Fusion of risk indicators aiming to predict near future traffic crash risks on motorways

Minh-Hai Pham, EPFL-LAVOC
André-Gilles Dumont , EPFL-LAVOC

Conference paper STRC 2011

STRC
11th Swiss Transport Research Conference
Monte Verità, Ascona, May 11 - 13, 2011

Fusion of risk indicators aiming to predict near future traffic crash risks on motorways

Minh-Hai Pham
EPFL-LAVOC
1015 Lausanne

Phone: +41 21 693 06 03
Fax: +41 21 693 63 49
email: minhhai.pham@epfl.ch

André-Gilles Dumont
EPFL-LAVOC
1015 Lausanne

Phone: +41 21 693 23 45
Fax: +41 21 693 63 49
email: andre-gilles.dumont@epfl.ch

Date

Abstract

Crashes are getting more attention in many countries. Crash prevention is vital in traffic safety especially on motorways where crashes are usually severe due to high speed. This paper proposes an approach for identifying traffic crash risk on the motorways in real-time using individual vehicle traffic data. The study site is on motorway A1 between Bern and Olten. Crashes interested in are rear-end and sideswipe.

Traffic conditions when there is no crash, called *non-crash cases*, are matched with traffic conditions leading to crashes, called *pre-crash cases*. After matching process, relevant non-crash cases are selected to compare with pre-crash cases whereas; irrelevant non-crash cases are not further considered. In reality, if a new traffic condition is not relevant to any pre-crash case; it can be declared as non-crash.

Thereafter, models are developed to distinguish pre-crash and relevant non-crash cases. Variables called *risk indicators* are generated and single-variable models are developed. Each risk indicator has certain performance in distinction of pre-crash and non-crash cases. To improve that performance, risk indicators are fused into a unique high performance model. Our results present high accuracy of more than 75% of non-crash and pre-crash cases identified.

Finally, to predict near future traffic crash risk in a real-time framework, traffic crash risk identified during the last intervals are used. The test with historical data shows that if two or three last intervals are identified as pre-crash, the chance for the coming interval to be pre-crash is high (more than 95%).

Keywords

Data Fusion – risk indicators – crash prevention – crash risk prediction

1. Background & Study Site

1.1 Background

Studies in traffic safety during the last ten years can be divided into two groups: aggregate and disaggregate based on the units of analysis according to (Golob and Recker 2003). For aggregate studies, units of analysis are crash frequencies whereas; for disaggregate studies, units of analysis are individual crashes. The present study focuses on identification of real-time crashes and provides prediction on the occurrence of individual crashes. Therefore, this study is classified into the group of disaggregate studies.

During the last decade, several disaggregate studies have been undertaken as presented in (Oh, Oh et al. 2001; Lee, Hellinga et al. 2003; Hourdakis, Garg et al. 2006; Abdel-Aty, Pande et al. 2007; Pande and Abdel-Aty 2007; Abdel-Aty, Pande et al. 2008; Hossain and Muromachi 2010; Pham, Bhaskar et al. 2011). However, in most of these studies, the performance of developed models in identifying real-time traffic crash risk is still low – lower than 70% (i.e. more than 30% crashes cannot be identified). Some models can do better with more than 70% out of crashes identified yet the cost for that higher accuracy is the higher rate of non-crash cases identified as crash cases (i.e. false alarm rate is more than 30%).

In this paper, we present results of our study aiming at identifying and predicting motorway traffic rear-end and sideswipe crashes by using risk indicators. In the second part of this section, the study site is introduced. In section 2, the methodology applied in this study is presented. Results and analyses are discussed in section 3. The conclusion is presented in section 4.

1.2 Study Site

A study site for the present study is selected on Swiss motorway A1 between two cities Bern and Zurich. The main criteria for study site selection is the simultaneous availability of individual vehicle data from double loop traffic detectors and crash records around the location of traffic detectors. The selected study site is presented in Figure 1. At the study site, there are two lanes per traffic direction.

