Data Mining Approaches to Analyze Road Traffic Accident Data

Dr. T. Anuradha [#1]

[#] *Associate Professor, Dept. of Computer Science, Dravidian University, Kuppam, A.P, India -517426,*

Abstract— **In recent years, the number of traffic accidents, especially severe vehicle crashes has been increasing because of rapid growth of volume of road traffic and rapid increase in speed of road traffic. The costs of fatalities and injuries due to road traffic accidents have raised great concern across the globe. Road traffic accident data are huge, heterogeneous and multidimensional. Data mining techniques have been applied to analyze road traffic data to identify various factors that influence injury severity, and to device road safety measures and strategies that reduce number and severity of road traffic accidents. This paper presents different data mining approaches to mine road traffic accident data.**

Index Terms— **data mining, road traffic accident data mining, road traffic accident severity**

I. INTRODUCTION

It is estimated by world Health Organization that 1.2 million deaths and 50 million injuries are caused by road traffic accidents each year. The cost of these deaths and injuries has a tremendous impact on the social and economical development of a location or country. Thus the methods that improve road traffic safety and reduce road traffic accidents and accident severity are of greater importance to traffic agencies and the public. Many data mining-related studies have been carried out to analyze road traffic accident data and to identify the significant factors that can help improve the overall driving safety, prevent accidents and also reduce injury severity.

Data Mining: Data mining is a major step of knowledge discovery in databases (KDD). KDD is the process of extraction of non-trivial, previously unknown, valid and potentially useful information from huge databases. Some of important data mining techniques are classification, clustering, association rule mining, outlier detection, segmentation, sequential pattern mining etc. This paper presents a few data-mining related studies that have been carried to perform analysis on road traffic accident data.

II. LITERATURE REVIEW:

Many studies have been carried out to identify the factors leading to severe road traffic accidents, and to reduce the number and the severity of the accidents by eliminating or controlling these factors. As the traffic accident data is huge, heterogeneous and high dimensional many researchers have followed data mining approaches to carry their studies.

Authors in [1] have applied artificial neural networks and decision trees to an actual data set containing traffic accident records from 1995 to 2000 from the National Automotive Sampling System (NASS) General Estimates System (GES). The data set was narrowed to head-on collisions only. The accident result variable 'injury severity' is given five label-values: no injury, possible injury, non-incapacitation injury, incapacitating injury and fatal injury. Neural networks and decision trees are applied to investigate their performances. The results showed that the decision tree approach outperformed neural networks. The experiments showed that the most important factors in fatal injuries are: driver's seat belt usage, light condition of the roadway, and driver's alcohol usage.

Miao et al. [2] investigated the application of machine learning approaches: neural networks, decision trees, hybrid decision tree neural network and support vector machines on the data from the National Automotives Sampling System (NASS) General Estimates System (GES). The data set contained data pertaining to driver only data related to other passengers' was not included. The inputs were like drivers' age, gender, alcohol usage, vehicle body type, vehicle age, vehicle role, initial point of impact, collision manner, travel speed etc. and output was injury severity and it has five values: no injury, possible injury, non-incapacitation injury, incapacitating injury and fatal injury. The above machine leaning approaches were applied to classify the severity of injuries more accurately. The results revealed that, for non-incapacitating injury, incapacitating injury and fatal injury classes, the hybrid approach performed better than neural network, decision trees and support vector machines; for no injury and possibly injury classes, the hybrid approach performed better than neural network. Decision trees could best model the no injury and possible injury classes.

Bedard et al. in their work [3] applied multivariate logistic regression to determine the independent contribution of driver, crash, and vehicle characteristics to drivers' fatality. The results showed that increased use of seatbelt, reduced vehicle speed, and reduced number and severity of driver side impacts might prevent drivers' fatalities.

Sohn et al. [4] considered two categories of severity of road traffic accident data: bodily injury and property damage. Data fusion, ensemble and clustering were applied to improve the

accuracy of neural networks and decision trees. Data set was divided into subsets by applying clustering algorithms and the each subset of the data was used to train neural networks and decision trees. The results showed that classification based on clustering works better, particularly if the data set has relatively large variation as in the case of Korean road traffic accident data set.

Feed-forward Multi-layer-perceptron with back propagation learning was used by Mussone et al. [5] to analyze vehicle accidents that occurred at road intersections in Milan, Italy. The neural network model had 10 input nodes and one output node called accident index node. The accident index was calculated as the ratio between the number of accidents for a given intersection and the number of accidents at the most dangerous intersection. Experiments showed that the highest accident index for running over of pedestrians occurs at intersections without signal at night time.

