AN OVERVIEW OF DATA MINING IN ROAD TRAFFIC AND ACCIDENT ANALYSIS

K. Jayasudha
Lecturer/Research Scholar,
Department of MCA,
K.S.R College of Engineering,
Tiruchengode – 637 215, Tamilnadu, India.
Email: jayasudhakaliannan@yahoo.com

Dr. C. Chandrasekar
Reader, Periyar University, Salem,
Tamilnadu, India
Email: ccsekar@gmail.com

ABSTRACT

Data mining is a promising area for dealing with the increased, stored data that has been generated in our times .It is the extraction of implicit, previously unknown and useful data. In this paper we have analyzed some of the data mining techniques, tools, applications and search engines for accident investigation and traffic analysis. Most of the accident investigation methodologies are based on scenarios of the accident occurrence and simulation of accident situation. The costs of fatalities and injuries due to traffic accident have a great impact on society. Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. In recent years, researchers have been utilizing real-life data in studying various aspects of traffic accidents. So measures have to be taken to reduce accidents. It is important that the measures should be based on scientific and objective surveys of the causes of accidents and severity of injuries. Our study highlights various tools, techniques and applications of data mining in accident analysis will eliminate deficiencies of other techniques but covers their advantages. Our main aim is to overcome the death rate and the increased rate of loss of lifes by means of using some tools, techniques or various algorithms in the field of data mining using the traffic data bases.

Keywords - mining, accidents, road safety

1. INTRODUCTION

Road safety experts and researchers deal with large volumes of quantitative information and collected statistics, in order to understand and estimate the social and economic cost of the accidents and to be able to introduce safety plans in order to prevent or reduce occurrences of accidents. The road traffic and accident statistics must be presented in such a way to make it easier to be both recognized and interpreted by a human operator. Previous works on accident analysis included statistical methods and formal techniques. Statistics tables and ordinary charting techniques are not sufficient for present day requirements and this causes difficulties in the effective visualization of results and patterns. Another disadvantage is that ordinary methods limit human involvement in the exploration tasks. In this paper we have analyzed various studies done by various authors in accident databases and also measures proposed to prevent the accident and its death rate.

Functionalities:

Data mining is the process of identifying valid and understandable patterns in the data set. It helps in extracting and refining useful knowledge from large data sets. Data mining tasks can be classified in to two categories *descriptive* and *predictive*. *Descriptive mining* tasks characterize the general properties of the data in the database. *Predictive mining* tasks perform inference on the current data in order to make predictions. Main functionalities of data mining are as follows: *Data characterization* which summarizes the target class(class under the study).*Data discrimination* which compares the target class with one or more sets. *Association analysis* which discovers association rules showing attribute-value conditions that occur frequently together in a given set of data.(using support and confidence).*Classification* is the process of finding a set of models that describes and distinguish data classes or concepts for the purpose of being able to use

the model to predict the class of objects whose class label is unknown. *Prediction* it is nothing but in many applications users may wish to predict some missing or unavailable data values (numerical value) is often specifically referred to as prediction. *Cluster analysis* which analyzes data objects without consulting a known class label. The objects are clustered or grouped based on the principle of maximizing the intra class similarity and minimizing the interclass similarity. *Outlier analysis:* A database may contain data objects that do not comply with the general behavior or model of the data. These data objects are called outliers. *Evolution analysis* which describes and models regularities or trends for objects whose behavior changes over time.