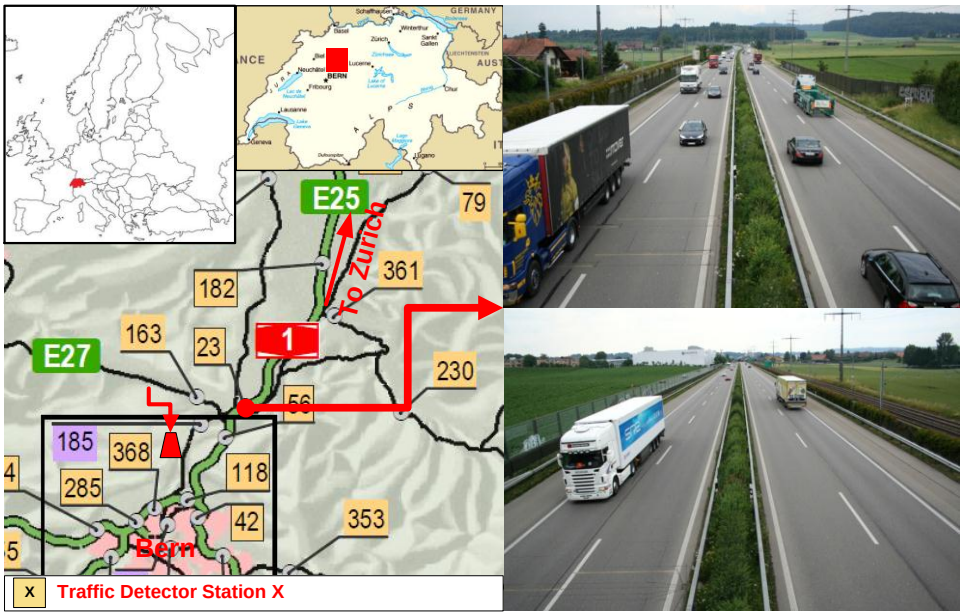

Figure 1: Study site

2. Methodology

2.1 Introduction

The methodological frame work of the current study is presented in Figure 2. At first, pre-crash and non-crash cases are defined assuring that pre-crash and non-crash cases are not overlapping. As pre-crash cases which occur before crashes are rear events, it is necessary to select relevant non-crash cases to be compared with pre-crash cases. Pre-crash cases together with matched non-crash cases make up Traffic Regimes. Under each Traffic Regime, models for identifying traffic conditions similar to pre-crashes cases are developed. To predict the near future crash risk, results from risk identification models under each Traffic Regimes are combined.

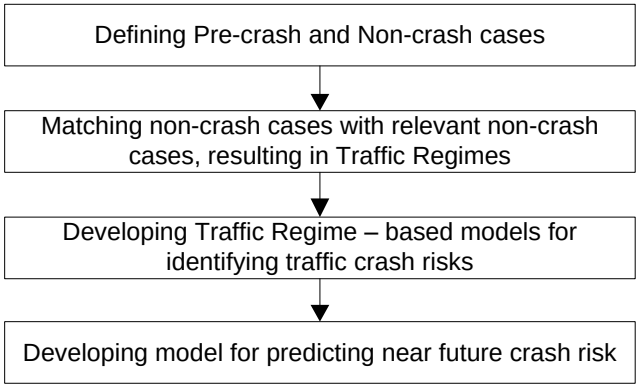

Figure 2: Methodological framework

2.2 Pre-crash and Non-crash Definition

When a crash occurs, the traffic evolution preceding the crash is of interest in the current study. Traffic evolution following the crash is not considered as the traffic could be deviated by the crash itself or intervened by the police or emergency services that come to the crash location. As illustrated in Figure 3, for a crash there is a *crash zone* including traffic evolution before and after the crash. Crash zone is divided into three parts: pre-crash buffer zone and pre-crash zone that precede the crash and post-crash zone that follows the crash. Outside of the crash zone is the non-crash zone. The post-crash zone is not considered in this study. The pre-crash zone contains traffic evolution leading to the crash. Therefore, the pre-crash zone is used as the container of pre-crash cases. The pre-crash buffer zone is the frontier between the pre-crash zone and the non-crash zone. The presence of the pre-crash buffer zone is necessary to avoid any potential overlapping between pre-crash and non-crash zones.