In [6] to evaluate the interaction of gender, age, type of crash, and occupant role in motor vehicle crash injuries leading to hospitalization, we analyzed 1997 Wisconsin hospital discharge data for patients with primary E-code diagnoses of motor vehicle injuries. The overall ratio of males to females (M/F ratio) hospitalized for motor vehicle crash injuries was 1.33 (95% confidence interval (CI): 1.26-1.41). The M/F ratio varied by type of crash and differed for passengers and drivers. For injuries sustained in collisions between vehicles, the M/F ratio was 0.96 (95% CI: 0.87-1.05); in loss of control accidents the M/F ratio was 1.95 (95% CI: 1.76-2.17). Within each type of crash, the M/F ratio for drivers was similar to that for the entire type; the M/F ratio for passengers was about half of the type total. Expressed as rates of hospitalization per 100,000 people in the general population, hospitalizations of drivers in collisions with another motor vehicle increased steeply in males, but not in females, beginning at about age 70. For drivers in loss of control crashes, male rates exceeded female rates in all age groups, with peaks in the groups 15-24 and 85-89. For passengers, injury rates from collisions with other motor vehicles were greater for females, especially in the elderly, and injury rates from loss of control crashes were similar for both genders, with peaks at 15-24 and 85-94. The higher fatality of men in loss of control motor vehicle crashes, compared to women, suggests an important area for further investigation.

In [7], a new mathematical model was developed to estimate average injury and fatality rates in frontal car-to-car crashes for changes in vehicle fleet mass, impact speed distribution, and inherent vehicle protection. The estimates were calculated from injury fatality risk data, delta-V distribution and collision probability of two vehicles, where delta V-depends on impact speed and mass of the colliding vehicles. The impact speed distribution was assumed to be unaffected by a change in fleet mass distribution. The results showed that safety in frontal crashes would improve 27-35% by a 10% increase in fatality risk parameters, which reflected substantial improvement in inherent vehicle protection. A 40% safety improvement was attained by a 10% impact speed reduction. Consequences of vehicle fleet mass were not as strong, but depended on the average mass ratio of the fleet. A reduction in mass range would be the most beneficial, while a

uniform mass reduction of 20% would increase the fatality rate by 5.4%. The model estimates trends in traffic safety and may help to identify priorities in active and passive safety.

The purpose of paper [8] is to develop a statistical model explaining the relationships between certain driver characteristics and behaviors, crash severity, and injury severity. Applying techniques of categorical data analysis to comprehensive data on crashes in Hawaii during 1990, we build a structural model relating driver characteristics and behaviors to type of crash and injury severity. The structural model helps to clarify the role of driver characteristics and behaviors in the causal sequence leading to more severe injuries. From the model we estimate the effects of various factors in terms of odds multipliers—that is, how much does each factor increase or decrease the odds of more severe crash types and injuries. We 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. Driver errors are found to have a small effect, while personal characteristics of age and sex are generally insignificant. We conclude with a discussion of our modeling approach and of the implications of our findings for appropriate traffic safety interventions and future research.

Injuries trends of passenger car drivers in head-on collisions are identified based on crash data extracted from the National Automotive Sampling System [9]. Annual injury incidence levels are estimated for years 1990–2007. Over that period, the number of crashes is predicted to rise by 71%. However, the number of serious injuries to drivers is expected to rise by only 41% and driver fatalities are anticipated to decrease by 9%. Meantime, the types of injuries suffered by drivers are changing. Year-to-year shifts in injury patterns result from changes in vehicle size classes within the US vehicle fleet population and increases in seat belt use and air bag availability. The effectiveness of air bags in saving lives is estimated to be 30%, and with more air bag-equipped cars on the road, the probability of sustaining a life-threatening head or a torso injury is reduced. Air bags, however, are not as effective in preventing upper and lower extremity injuries, and thus arm and leg injuries will become more prevalent in years to come.

Hanrahan R B, Layde P M et al. [10] used Wisconsin Crash data from 2002 to 2004 to study the association of driver age with traffic injure severity. The data set contained 602,964 records pertaining to drivers' involved in a motor vehicle crash. Logistic regression model was applied to this data set and the results showed that the drivers older than 85 years had the highest risk for severe injury and fatality.

In [11] to identify the important factors contributing to road accidents in Britain, a large data set containing 9297 different records with 52 different attributes is analyzed. From this data set they have evaluated first simply the percentage of fatalities in the data. They also interrogated the percentage of accidents involving one or more children whose age is greater than 0 and less than 16. Analysis is also performed to see what gender age group was involved in the most accidents, whether there were any trends in the types of vehicles that were involved in the accidents. Finally classification was performed on data to see a common occurrence on today's roads, multiple car accidents.

