Volume 8, No. 5, May – June 2017

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Modified SVD-PCA Technique for Traffic Incident Detection

Shagufta Afreen
Dept. of C.S.E
Jamia Hamdard
New Delhi, India

Suraiya Parween
Dept. of C.S.E
Jamia hamdard
New Delhi, India

Abstract: In Real time traffic incident detection is acute for increasing safety and mobility on freeways. There has been incident detection approaches built on traffic behavior or mathematical models projected for this task. Though, earlier incident detection methods are partial in unique recurrent and non-recurrent congestions. The difficulty of current methods makes them insufficient to handle the real time task. In this paper, a novel approach for detecting incidents is proposed. The algorithm properties of singular value decomposition (SVD) and principle component analysis (PCA) matrix whose elements are local energies of coefficients at different scales. It uses the diagonal, left and right singular matrices obtained in SVD to determine the number of scales of self-similarity for road density, no. of vehicle and speed limit of location and scales of anomaly in data, respectively. Our simulation work on UK based authentic data sets validates that the technique achieves better detection rate that can be able to find if results nearly to zero for true and false accident rate for our hybrid approach NB based SVD technique better than the existing approach for self-similar data used C#.net based application.

Keywords: Traffic incident, SVD, PCA, NB classifier etc

I. INTRODUCTION

Initial detection of traffic incidents can minimize the delay knowledgeable by drivers, wasted fuel, emissions, and lost productivity, while also decreasing the possibility of secondary collisions [1]. Traffic incident recognition is thus a critical issue and it is significant to develop machineries to detect traffic incidents as initial as probable. The incident detection problem has established great interest from investigators and numerous incident detection methods have been established. As in [2] they have widely reviewed numerous approaches to incident detection. Current incident detection approaches fall into the subsequent foremost groups: pattern recognition, time series analysis, Kalman filters, partial least squares regression [3, 4], and data mining machineries, that contain NN, fuzzy logic, support vector machines [5, 6], rough set [7], ensemble learning [8, 9], and decision tree learning [10]. Data mining tools has been presented for most capable methods of the incident detection approaches. The characteristic classifiers in data mining, such as decision tree inductive schemes and neural networks, are calculated to enhance complete accuracy without accounting for the comparative circulation of every class. Such as an outcome, these classifiers incline to neglect minor classes while directing on classifying the huge classes precisely [11]. Unfortunately, these classifiers achieve poorly in incident detection since the real-world traffic information suffers from class differences that classically cover much fewer incident cases than incident-free cases. Such circumstances pose experiments for these classifiers. Numerous explanations to the class imbalance difficult have been formerly offered both at the data and algorithmic levels. At the information level, resampling approaches are usually used to address the class imbalance difficult [12, 13]. Though such methods can be very modest to appliance, tuning resampling approaches to be actual use is not an easy task. At the algorithmic level, cost-sensitive learning [14], one-class classifiers [15], and ensemble-based classifiers, like as Boosting and Adaboost [16, 17], are very famous methods for solving dataset imbalance difficulties.

II. RELATED RESEARCH WORK

Nowadays, incident detection has involved much consideration in freeway systems control to decrease traffic delay, improvement road safety, capability and real time traffic control as once freeway and major incidents happen, it cause congestion and flexibility loss, if it will be not fixed on time, they can cause second traffic accidents. The pattern recognition methods contain the processes which help to sense incidents, time series (Angshuman, 2004), filtering, fuzzy set(Edmond Chin-Ping Chang and Kunhuang Huarng, 1993), and ANN (Edmond Chin-Ping Chang, 1992) (Wang et al., 2007) Dissimilar sensors, vehicles can be prepared with sensing devices with high processing power, high cost and weight such as GPS, chemical spill detectors, video cameras, vibration sensors, acoustic detectors, etc. subsequently they are not typically controlled through energy and size constrictions. VANETs deployment scenario is dissimilar than traditional WSN deployment setups since vehicles expose imperfect mobility design reason behind to street shapes, connections, speed control, vehicle size. If vehicles are prepared with network cards, then the wireless access the networks like DSRC/WAVE, Cellular, Wifi, WiMAX, etc. (Lee and Gerla, 2010)