The non-crash zone includes traffic evolution that is not within crash zones of all recorded crashes, including crashes which are not rear-end or sideswipe.

In the current study, the duration of pre-crash buffer zone and pre-crash zone is fixed at 30 minutes. According to our analysis, for some crashes, the traffic can only be recovered to normal state at three hours after the crash. Therefore, to avoid any possible post-crash effect, the Post-crash zone is fixed at 210 minutes.

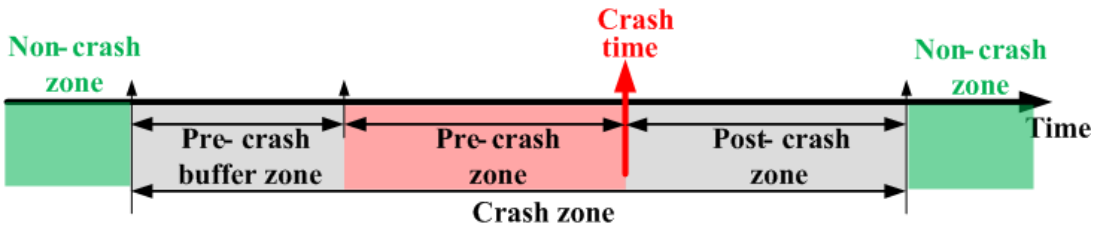

Figure 3: Definition of Pre-crash and Non-crash cases

Here, a traffic case is defined as traffic evolution during 5-minute interval. A pre-crash case is a traffic case occurring within pre-crash zones. As a pre-crash zone is a 30-minute interval, there are six pre-crash cases within a pre-crash zone. A non-crash case is a traffic case occurring within non-crash zones. In the current study, the proportions of traffic evolution that are unconsidered include all crash zones of crashes which are not rear-end or sideswipe and all pre-crash-buffer zones and post-crash zones of rear-end and sideswipe crashes.

To conform to our previous work presented in (Pham, Bhaskar et al. 2011), we call non-crash cases as *NTS* and pre-crash cases as *PTS*. Both PTS and NTS are characterized by 21 variables: Time of the day (X1); Day of the week (X2); Flow, average speed, average headway, occupancy, headway variation, speed variation and percentage of heavy vehicles on the left (from X3 to X9, respectively) and on the right (from X10 to X16, respectively) lanes; speed difference between two lanes (X17); flow change, speed change on the left (X18 and X19, respectively) and the right (X20 and X21, respectively).

2.3 Pre-crash and Non-crash Matching

As crashes are rear events, the number of pre-crash cases is much lower than the number of pre-crash cases. As stated in (Pham, Bhaskar et al. 2011), the models developed based on the imbalanced data sets might have high accuracy with training data, yet low prediction performance when used with new data. Moreover, as rear-end and sideswipe crashes only occur under certain traffic conditions, i.e. when the traffic flow is high, many non-crash cases such as midnight traffic might not be appropriate for being compared with pre-crash rush-hour traffic. Therefore, it is necessary to sample non-crash cases such that non-crash and pre-crash cases are comparable.

Figure 4 presents the process for matching NTS with PTS. The process includes clustering NTS into groups called traffic regimes – TR represented by cluster centers and classifying PTS into obtained TR such that finally, under each TR, there are a set of NTS and a set of PTS. To cluster NTS, NTS data are transformed using Principal Component Analysis (PCA) to reduce the number of dimensions facilitating the clustering step. Eigen vectors used in PCA transformation of NTS are also used to transform PTS during classification step.