Logistic regression models were employed to develop crash-related injury prediction models. Singleton, Qin et al [12], analyzed traffic crash data in Kentuchy 2000-2001 using logistic regression and concluded that concluded that the occupant's risk factors for high level of injury severity were age, female gender, and restraint non-use, ejection from the vehicle and driver imparity.

In [13], the authors used logistic regression model to quantify the association of driver's age with traffic injury severity. Winconsin crash data from 2000 to 2004 was used to study 602,964 drivers encountered mot, or vehicle crash, and discovered that the oldest drivers, especially those older than 85 years had high risk of severe injury and fatality

Artificial neural networks were applied to analyze road traffic accident data. In [14], to model potentially non-linear relationships between the injury severity levels and crash-related factors, authors used a series of neural networks. In [15], artificial neural networks are employed to model the relationship between driver injury severity and the number of crash factors.

III. CONCLUSION

The costs of deaths and injuries due to road traffic accidents have a tremendous impact on the socioeconomic development of a country. It has been acknowledged as a global phenomena and traffic safety has gained a serious concern globally. Many researchers have analyzed road traffic accident data to identify key factors of road traffic accident severity. Many data mining approaches have been used to analyze traffic accident data. This paper presents some of the works done in mining traffic accident data.

REFERENCES

[1] Miao M Chong, Ajith Abraham, Marcin Paprzycki, "Traffic Accident Analysis Using Decision Trees and Neural Networks", IADIS International Conference on Applied Computing, Portugal, IADIS Press, Pedro Isaias et al. (Eds.), ISBN: 9729894736, Vol. 2, pp. 39-42, 2004.

[2] Miao M Chong, Ajith Abraham, Marcin Paprzycki, "Traffic Accident Data Mining Using Machine Learning Paradigms" Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, ISBN 9637154302, pp. 415-420, 2004.

[3] Bedard M, Guyatt G H, Stones M J and Hireds J P, "The Independent Contributions of Driver, Crash, and Vehicle Characteristics to Driver Fatalities", Accident Analysis and Prevention, Vol. 34, pp 717-727, 2002.

[4] Sohn S Y and 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, issue 1, Feb 2003, pp 1-14.

[5] Mussone L, Ferrari A and Oneta M, "An analysis of urban collision using an artificial intelligence model. " Accident analysis and pretension, vol. 31 1999, pp 705-718

[6] Tavris D R, Kuhn E M and Layde P M, "Age and Gender Pattern in Motor Vehicle Crash injuries: Importance of Type of Crash and Occupant Role", Accident Analysis and Prevention, Vol 33, 2001, pp 167-172

[7] Buzeman D G, Viano D C & Lovsund P, "Car Occupant Safety in Frontal Crashes: A Parameter Study of Vehicle Mass, Impact Speed, and Inherent Vehicle Protection", Accident Analysis and Prevention, Vol. 30, No. 6, pp 713-722, 1998

[8] 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.

[9] Martin P G, Crandall J R & Pilkey W D, " Injury Trends of Passenger Car Drivers in the USA", Accident Analysis and Prevention, Vol.32, 2000, pp 541-557

[10] Hanrahan R B, Layde P M, "The Association of Driver Age with Traffic Injury Severity in Wisconsin", Traffic Injury Prevention, Vol. 10, No. 4, 2009, pp 361-367

[11] Mohamad Saraee, Jonathan Kerry, Michelle Lloyd, Christine Markey, "Data Mining Applications: Case of Road Traffic Accidents in the UK West Midlands Area 2000"

[12] Singleton M., Qin H., Luan J., "Factors Associated with Higher Levels of Injury Severity in Occupants of Motor Vehicles that were Severely Damaged in Traffic Crashes in Kentucky, 2000-2001", Traffic Injury Prevention, Vol. 5, N0. 2, pp 144-150

[13] Hanrahan R B., Layde P M., et al., "The Association of Driver Age with Traffic Injury Severity in Wisconsin", Traffic Injury Prevention, Vol 10, No. 4, pp 361-367, 2009

[14] Delen D., Sharda R., Bessonov M.,"Identifying Significant predictors of injury severity in traffic accidents using a series of artificial neural networks", Accident Analysis and Prevention, Vol. 38, No. 3, pp434-444, 2006

[15] Abdelwahab H T., Abdel Aty M A., "Development of Artificial Neutral Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections", Transportation Research Records, Vol. 1746, No. 1, pp. 6-13, 2001