2. RELATED RESEARCH

Kim[1] developed a log-linear model to clarify the role of driver characteristics and behaviors in the causal sequence leading to more severe injuries. They found that driver behaviors of alcohol or drug use and lack of seat belt use greatly increase the odds of more severe crashes and injuries. Shankar [2] applied a nested logic formulation for estimating accident severity likelihood conditioned on the occurrence of an accident. The study found that there is a greater probability of evident injury or disabling injury/fatality relative to no evident injury if at least one driver did not use a restraint system at the time of the accident. Dia used real-world data for developing a multi-layered NN freeway incident detection model [3]. They compared the performance of the neural network model and the incident detection model in operation on freeways. Yang [4] used NN approach to detect safer driving patterns that have less chances of causing death and injury when a car crash occurs. Evanco [5] conducted a multivariate population-based statistical analysis to determine the relationship between fatalities and accident notification times . The analysis demonstrated that accident notification time is an important determinant of the number of fatalities for accidents on rural roadways. Mussone [6] used neural networks to analyze vehicle accident that occurred at intersections. Results showed that the highest accident index for running over of pedestrian occurs at non-signalized intersections at nighttime. The next major approach focuses on the area of Advanced Driving Assistance Systems (ADAS) [7]. These systems concentrate on attempting to prevent specific damaging scenarios, such as vehicles rear-end collision and lane deviation. They are mostly used in Smart Cars, and they work by mining data obtained from various sensors in the car. Different data mining techniques are used in an attempt to predict a driver's moves, so that unsafe actions can be rectified, or prevented. The use of data mining to improve road safety can be categorised into two major approaches. The first approach concentrates on mining crash data, which includes various attributes relating to both driver and vehicle at the time of the crash [8,9,10]. The focus is on analyzing the data for the purpose of discovering useful, and potentially actionable, information. In [11] crash data was mined to identify the driver and vehicle attributes which are the main causes for road accidents. Principal Component Analysis was used to emphasize the relationships between characteristics such as age, gender and vehicle type, to the crash variables. Bedard . [12] applied a multivariate logistic regression to determine the independent contribution of driver, crash, and vehicle characteristics to drivers' fatality risk. It was found that increasing seatbelt use, reducing speed, and reducing the number and severity of driver-side impacts might prevent fatalities. Ossiander [13] used Poisson regression to analyze the association between the fatal crash rate (fatal crashes per vehicle mile traveled) and the speed limit increase and found that the speed limit increase was associated with a higher fatal crash rate and more deaths on freeways . There are many research papers about spatial data mining [14] technologies of traffic data, for instance optimization of travel plan, visualization of traffic jam, GPS assisted navigation, road design, and ICT-assisted traffic congestion. The use of data mining to improve

road safety can be categorized into many major approaches and one such approach concentrates on mining crash data, which includes various attributes relating to both driver and vehicle at the time of the crash [15] The focus is on analyzing the data for the purpose of discovering useful, and potentially actionable, information. Abdel-Aty [16] used the Fatality Analysis Reporting System (FARS) crash databases covering the period of 1975-2000 to analyze the effect of the increasing number of light truck vehicle (LTV) registrations on fatal angle collision trends . They investigated the number of annual fatalities that result from angle collisions as well as collision configuration (car-car, car-LTV, LTV-car, and LTV-LTV). Sohn et al.[17] applied data fusion, ensemble and clustering to improve the accuracy of individual classifiers for two categories of severity (bodily injury and property damage) of road traffic accident.

3.ROAD SAFETY DATABASES

In developing an accident database at the national level, efforts were made to use the technologies available that are able to assist in developing and implementing road safety programs and enable planning at both the local and national levels. The study of these databases showed that a tremendous effort had been made by different international and regional organizations. The main objectives of these databases were to explain road accidents in compatible and homogenous formats and to reduce the effort and time spent by researchers and end users in collecting accident statistics. International and regional road safety organizations use these databases to publish annual reports and statistics, according to predefined user requirements with agreed variables and indicators. Research in this area is devoted to developing database schema that represent the road accident data and other relevant statistics (population censuses, socioeconomic and energy related data) that is useful for users and decision-makers. Technically, all available road databases collect and process data in different ways . Most of the accident variables used in different databases are selected to give full descriptions of the crash type and persons involved . For the purpose of the research, the data set used comes from international road databases that have good quality information and have been in existence for a significant length of time.

IRTAD (http://www.irtad.net/)

The International Road Traffic and Accident Database (IRTAD) is an international database that gathers data on traffic and road accidents (accident and victims as well as exposure data are collected on a continuous basis) from the Organization for Economic Co-operation and Development (OECD) member countries. The main part of the database, with around 500 data items, includes aggregated data on injury accidents, road fatalities, vehicle population, and network length.

GLOBESAFE (http://www.globesafe.org/)

GLOBESAFE is a database and a platform used to share road safety information among road safety organizations. This database offered data for the nine ASEAN countries from 1994 onwards.

ACCIDENT RESOURCE
(http://www.accident-resource.com/)

The ARC Network is the number one web site for traffic accident reconstruction on the Internet. It hosts an expert directory of over 500 expert witnesses and contains thousands of pages relating to all areas of accident reconstruction. It provides valuable information regarding all aspects of all traffic accidents, including motorcycle accidents, pedestrian accidents, auto accidents, commercial vehicle accidents, bicycle accidents and more.