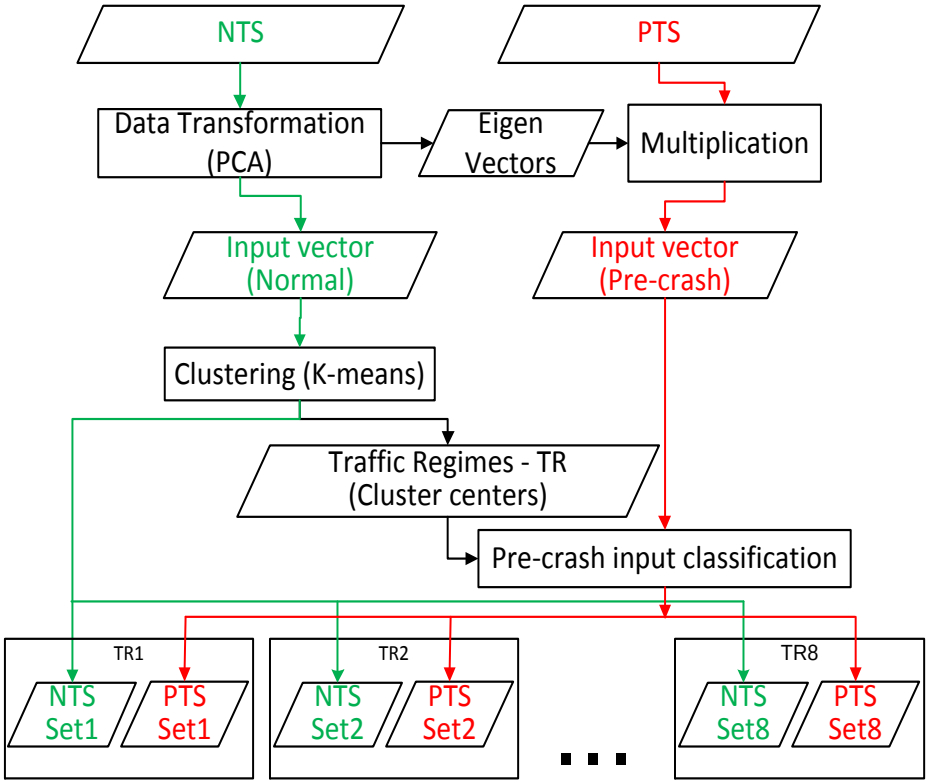

Figure 4: Matching NTS with PTS

2.4 Risk Identification Models

Two risk indicators are used in this study: the inverse of Time-To-Collision – TTC (Hayward 1971) and Platoon Barking Time Risk – PBTR (Pham, Mouzon et al. 2008). These indicators are calculated based on individual vehicle data. To aggregate the indicators for 5-minute intervals, the statistics of the indicators during 5 minutes are mean and standard deviation. The statistics of risk indicators are used as variables in the models for identifying traffic risks. Two risk indicators make up four variables. Together with 21 variables used for representing NTS and PTS, the total number of variables that can be used for developing risk identification models under each traffic regimes is 25.

Under each regime, single-variable models for 25 variables are developed to distinguish NTS and PTS belonging to that regime. After that results of single-variable models (M1 to M25)

are fused to generate a unique risk identification model, called fused Risk Identification Model – RIM under that regime as illustrated in Figure 5.