When they apply data mining methods to classic traffic accident data records can help to recognize the features of drivers' behavior, roadway ailment and weather ailment that were causally associated with altered injury severity. This can help decision creators to communicate better traffic security control policies. Roh et al. [22] demonstrated how statistical approaches based on focused graphs, created over data for the current period, may be valuable in modeling traffic fatalities by associating models quantified using focused graphs to a classical, based on out-of-sample forecasts, initially established by Peltzman [23]. The absorbed graphs model outperformed Peltzman's model in root mean squared prediction error. Ossenbruggen et al. [24] used a logistic regression model to classify statistically important factors that forecast the likelihoods of crashes and damage crashes pointing at using these models to achieve a risk assessment of a given region. These models were functions of factors that define a site by its land use activity, roadside strategy, use of traffic control strategies and traffic exposure. Their study demonstrated that village sites are less hazardous than residential and shopping sites. Abdalla et al. [25] deliberate the association among casualty frequencies and the distance of the accidents from the regions of residence. As power has been expected, the subject occurrences were higher earlier to the regions of residence, most probably due to higher coverage. The work exposed the casualty rates among residents from areas ordered as moderately deprived were meaningfully higher than those from comparatively rich areas. In [26] author gives the statistical properties of four regression models: two traditional linear regression models and two Poisson regression simulations in positions of their capability to model vehicle accidents and highway geometric strategy relations. Roadway and truck accident data from the Highway Safety Information System (HSIS) have been working to demonstrate the use and the limits of these models. It was established that the predictable linear regression models lack the distributional property to define sufficiently random, discrete, nonnegative, and classically infrequent vehicle accident events on the road. The Poisson regression models, on the other hand, possess most of the necessary statistical possessions in emerging the relationships.

III. SYSTEM MODEL

In our proposed system mode have refine data processing and evaluate better accuracy with similar diagonal matrix and it's all status attribute:

a. Find frequencies for all road density with same speed limit.
b. Calculate counter for next decomposition that have to use our proposed approach

c. Then apply Singular value decomposition of a matrix $M \in C^{mxn}$ is given by for all fix state after pre-processing of dataset and calculate U,V matrix , if all speed rate get same for U,V matrix then M non-negative matrix get consider

$$M=U\Sigma V^H \qquad (1)$$

Where $U \in C^{mxm}$ and $V \in C^{mxn}$ are unitary matrices and $\Sigma \in R^{mxn}$ *is a real non-negative*

d. Now non negative matrix compute , $M^H = V\,\Sigma^T U^H$, now put U,V for condition $m \geq n$ without loss of generality. The singular values may be arranged in any order, for if $P \in R^{mxm}$ and $Q \in R^{nxn}$ are permutation matrices such that $P\Sigma Q$ remains "diagonal", then

$$M=(UP^T)(P\Sigma Q)(Q^T V^H) \qquad (2)$$

is also an SVD. It is customary to choose P and Q arranged in non-increasing order:

$$\sigma_1 \geq \ldots \geq \sigma_r > 0,\ \sigma_{r+1} = \_\ \_\ \_ = 0, \qquad (3)$$

Where r = rank (M).

e. If the matrices for all road density are U, Σ and V for velocity are partitioned by columns as

$$U = [u_1, u_2, \_\ \_\ \_, u_m],\ \Sigma = \text{diag}\,[\sigma_1, \sigma_2, \_\ \_\ \_, \sigma_n] \text{ and } V = [v_1, v_2, \_\ \_\ \_, v_n], \qquad (4)$$

f. The σ_i is the i^{th} singular value of M, and u_i and v_i are the left and right singular vectors for road density only corresponding to σ_i. If M is real, then the unitary matrices U and V are real and hence orthogonal.
g. A memory is initialized with $M = \{I_M, A, I_N\}$ where I_L stands for a L × L identity matrix. During the bidiagonalization phase, Givens rotations are successively applied to A from the left-hand side (LHS) and from the right-hand side (RHS), such that the M × N dimensional inner matrix A gets bidiagonal and real-valued (denoted by B_0)
h. Diagonalization: The diagonalization phase consists of multiple diagonalization steps (indicated with k) and ensure convergence of the diagonalization phase and to reduce the overall computation time of the SVD, each diagonalization step of first Given rotation is performed with a modified input vector $[x\ y]^T$, where $y = t_{12}$ and $x = t_{11} - \mu$ uses the Wilkinson shift [7].