CRASHDATARETRIEVAL*(BlackBox)*
(www.cdr-system.com)

This web site is dedicated to the crash data retrieval system. This is the software system kit that allows people to download information from a vehicles "black box". This web site helps to

understand what is going on with the new technology. The CDR System is a tool that provides access to vehicle crash data that may be stored in late model passenger cars, light trucks. Crash Data Retrieval System is the essential tool used for

- better case evaluation
- Collection of additional evidence
- Settlement decisions
- Subrogation
-

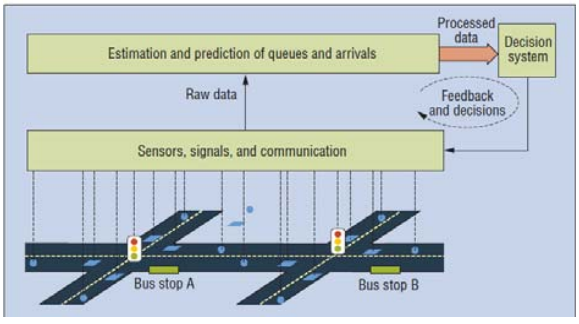
http://www.indiandrivingschools.com/accidents-on-indian-roads.html This web site clearly explains and gives details regarding the road safety ,traffic rules, what are all the conditions to be followed for driving ,details of the traffic authorities, in various types of road ranges.

4. TOOLS USED

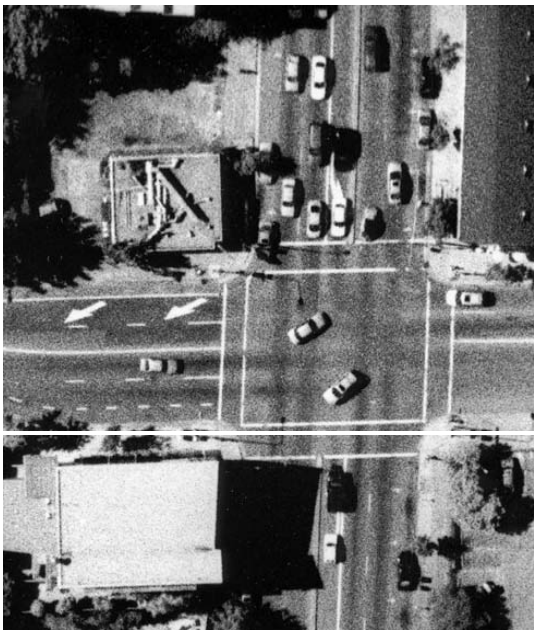
- Orange
- Tanagra
- Clementine(SPSS PASW Modeler)
- Rapid Miner
- SAS
- SAS Enterprise Miner
- Darwin
- Data cruncher
- Enterprise miner
- Gain smarts
- Intelligent miner
- Mine set
- Model Quest
- Salford CART
- Olpars
- Beyesia
- Scenarion
- S-plus

5. PROPOSED WORK

With the rapid increasing popularity of the WWW, Websites are playing a crucial role to convey knowledge and information to the end users. Discovering hidden and meaningful information about Web users usage patterns is critical to determine effective marketing strategies to optimize the Web server usage for accommodating future growth. Most of the currently available Web server analysis tools provide only explicitly and statistical information without real useful knowledge for Web managers. The task of mining useful information becomes more challenging when the Web traffic volume is enormous and keeps on growing. In this paper, we propose a alert model to discover and analyze useful knowledge from the available Web data. We made use of the classification information generated by a self organizing map for pattern analysis to capture the chaotic trend to provide short-term (hourly) and long-term (daily) Web traffic trend predictions like fatal accidents, seriously injured people, slightly injured people, material damage. Our traffic networks equipped with a number of measurement devices of various kinds(inductive loops, video sensors, radar detectors) that deliver real-time information about the current traffic conditions in corresponding locations. Our proposed model will gives alert whenever the vehicle crosses the danger zone either in audio or video format. Also we aimed to extend it by sending alerts from one vehicle to another. We can make use of the sensors to gather the details with the help of the GPS technology.it was collected in the data base and gathered in it.then data was classified using the decision support systems and identify the exact decision to be given to the driver as the alert.it is one among the useful way to prevent accident rates by giving alerts to the driver. A proposed sample diagram is shown in the following figure



It is identified that accident rates are more at the intersections than in the highways or main roads. Sample traffic flow at the intersections was shown below. It was the aerial view of the photograph taken.