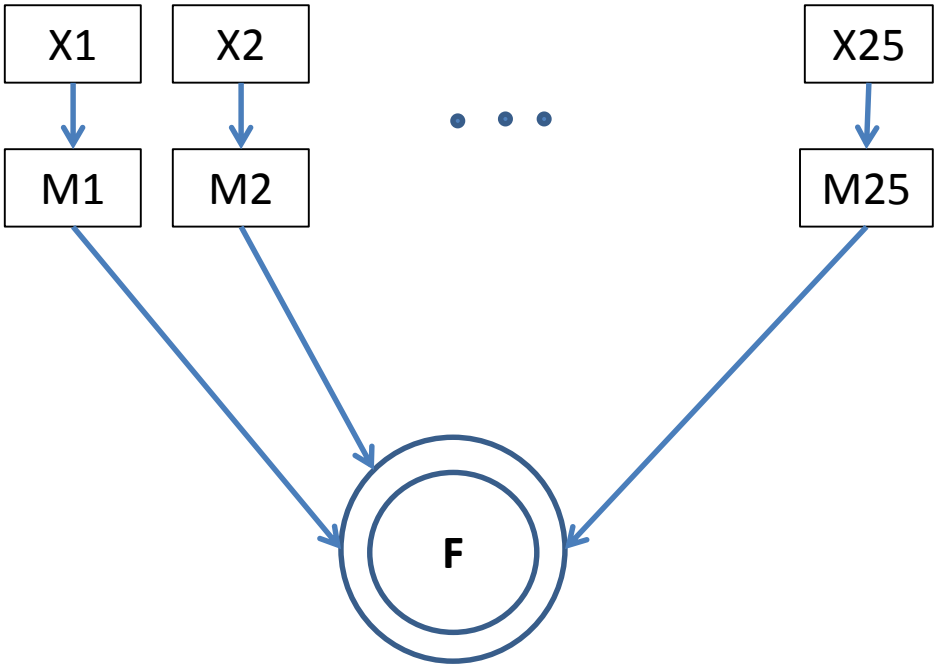

Figure 5: Fusing single-model process

The fusion process is based on a Dempster-Shaffer framework.

2.5 Risk Prediction Model

Once the traffic crash risk during the last 5-minute interval is identified in a real-time framework, near future crash risk can be predicted. Figure 6 present an example of the decision that needs to be done when the risk of the last six 5-minute intervals is identified: PTS-PTS-NTS-PTS-PTS-PTS. With that pattern, what can be predicted?

By testing the sensitivity of historical data, the number of 5-minute intervals that is necessary for predicting near future risk is decided.

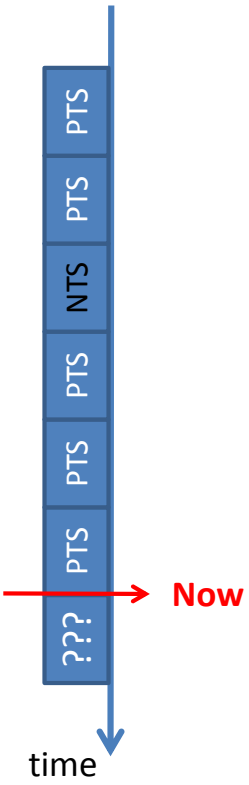

Figure 6: Traffic crash risk prediction process

3. Results and Analyses

3.1 Traffic Regimes

Based on the clustering error, the number of traffic regimes equal to 8 is chosen. Eight regimes are named from A to H. This is because selecting more traffic regimes does not reduce significantly the clustering error. Figure 7 presents the results of NTS and PTS matching process. Under each of eight traffic regimes, there is certain number of NTS. However, under regimes A and F, there is no PTS classified in. Under regimes D and E, the number of PTS is also low whereas; the number of PTS is higher under regimes B, C, G, and H.