$$\mu = a_n + c - \text{sign}\,(c)\sqrt{a_{n+}^2 b_{n-1}^2} \qquad (5)$$

With $c = \frac{1}{2}(a_{n-1} - a_n)$, TB_K^H, B_k, and the trailing non-zero sub-matrix of T corresponds to

$$T\,(n-1{:}\,n,\ n-1{:}\,n) = \begin{pmatrix} a_{n-1} & b_{n-1} \\ b_{n-1}^* & a_n \end{pmatrix} \qquad (6)$$

$$A = \begin{bmatrix} C & C & C \\ C & C & C \\ C & C & C \end{bmatrix} \xrightarrow{LHS} \begin{bmatrix} R & C & C \\ 0 & C & C \\ 0 & C & C \end{bmatrix} \xrightarrow{RHS} \begin{bmatrix} R & R & 0 \\ 0 & C & C \\ 0 & C & C \end{bmatrix} \xrightarrow{LHS} \begin{bmatrix} R & R & 0 \\ 0 & R & C \\ 0 & 0 & C \end{bmatrix} \xrightarrow{RHS} \begin{bmatrix} R & R & 0 \\ 0 & R & R \\ 0 & 0 & R \end{bmatrix} \xrightarrow{LHS} \begin{bmatrix} R & R & 0 \\ 0 & R & R \\ 0 & 0 & R \end{bmatrix} = B_0$$

(7)

Diagonalization

$$B_k = \begin{bmatrix} d_1 & f_1 & 0 \\ 0 & d_2 & f_2 \\ 0 & 0 & d_3 \end{bmatrix} = \begin{bmatrix} x & y & 0 \\ 0 & R & R \\ 0 & 0 & R \end{bmatrix} \xrightarrow{RHS} \begin{bmatrix} R & R & 0 \\ R & R & R \\ 0 & 0 & R \end{bmatrix} \xrightarrow{LHS} \begin{bmatrix} R & R & R \\ 0 & R & R \\ 0 & 0 & R \end{bmatrix} \xrightarrow{RHS} \underset{B_{k+1}}{\begin{bmatrix} R & R & 0 \\ 0 & R & R \\ 0 & R & R \end{bmatrix}} \xrightarrow{LHS} \begin{bmatrix} R & R & 0 \\ 0 & R & R \\ 0 & 0 & R \end{bmatrix} =$$

(8)Fig. Bidiagonalization and Diagonalization phases' illustration for SVD for a complex valued 3x3 matrixes

[7]. The transformation (or mapping) of a dataset onto a new set of principal axes or components. These axes are ordered by the amount of variation in all attributes that they capture in the data. Namely, the maximum amount of variation that is possible to represent on a single axis is captured by the first principal axis . Each of the remaining.

i. The number of traffic detectors located on a subset of the total set of arterial links of an urban network, t be the number of successive class of dataset in which the detector data are collected, and τ be the number of time intervals wherein each day is partitioned.
j. Now singular value corresponding to the i-th principal axis. The magnitude of singular values demonstrates the overall variation attributable to each particular principal component and, hence, the potential to reconstruct total traffic data using a smaller number of dimensions.
k. The probabilities of classes in given a text data by the help of combined probabilities of disputes and groups. It is built on the notion of word individuality. The preliminary opinion is the Bayes formula for provisional possibility, asserting which, for an assumed information opinion x and class C:
l. In particular, we need to find out

- how often each word appears in the true rate
- how often each word appears in the false rate
- how many true terms available
- how many false terms available
- total number of attribute in dataset.

