{detected}$ is the time interval between the onset of an incident and the instant when the alarm is triggered, t_{onset} is the time when the incident actually occurs, and is the number of detected incidents.

Experimental results

In our experiment, the first 165 groups of data are used as training MPSVM samples and the other data are used as testing samples. The choice of the kernel and regularizing parameters was determined by evaluation performance on a validation set. 70% of the training set is used for training binary SVM classifiers and the rest 30% of the training set is used as validation set. The whole experiment is repeated 65 times. At the same time, the same data is trained and tested with rough sets approach (RS) and the multi-layer feed forward neural network (MLF) algorithm. The comparison of the experiment results between the three algorithms is shown in Tab. 1.

Tab.1 Performance comparison SVM/RS/MLF

Scheme	DR (%)	FAR (%)	MTTD (min.)	CR (%)
SVM	99.54	5.83	1.89	85.16
RS	97.13	15.72	2.56	77.43
MLF	94.76	18.64	3.78	72.73

From Tab.1, it is obvious that SVM classifier gave high DR, it yielded low FAR to be accepted in practice. Most of the incidents could be detected accurately by SVM classifier, it means that SVM produces significantly higher DR at the same low FAR, thus, SVM classifier outperforms the other two classifiers. It has revealed that the SVM approach can effectively detect the highway incident, and could be a potential tool for AID problem.

Tab.2 Comparison of experiment results

Scheme	Mono sensor data			MPSVM	DS-MPSVM
	LOOP	AVI	FCD		
Minimal accuracy	49.83	56.53	58.67	78.42	78.42
Average accuracy	92.38	85.87	84.43	94.54	96.78
Standard deviation	0.096	0.136	0.118	0.084	0.073

The experimental results of several methods are given in Tab. 2, three types of MSVM methods, i.e., the standard MSVM, MPSVM and DS-MPSVM, are trained for each sampling point, respectively. The testing accuracy of three individual MSVM classifiers corresponding to the three sampling points is given in loop detector, SVI and FCD columns. Tab. 2 shows the classification accuracy obtained by our proposed methods (DS-MPSVM) is satisfactory. We also find the robustness of traffic incident detection is also improved by using our proposed methods.

VI. CONCLUSION

In order to minimize traffic delays, improve road capacity and safety, traffic incidents need to be detected as earlier as possible, therefore, automatic incident detection has become an important component of a modern traffic monitoring system. To deal with distributed multi-source multi-class problem, this paper investigates a new traffic incident pattern recognition approach based on Demspter-Shafer evidence theory used to combine multiple MPSVM classifiers corresponding to the data sets from different sensor sources, the final decision is based on the maximal belief principle. The experiment results confirm that DS-MPSVM is a superior pattern classifier for highway incident detection. The accuracy and robustness of fault diagnosis is improved.

REFERENCES

[1] A. Abdel, R. Kitamura, P. Jovanis, “Using stated preference data for studying the effect of advanced traffic information on drivers’ route choice,’’ *Transportation research* C, vol. *5,* pp. 39-50, 1997.

[2] Cassidy M.J.1; Anani S.B.; Haigwood J.M., “Study of freeway traffic near an off-ramp”, Transportation Research Part A: Policy and Practice, Vol.36, No.6, pp. 563-572(10), 2002

[3] Yuh-Horng Wen; Tsu-Tian Lee; Hsun-Jung Cho, Hybrid models toward traffic detector data treatment and data fusion, Proc. of IEEE Conf. on Networking, Sensing and Control, pp.525–530, 2005

[4] C. Li, R. Yang, and F. Jin, “Data-fusion prediction of traffic information based on artificial neural network,” *Systems Engineering*, vol.22, no.3, pp.80-83, 2004.

[5] Haihong Liu, Xiaoyuan Wang, Derong Tan, Lei Wang, "Study on Traffic Information Fusion Algorithm Based on Support Vector Machines," Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06), pp. 183-187, 2006

[6] H. J. Payne, E. D. Helfenbein, and H. C. Knobel, “Development and testing of incident detection algorithms,”, Federal Highway Administration Rep.FHWA-RD-76-20, Research Methodology and Results 2, 1976.

[7] S. R. Ahmed and A. R. Cook, “Application of time-series analysis techniques to freeway incident detection,” *Transportation Res. Rec. 841*, pp. 19–21, 1982.

