International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

# Smart Predictor: A Road Incident Prediction System using Hybrid Data Mining Approach on Traffic Data

**Omkar Haldankar[1], P. S. Desai[2]**

[1]Department of Computer Engineering, SKNCOE, Pune, Savitribai Phule Pune University, Pune, India

[2]Professor, Department of Computer Engineering, SKNCOE, Pune, Savitribai Phule Pune University, Pune, India

**Abstract:** ***Tackling urban road congestion by means of ITS technologies, involves a number of key challenges. One such challenge is related to the accurate detection of traffic incidents in urban networks for more efficient traffic management. It combines K-means Clustering algorithm and Naive bayes techniques in light of the fault diagnosis theory. In the core of the proposed approach lies a more efficient feature extraction technique, based on the dynamic characteristics of data corresponding to those vehicles that are involved in incidents. In the core of the proposed approach lies a more efficient feature extraction technique, based on the dynamic characteristics of data corresponding to those vehicles that are involved in incidents. Our work observe show dynamic aspects of measured data can be exploited for extracting features that result in measurable improvement of the incident detection rate by the application of a Naive bayes.***

**Keywords**: Naïve-bayes, K-Means clustering, Automatic Incident Detection (AID), Machine Learning, Time Series Analysis

## 1. Introduction

TRAFFIC incidents are a critical problem in modern transportation systems. They cause congestion on the road and reduce the efficiency of transportation systems. According to, incidents cause 12 to 33 percent of congestion in major developed countries. Thus, automatic incident detection (AID) is a crucial component of intelligent transportation systems (ITSs) and has attracted the interest of many researchers.

Moreover, it has an economical and environmental impact, as it is directly associated with fuel consumption and air pollution. It is proven that a significant rate of congestions is produced due to incidents. By incidents, we refer to unusual events that take place in roads and can interrupt the smooth traffic flow. Examples of such incidents include car accidents, interruptions due to extreme weather conditions or high concentration of people who demonstrate, etc. The financial cost and the discomfort of thousands drivers caused by incidents, are so big that impose the need for more effective detection and immediate treatment.

In AID research, incidents usually refer to any nonrecurring events that may disrupt normal traffic flow and reduce the traffic capacity of a road. Effective and efficient detection of incidents can expedite response and intervention by administrative agencies and reduce the possible loss caused by incidents in terms of life, Money and Time.

## 2. Motivation

Tackling urban traffic has a great societal impact, as it affects the quality of transportation in cities for citizens. Moreover, it has an economical and environmental impact, as it is directly associated with fuel consumption and air pollution. It is proven that a significant rate of congestions is produced due to incidents. By incidents, we refer to unusual events that take place in roads and can interrupt the smooth traffic flow. Examples of such incidents include car accidents, interruptions due to extreme weather conditions or high concentration of people who demonstrate, etc. The financial cost and the discomfort of thousands drivers caused by incidents, are so big that impose the need for more effective detection and immediate treatment.

## 3. Literature Survey

**(I) Application Context and Data Source**

AID is mainly applied in two traffic contexts: freeways and urban roads. In general, freeways have uninterrupted traffic flow and relatively homogenous traffic patterns. Urban roads are usually characterized by irregular traffic flow, which is caused by turns and traffic controls. Data needed for AID are usually collected from three sources: static sensors, dynamic sensors, and traffic cameras. (Here, we separate traffic cameras from the other two types of sensors because of their unique features.) Static sensors mainly include loop detectors, which can provide numerical traffic information, such as volume, occupancy, and speed. Dynamic sensors are usually installed on probe vehicles to provide a continuous stream of measurements of the streets that they pass by. Traffic cameras provide video clips of the roads. Videos from traffic cameras can be either converted into numerical variables or directly used for detecting incidents with the help of computer vision techniques.

(ii) Comparative Algorithms: Based on the comparison between the occupancy values that occur in two consecutive roads. The first one is the road in question (i.e., the one that is checked for incident occurrence) and the second one is the next road in the same flow direction. If the value of the occupancy of the next road is much lower than the occupancy of the current road, then it is assumed that an incident occurs.

(iii) Statistical algorithms: Calculate statistical differences from the occurring data and the predicted or estimated data.

(iv) Time series algorithms: Time series algorithms consider that the traffic data can be represented by a standard pattern over time. In order to calculate this pattern, time series models are applied. Thus, after calculating the regular data model for any considered time interval ahead, simple comparison with the real data evolving pattern may reveal an incident by detecting any deviation from the regular model.