The result of classification, we will use the Maximum Likelihood approach which means that you calculate the probability of the text being in each class and output the class corresponding to the highest probability. More specifically, you will calculate:
And output positive if Pr(*False rate* | *f*) was max or false rate if Pr(*True rate of accident* | *f*).

m. An incident has occurred on the road or not then a strap needs to detect whether a peak with a certain value has occurred in any row/lane of the Road. One way to identify such a peak may be described as follows:

1. Compute the average (μ) and standard deviation (σ) for Count Values for each row in the Road incident data. i.e. for each lane.
2. Find the minimum Count, Countmin
3. Use the idea of band pass filter to take away regular oscillation and fluctuation from the values. Actually, this is very important step as we are concerned in large negative peak at some positions given that other positions have normal counts.
4. If $\mu - \sigma -$ Countmin $>$ K then alarm will be raised for an incident, where K is a detection threshold that determines how conservative the detection should be. The larger the detection threshold, the more conservative the detection is, the more detection time is needed and less false alarms is generated.

n. Combined 5 Rules: for our proposed approach get used probability combination rules showing [12] is product rule, the sum rule, the min rule, max rule, and PCANBTree Classifier Rule have been proposed .

.

IV. RESULT AND DISCUSSION

Using the estimated traffic model, we evaluated the performance of the proposed method. Our detection method gives an alert when the divergence of an input set of observation values for the estimated traffic model exceeds a given threshold. We regarded a set that includes at least one value labeled as an incident to be a truly inconsistent set. The two proposed algorithms require parameters N or T. The results shown in Section

Figure 1: Load the train data

Figure 1 indicated that the traffic patterns are different for each route, section, and time, and therefore the detection threshold must be changed accordingly. Although the granularity of such a threshold tuning should depend on the amount of data, we know of no method to determine the appropriate granularity.

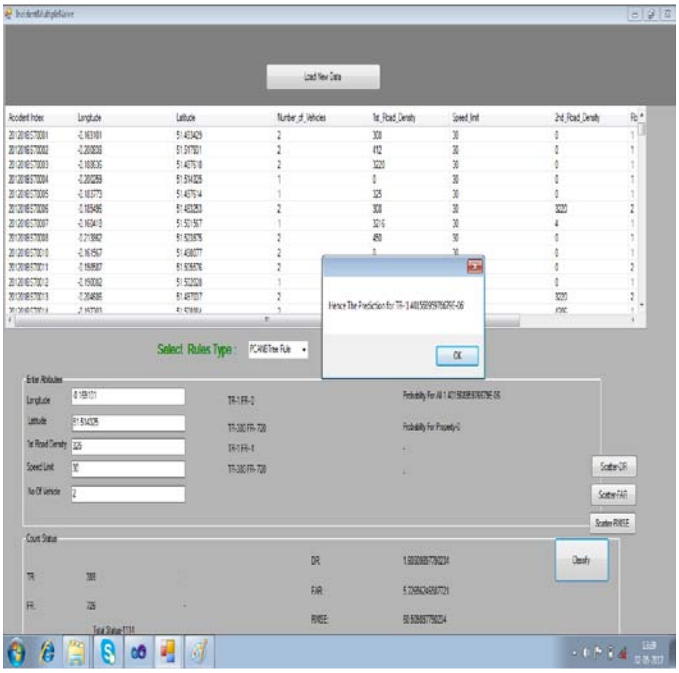


Figure 2: Rule selection and our proposed Rule

In this figure, we select the longitude, latitude, road density, speed limit and no. of vehicles.
After this we select the rule as PCANB Tree Rule and then get DR value=1.60, FAR= 5.72 and RMSE= 60.50. So we conclude this rule is able to find more RMSE and FAR and give best accuracy for detection rate.

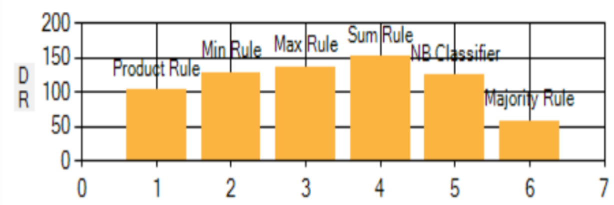


Figure 3: Performance improvement with Hybrid-NB for Traffic Incident detection

Above figure shows the Hybrid-NB for traffic incident detection. This technique gives the best performance in sum rule for DR amount of instances that were incorrectly classified as incident instances based on the total instances in the testing set. Out of the total number of applications of the model to the dataset, DR is calculated to determine how many incident detection rate are same.