[8] A. S. Willsky, E. Y. Chow, S. B. Gershwin, C. S. Greene, P. K. Houpt, and A. L. Kurkjian, “Dynamic model-based techniques for the detection of incidents on freeways,” *IEEE Trans. Automat. Contr.*, vol. 25, pp. 347–360, 1980.

[9] Vapnik V N, Statistical learning theory. Springer-Verlag, New York, 1998

[10] X. Zhang, “Introduction to statistical learning theory and support vector machines,” Acta Automatica Sinica, vol.26, no.1, pp.32-42, 2000

[11] Shafer, G.: A Mathematical Theory of Evidence, Princeton, NJ: Princeton University Press, (1976).

[12] Fabre, S., Appriou A., Briottet X.: Presentation and Description of Two Classification Methods using Data Fusion based on Sensor Management. Information Fusion. pp:49-712, 2001

[13] Rottensteiner F., Trinder J., Clode S., Kubik K.: Using the Dempster - Shafer Method for the Fusion of LIDAR Data and Multi - Spectral Images for Building Detection. Information Fusion, pp: 283-300, 2005

[14] Basir O., Karray F., Zhu H.: Connectionist-Based Dempster–Shafer Evidential Reasoning for Data Fusion. IEEE Transactions on Neural Networks. pp: 1513-1530, 2005

[15] Fan X., Huang H., Miao Q.: Agent-based Diagnosis of Slurry Pumps Using Information Fusion. Proceedings of the International Conference on Sensing, Computing and Automation. ICSCA 06. pp:1687-1691, 2006

[16]Xianfeng Fan, Hong-Zhong Huang, Qiang Miao, "Evidence Relationship Matrix and Its Application to D-S Evidence Theory for Information Fusion", LNCS IDEAL, pp:1367-1373, 2006

[17] Guan, J.W., Bell, D.A.: Evidence Theory and its Applications (Vol.1). North-Holland-Amsterdam, New York (1992)

[18] Yen, J.: GERTIS: A Dempster-Shafer Approach to Diagnosing Hierarchical Hypotheses.Communications of the ACM 5 32, pp: 573-585, 1989

[19] Beynona, M., Coskerb, D., Marshallb, D.: An Expert System for Multi-criteria Decision Making Using Dempster-Shafer Theory. Expert Systems with Applications 20 (2001) 357-367

[20] J. C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," *In: Advances in Large Margin Classifiers, A. J. Smola, P. Bartlett, B. Scholkopf, D. Schuur-mans, eds.*, MIT Press. , 1999,

[21] T. Hastie, R. Tibshirani, Classification by Pairwise Coupling. Technical Report, Standford University and University of Toronto, 1996.

[22]C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Mining Knowledge Discov. 2 (1998) 121–167

[23] Lin H T, Lin C J, Weng R C. A note on platt's probabilistic outputs for support vector machines. National Taiwan University, Taibei, 2003

[24] Hsu, C.-W., Lin, C.-J.: A Comparison of Methods for Multi-class Support Vector Machines. IEEE Transactions on Neural Networks 13(2) (2002) 415-425

[25]Lee, Y., Lin, Y., Wahba, G.: Multicategory Support Vector Machines: Theory and Application to the Classification of Microarray Data and Satellite Radiance Data. Journal of the American Statistical Association 99(465) (2004) 67-81

[26] Zhonghui Hu, Yunze Cai, Ye Li, Yuangui Li, and Xiaoming Xu, Data Fusion for Fault Diagnosis Using Dempster-Shafer Theory Based Multi-class SVMs, ICNC2005, pp:175-184, 2005

[27]ACEVEDO F. J. ; MALDONADO S. ; DOMINGUEZ E. ; NARVAEZ A. ; LOPEZ F. Probabilistic support vector machines for multi-class alcohol identification, Sensors and actuators. B, Chemical, Vol.122, No.1, pp:227-235, 2007

[28]WASKE Bjom, ATLI BENEDIKTSSON Jon, Fusion of Support Vector Machines for Classification of Multisensor Data, IEEE transactions on geoscience and remote sensing , Vol.45, No.12, pp: 3858-3866, 2007

[29]Samant A.and Adeli H..Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis[J].Computer-Aided Civil and Infrastructure Engineering, 2000,13(4):241～250.