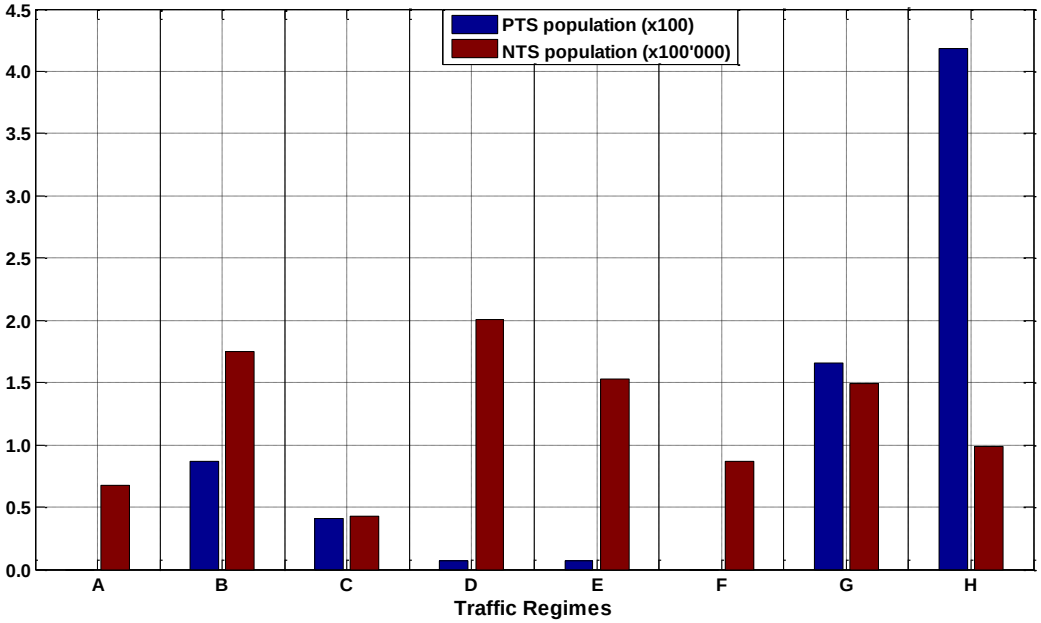

Figure 7: Distribution of NTS and PTS under eight traffic regimes

In a real-time framework, whenever a new traffic case is classified into one of regimes A, B, D, and F, that traffic case can be declared as "risk free". If the traffic case is classified into one of regimes B, C, G, and H, the fused RIM can be used to identified whether that traffic case is risky or not.

3.2 Fused Risk Identification Models

Single-variable models are developed using Random Forest (Breiman 2001) to classify a traffic case into one of two class NTS or PTS. For a traffic case, there are 25 results from 25 single-variable models. The Fused RIM result is decided by voting single-variable results.

Table 1 presents the results obtained with Fused RIM for four traffic regimes B, C, G, and H. With data including NTS and PTS divided into three smaller data sets: training and validation the performance of the models is reduced for the validation data set yet still at high level.

Table 1: Fused RIM results

Traffic Regime	Training (%)		Validation (%)	
	NTS	PTS	NTS	PTS
B	99.35	100.00	95.53	83.33
C	95.64	100.00	90.91	90.00
G	92.70	100.00	87.05	87.80
H	93.59	100.00	83.35	75.31
All four regimes	95.67	100.00	89.83	83.62

3.3 Crash Risk Prediction Model

Fused RIMs provide the affirmation of traffic crash risk for a traffic case that has occurred. In a real-time framework, the affirmation of traffic crash risk for several last traffic cases can be employed to predict the coming traffic case. Call the false alarm rate the percentage of NTS that are wrongly predicted as PTS and missed alarm rate the percentage of PTS identified as NTS. The sensitivity test with historical PTS shows that there is a tradeoff between the false alarm rate and missed alarm rate: it is impossible to minimize false alarm rate and miss alarm rate at the same time.

As illustrated in Figure 8, the number of the last traffic cases considered to predict the coming traffic case, called the length of risk memory, ranges from 1 to 5. The optimum length of risk memory can be 2 or 3 as with this length, the false alarm rate and the missed alarm rate are although not the lowest yet acceptably low compared to other length of risk memory.