In general, detection rates (DR), a false alarm rate (FAR) and root mean square error (RMSE) indexes to evaluate an incident detection algorithm. In this paper, we use C# dotnet framework software to make a computer simulation for Improved NB-tree algorithm.

For false alarm rate derived formula =(No of incident false alarm /Total no of non- incident case)x100%

Below figure is able to find more FAR rate using NB-classifier. This figure suggests a traffic incident detection algorithm based on longitude and latitude. And hybrid classifier rate 154.2 as compare to Sum rule of our existing author study i.e 150.5 The simulation result shows that the algorithm has high DR and low FAR.

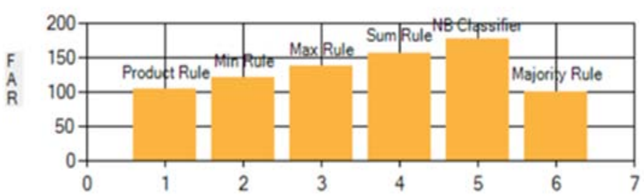


Figure 4: Performance improvement in FAR using Hybrid classifier

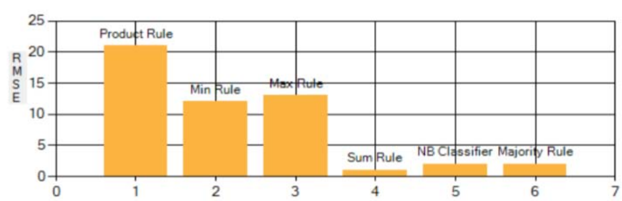


Figure 5: Performance improvement in RMSE using Hybrid classifier

Based on the observed and predicted results, the root mean squared error (RMSE) and MAPE were calculated to investigate the accuracy of the predictions. RMSE is defined as follows:

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Oi-Pi)^2}$$

where O_i is the observed duration for i^{th} traffic incident; P_i is the predicted duration for ith traffic incident; and n is the number of traffic incident records to be predicted.

V. LIMITATIONS

For this work in real area traffic data from a traffic circle, alterations allowing to the limitations in the mining of data from video files were completed. These can be categorized into two main collections:

• Primary limits; recognized through the conditions of the quality-checking process labeled before (long enough routes in time and space, when speeds under 200 km/h then single pass concluded the connection and a insignificant percentage of location points positioned inside the road limitations). It means automatic traffic incident still suffer from the quality checking process of vehicle.

• Secondary limits; In current time data connected to suitable substances. In most applicable cases contain routes which are quicker to every other than in authenticity and incorrect approximations of the size and headline angles of vehicles. Most traffic incidents occur when the vehicle close to each other on roadway.

VI .CONCLUSION

This paper presents the possibilities of achieving maximally effective and efficient development and implementation of the real-time traffic incident problems. It emphasizes the special significance on the timely incident detection. The efficient management of available information, data exchange as well as intelligent real-time decision-making can reduce the counter value and traffic incidents, which prevent secondary incident The system performance improve significantly by the approach based on the intelligent transport system model. The main performance characteristics taken into consideration are response time and reduction of harmful consequences from incidents. These all work done using NB-tree, PCA and SVD which achieve better improvement.

The future work should study the possibilities of different realizations of the algorithm for traffic flow variables estimation. Therefore, a promising approach is based on the implementation of neuro-fuzzy estimators, which may also include the incident detection algorithms.