(v) Machine Learning: AID algorithms in the ML category usually consider incident detection as a binary classification problem, where each instance is the traffic on a road segment at a particular time. Features describing the instance are generated and fed into an ML classifier to decide whether an incident happened or not.

(vi)Car-Following Data Collection Methods: Car-Following studies typically collect vehicle trajectory data through various means, including naturalistic, simulator, and video data collection methods.

(vii) Car-Following Models:
Car-following models are designed to process various stimuli, such as the headway distance between vehicles, and produce a response or action, such as the driver decelerating to maintain a certain following distance. The main categories of car following are action-point, linear models, non-linear models.

## 4. Design Techniques

### (1) K-Means Clustering

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation maximization mechanism allows clusters to have different shapes. Given a set of observations (x1, x2, …, xn), where each observation is a d-dimensional real vector, kmeans clustering aims to partition the n observations into k (≤ n) sets S = {S1, S2, …, Sk} so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find:

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2$$

where μi is the mean of points in Si.

### (2) Naïve bayes Algorithm

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness and diameter features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood in other words; one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests. An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters necessary for classification.

## 5. Conclusion

A By using hybrid approach, we can get most accurate solutions for traffic incident detections. We have proposed a hybrid approach that combines K-means clustering and Naïve-Bayes to address the incident detection problem in ITS. We presented an AID system based on a SVM implementation that shows improved performance after selecting appropriate features. The purpose of our research was to select those SVM features and parameters that provide the most accurate solutions for traffic incident detection. After experimenting with several combinations of features, we confirmed that adding the average deviation of the speed of each road that is examined for detecting incidents, as a new feature, decreases the value of MTTD, hence the incidents are detected faster.

## 6. Acknowledgement

I wish to thank Prof P. S. Desai for giving me proper guidance and suggestions regarding this work.

## References

[1] J.Wang, X.Li, S.Liao, and Z.Hua, A Hybrid Approach for Automatic Incident detection, IEEE Transactions on Intelligent Transportation Systems, vol.14, 2013.

[2] Evangelos Gakis, Dionysios kehagias and Dimitrios Tzovaras, Mining Traffic Data for Road Incidents Detection, IEEE 17th International Conference on Intelligent Transportation Systems(ITSC), 2014.

[3] B. Chidlovskii and E. Cardenas Sanchez, Traffic Delay Detection by mining Ticket Validation Transactions, IEEE 17th International Conference on Intelligent Transportation Systems(ITSC), 2014.

[4] X.Wen, L.Shao, Senior Member, IEEE, Wei Fang and Yu Xue, Efficient Feature Selection and Classification for vehicle detection, IEEE Transaction on Circuits and System for Video Technology.

[5] Bryan Higgs and Montasir Abbas, Member, IEEE, Segmentation and Clustering of Car-Following Behaviour: Recognition of Driving Patterns, IEEE Transactions on Intelligent Transportation Systems, Vol.16, NO1, february 2015.

[6] M.Wu, S.Lam, Member, IEEE, and Thambipillai Srikanthan, Senior Member, IEEE, Nonparametric Technique Based High-Speed Road Surface Detection.

[7] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models, " *J. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 53–72, May 2009.

[8] W. Wang, S. Chen, and G. Qu, "Incident detection algorithm based on partial least squares regression, " *Transp. Res. C, Emerg. Technol.*, vol. 16, no. 1, pp. 54–70, Feb. 2008.

[9] H. M. Wen, Z. S. Yang, G. Y. Jiang, and C. F. Shao, "A new algorithm of incident detection on freeways, " in *Proc. IEEE Int. Conf. Veh. Electron.*, 2001, pp. 197–202.

[10] R. Bryant, R. H. Katz, and E. D. Lazowska, "Big- data computing: Creating revolutionary breakthroughs in commerce, science and society, " in *Computing Initiatives for the 21st Century*. Washington, DC, USA: Comput. Res. Assoc., Dec. 2008.

[11] S. Chen, W.Wang, and H. VanZuylen, "Construct support vector machine ensemble to detect traffic incident, " *Expert Syst. Appl.*, vol. 36, no. 8, pp. 10 976–10 986, Oct. 2009.

[12] F. Yuan and R. L. Cheu, "Incident detection using support vector machines, "*Transp. Res. C, Emerg. Technol.*, vol. 11, no. 3/4, pp. 309–328, Jun.–Aug. 2003.