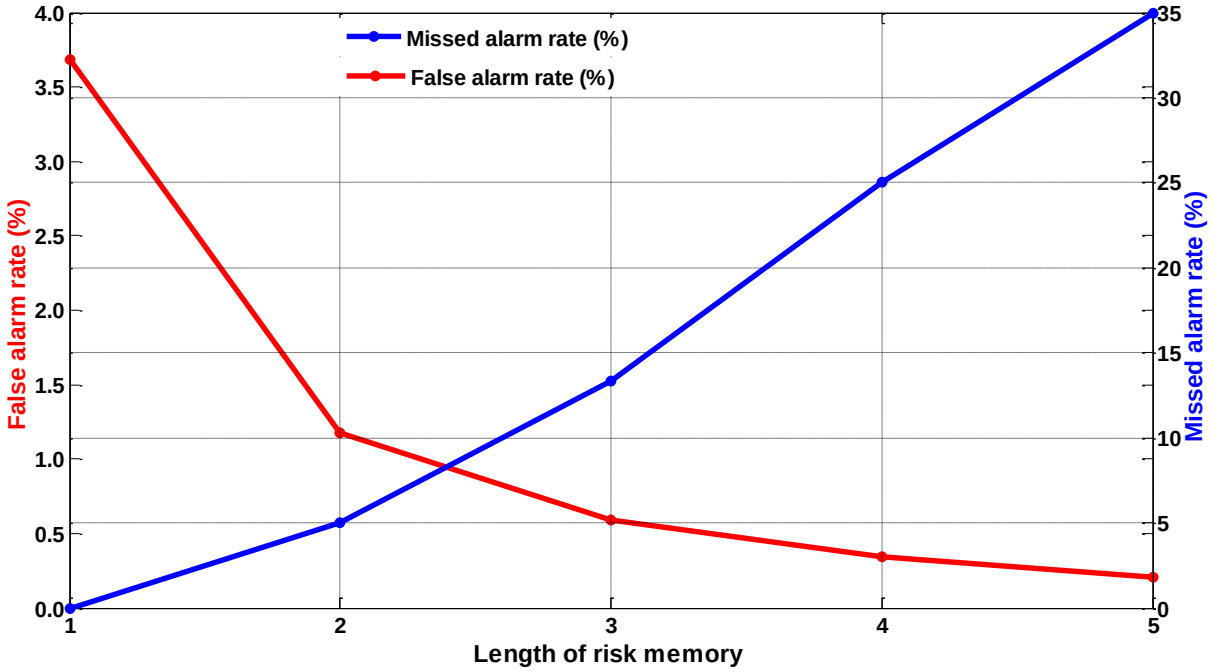

Figure 8: Tradeoff between false alarm rate and miss alarm rate

If the length of risk memory is fixed at 2, more than nearly 99 NTS are correctly predicted and nearly 95% PTS are correctly predicted.

4. Conclusions

This paper describes our on-going work on developing models for predicting near future traffic crash risk on the motorway. Data fusion technique is used to combine results of single-variable risk identification models. Among 25 variables, four variables are extracted from risk indicators. Our results are encouraging and show that nearly 95% risky traffic cases can be predicted with the cost of more than 1% of the time, the non-crash traffic is classified as pre-crash.

In reality, the false alarm rate of more than 1% is still high. Therefore, it is necessary to reduce the false alarm rate with the same or reduced missed alarm rate. One solution can be using other existing risk indicators or developing new risk indicators.

References

Abdel-Aty, M., A. Pande, et al. (2008). "Assessing Safety on Dutch Freeways with Data from Infrastructure-Based Intelligent Transportation Systems." Transportation Research Record: Journal of the Transportation Research Board **2083**(-1): 153-161.

Abdel-Aty, M., A. Pande, et al. (2007). "Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety on Freeways." Journal of Intelligent Transportation Systems: Technology, Planning, and Operations **11**(3): 107 - 120.

Breiman, L. (2001). "Random Forests." Machine Learning **45**(1): 5-32.

Golob, T. F. and W. W. Recker (2003). Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. Reston, VA, United States, American Society of Civil Engineers.

Hayward, J. C. (1971). Near misses as ameasure of safety at urban intersections.

Hossain, M. and Y. Muromachi (2010). Evaluating Location of Placement and Spacing of Detectors for Real-Time Crash Prediction on Urban Expressways. 89th TRB Annual meeting, Washington DC.

Hourdakis, J., V. Garg, et al. (2006). "Real-Time Detection of Crash-Prone Conditions at Freeway High-Crash Locations." Transportation Research Record: Journal of the Transportation Research Board **1968**(-1): 83-91.

Lee, C., B. Hellinga, et al. (2003). Real-time crash prediction model for application to crash prevention in freeway traffic. Washington, DC, United States, National Research Council.

Oh, C., J.-S. Oh, et al. (2001). Real-time Estimation of Freeway Accident Likelihood. 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.

Pande, A. and M. Abdel-Aty (2007). "Multiple-Model Framework for Assessment of Real-Time Crash Risk." Transportation Research Record: Journal of the Transportation Research Board **2019**(-1): 99-107.

Pham, M.-H., A. Bhaskar, et al. (2011). Methodology for Developing Real-time Motorway Traffic Risk Identification Models Using Individual Vehicle Data. 90th Transportation Research Board annual meeting, Washington DC.

Pham, M.-H., O. d. Mouzon, et al. (2008). Sensitivity of risk indicators under motorway traffic regimes clustered by self-organizing map. 7th European Congress on ITS, Geneva, Switzerland.