REFERENCES

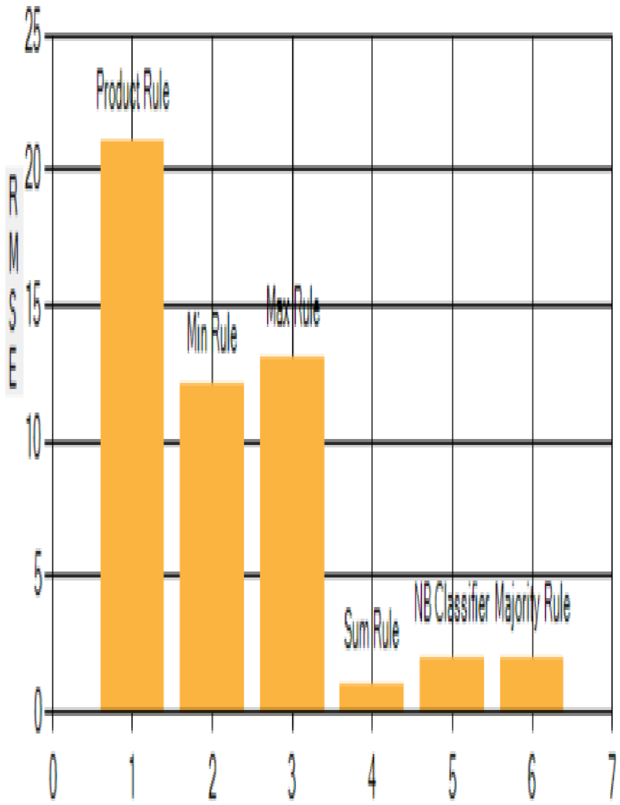
[1] B. Hellinga and G. Knapp, “AVI based freeway incident detection,” in Proceedings of the TRB Annual Meeting, pp. 1–26, 2000

[2] J. Black and I. Sreedevi, “Detection Algorithms,” http:// www.calccit.org/itsdecision/serv and tech/Incident management/incident management overview.html

[3] W. Wang, S. Chen, and G. Qu, “Incident detection algorithm based on partial least squares regression,” Transportation Research Part C, vol. 16, no. 1, pp. 54–70, 2008

[4] W. Wang, S. Chen, and G. Qu, “Comparison between partial least squares regression and support vector machine for freeway incident detection,” in Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems, ITSC 2007, pp. 190–195, Seattle, Wash, USA, October 2007.

[5] F. Yuan and R. L. Cheu, “Incident detection using support vector machines,” Transportation Research Part C, vol. 11, no. 3-4, pp.



[16] R. E. Schapire, “A brief introduction to boosting,” in Proceedings of the 16th International Joint Conference on Artificial Intelligence, pp. 1401–1406, 1999.

[17] Y. Freund and R. Schapire, “Experiments with a new boosting algorithm,” in Proceedings of the 13th International Conference on Machine Learning, M. Kaufmann, Ed., pp. 148–156, SanFrancisco, Calif, USA, 1996.

[18] Angshuman, G., 2004. An Incident Detection Algorithm Based On a Discrete State Propagation Model of Traffic Flow

[19] Edmond Chin-Ping Chang, Kunhuang Huarng, 1993. Fuzzy Set Applications For Freeway Incident Detection. IEEE, pp. 439-443

[20] Edmond Chin-Ping Chang, 1992. A Neural Network Approach to Freeway Incident Detection. IEEE, pp. 641-647.

[21] Wang, W., Chen, S., Qu, G., 2007. Comparison between Partial Least Squares Regression and Support Vector Machine for Freeway Incident Detection. IEEE, pp. 190-195.

[22] Roh J.W., Bessler D.A. and Gilbert R.F., Traffic fatalities, Peltzman’s model, and directed graphs,
Accident Analysis & Prevention, Volume 31, Issues 1-2, pp. 55-61, 1998.

[23] Peltzman, S., The effects of automobile safety regulation. Journal of Political Economy 83, pp. 677–725, 1975.

[24] Ossenbruggen, P.J., Pendharkar, J. and Ivan, J., Roadway safety in rural and small urbanized areas. Accid. Anal. Prev. 33 4, pp. 485–498, 2001.

[25] Abdalla, I.M., Robert, R., Derek, B. and McGuicagan, D.R.D., An investigation into the relationships between area social characteristics and road accident casualties. Accid. Anal. Prev. 29 5, pp. 583–593, 1997

[26] Miaou, S.P. and Harry, L., Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 6, pp. 689–709, 1993.

[27] Lee, U., Gerla, M., 2010. A survey of urban vehicular sensing platforms. Computer Networks
54, 527–544.