6. CONCLUSION

The aim of this study was to show the applications of data mining techniques in the feild of accident investigation. It was done by reviewing various papers. Association rules are proposed to use to discover the patterns and rules that cause the occurrence of accidents. Our method is at early stages of development. We are currently enhancing it by considering several issues, variation in crash occurrence may have some consequence for traffic safety measures in some counties. It is beyond the scope of this study to determine which traffic crash excesses are due to poor roadway conditions or the necessity of stop signs or traffic lights.. The determination of specific precautionary measures to reduce crashes requires study of other factors such as the identification of specific road sections that need work, etc.. A systematic method for updating the index year after year could be considered. Additionally, further explorations of traffic safety data using data mining techniques are warranted.

REFERENCES

[1]. Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention*, Vol. 27, No. 4, 1995, pp. 469-481.

[2]. Shankar, V., Mannering, F., & Barfield, W., Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol. 28, No. 3, 1996, pp.391-401.

[3]. Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 313-331.

[4]. Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns by A Neural Network Approach. ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining, Vol. 9, pp. 839-844, Nov. 1999.

[5]. Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. Accident Analysis and Prevention, Vol. 31, 1999, pp. 455-462.

[6]. Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31, 1999, pp. 705-718.

[7]. Oliver, N. and Pentland, A. P., Graphical Models for Driver Behavior Recognition in a SmartCar, MIT IEEE intelligent vehicles symposium, 2000.

[8]. Tavris, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. Accident Analysis and Prevention, Vol. 33, 2001, pp. 167-172.

[9]. Singh, S., 'Identification of Driver and Vehicle Characteristics through Data Mining the Highway Crash', National Highway Traffic Safety Administration, USA, 2001.

[10]. Singh, S., 'A Sampling Strategy for Rear-End Pre- Crash Data Collection', National Highway Traffic Safety Administration, USA, 2001.

[11]. Singh, S., 'A Sampling Strategy for Rear-End Pre- Crash Data Collection', National Highway Traffic Safety Administration, USA, 2001.

[12]. Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident analysis and Prevention*, Vol. 34, 2002, pp. 717-727.

[13]. Ossiander, E. M., & Cummings, P., Freeway speed limits and Traffic Fatalities in

Washington State. Accident Analysis and Prevention, Vol. 34, 2002, pp. 13-18.

[14]. S. Shekhar and S. Chawla, "Spatial Databases: A Tour," Prentice Hall, 2002.

[15]. Flach, P.A., Mladenic, D., Moyle, S., Raeymaekers, S., Rauch, J., Rawles, S., Ribeiro, R., Sclep, G., Struyf, J., Todorovski, L., Torgo, L., Blockeel, H., Wettschereck, D., Wu, S., Gartner, T., Grobelnik, M., Kavsek, B., Kejkula, M., Krzywania, D., Lavrac, N. and Ljubic, P., 'On the road to knowledge: mining 21 years of UK traffic accident reports'., Data Mining and Decision Support: Aspects of Integration and Collaboration, pages 143-155. Kluwer Academic Publishers,January 2003

[16]. Abdel-Aty, M., & Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. *Accident Analysis and Prevention*, 2003.

[17]. Sohn, S. Y., & Lee, S. H., Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 4, issue1, February 2003, pp. 1-14.

[18]. Ching-Yao Chan, David Marco,Traffic Monitoring at Signal-Controlled Intersections and Data Mining for Safety Applications, 2004.

Biography

Ms.K.Jayasudha received the B.Sc.degree in Computer Technology in 2001 and M.C.A Degree in 2005 and the M.E. degree in Computer science and Engineering in 2007 from K.S.Rangasamy college of Technology. She has three years of teaching experience. She is currently working as Lecturer at K.S.R college of Engineering .She has registered for PhD in Anna University, Coimbatore, India, under the supervision of Dr.C.Chandrasekar. Her area of interest includes data mining, networks etc.,

Dr.C.Chandrasekar obtained his Ph.D in the year of 2005 . He is in teaching profession for the past 18 years in reputed engineering colleges and universities.. He is guiding more than 30 Ph.D Scholars . He is presently working as Reader at Periyar University Salem. He has presented many papers in National Conferences International Conferences and published many papers in International Journals. His area if interest includes data mining, Image processing, Mobile Ad Hoc Networks, Network Security etc.,