Factors affecting road crash modeling

Malaya Mohanty*; Ankit Gupta
National Institute of Technology Hamirpur, Hamirpur, India

Abstract

Road accident fatalities have been on an increasing trend for the last decade or so in India. Hence traffic safety management has emerged as a topic of discussion for researchers all over the world. Hence accident modelling on different factors causing them has to be conducted. Accident modelling helps us to know the real causative agents behind an accident to occur. The effect of one cause can be greater than the other. And those causes can only be known from accident modelling. In this paper we have tried to divide this accident modelling techniques into two different categories based on the location of road i.e. accidents on urban roads and on rural roads. In both urban and rural road accident studies it was seen that mainly regression techniques like linear, multi-linear, logit and poisons regression have been used for modelling the road crashes. It was also marked that mostly authors have tried to research on one cause and go deep into it rather considering all factors at a time. From the studies it was found that speed and age along with gender has been the area of study for accident causes in urban areas whereas in rural roads mostly all authors have limited their studies to speed on roads and has been noted as the major cause of accidents in rural areas. This paper has tried to review as much papers as possible and various gaps in research along with future scope of study in this area has been indicated. Starting from the basic models like negative binomial/Poisson's model to the logistic and linear regressions to the new modeling techniques involving genetic mining and fuzzy logics have been discussed explicitly in the paper.

Introduction

Road crashes have been in on an increasing trend in the last decade or so. This has led the researchers to think of this problem and find possible causes and precautionary measures to prevent crashes from happening. This field of transportation engineering is more commonly recognized as traffic safety and management. These researches have led to development and discovery of new models predicting road crashes accurately. This paper combines many important models and discusses the theory involving the discovery to that model. It also compares and contradicts the models developed by different researchers.

Road accidents are very common all over the world and annual global road crash statistics (Association for Safe International Road Travel, 2013) states that, nearly 1.3 million people die in road crashes each year, on average 3,287 deaths a day with an additional 20-50 million are injured or disabled. More than half of all road traffic deaths occur among young adult ages between 15 to 44 years. Road traffic crashes rank as the 9th leading cause of death and account for 2.2% of all deaths globally. Road crashes are the leading cause of death among young people ages between 15 to 29, and the second leading cause of death worldwide among young people ages between 5 to 14 years. Unless action is taken, road traffic injuries are predicted to become the fifth leading cause of death by 2030.

When India is concerned, it is a part of human tragedy. Along with monetary losses it also leads to human sufferings, untimely death, injuries and loss of potential income. During the calendar year 2010, number of road accidents in India is around 5 lakhs and number of deaths due to those accidents is 1.3 lakhs. Number of injuries due to those accidents is 5.2 lakhs. If the age group and the accident data are compared, it is seen that 55% of road accident victims fall in the age group of 25-65 years while out of rest 45%, 40% of road accident victims come from the age group of 16-24 years. It can be concluded that the adolescents are very much prone to and contribute to most of the accidents in India.

Hence, traffic accidents and their safety is a major area of research. So, in this paper, some important models developed for traffic safety along with researches done on the topic are studied and are reviewed thoroughly. At first the general factors affecting the road crashes and general models developed for predicting road crashes are discussed in brief. It follows the literatures of road crashes in urban and rural roads and then a discussion involving comparison and contradiction of models. Also the future scope of the study is discussed in detail.

This paper has the following structure. In Section 1, we present the factors responsible for crashes; in Section 2, we present some models for traffic safety; In Section 3, we present a discussion of the accident models. Finally, we present the Conclusions of the paper.
1. Factors responsible for crashes

Traffic safety and accident studies have been in the research area for last two decades extensively as the rise of accidents have been alarming across the world. From the works done by researchers, it can be said that traffic accidents are caused due to mainly 3 factors i.e.:
- Personal or human behavioural factors
- Road and Environmental factors
- Vehicle factors

Personal or human factors mainly include the age of driver or victim, gender of the victim, was he drunk while driving, etc. Similarly, environmental factors include the general factors of climate and environment, lighting conditions of road, time of accident, i.e. day or night, pavement conditions, etc. Road geometric factors include the type of junction or intersection, then horizontal slope, curves, etc. present on the road, due to faults of which, accidents may occur. At the end come traffic factors. This mainly includes the speed, density, traffic flow parameters that may lead to accidents.

2. Models for traffic safety

Many models have been devised by the researchers in past for accident safety, causes of accidents safety, accident severity crashes, etc. and also precautionary measures have been stated. Though the most common models used are the regression models, but there are many other techniques that have been used in the modeling by the researchers. Some of them are:
- Genetic mining approach
- Logit models, both multinomial and binomial
- Regression models which includes various types like linear, non-linear, logistic regression techniques
- Bayesian-cohort model, etc.

This paper has divided the traffic safety models mainly into two parts under which they will be studied. They are:
- Accident study in urban roads
- Accident study in rural roads

2.1. Accident study in urban roads

Fridstrom et al. (1995) measured the contribution of Randomness, Exposure, Weather and Daylight to the variation in Road Accident counts. Using a generalized Poisson regression model, the variation in accident counts in 4 European countries were calculated into parts attributable to randomness, exposure, weather, daylight, or changing reporting routines and speed limits. A set of specialized goodness-of-fit measures were also developed by them. Pure randomness was seen to “explain” a major part of the variation in smaller accident counts (e.g. fatal accidents per county per month), while exposure was the dominant systematic determinant. The relationship between exposure and injury accidents appears to be almost proportional, while it was less than proportional in the case of fatal accidents or death victims. Together, randomness and exposure accounted for 80% to 90% of the observable variation in their data sets.

Graham and Glaister (2003) examined the role of urban scale, density and land-use mix on the incidence of road pedestrian casualties. The study used English census wards as the spatial unit of study and developed negative binomial models to carry out the analysis. The study concluded that the incidence of pedestrian casualties and Killed and Seriously Injured (KSI) were higher in residential areas than in business areas. In addition, the relationship between urban density and pedestrian casualties were found in quadratic form with incidents reduced in highly populated wards.

Griebe (2003) described some of the main findings from two separate studies on accident prediction models for urban junctions and urban road links described in (Griebe and Hemdorff, 1995, 1998). They established simple, practicable accident models that can predict the expected number of accidents at urban junctions and road links as accurately as possible. The models can be used to identify factors affecting road safety and in relation to ‘black spot’ identification and network safety analysis undertaken by local road authorities. The accident prediction models are based on data from 1036 junctions and 142km road links in urban areas. Generalised linear modelling techniques were used to relate accident frequencies to explanatory variables. The estimated accident prediction models for road links were capable of describing more than 60% of the systematic variation (‘percentage-explained’ value) while the models for junctions had lower values.

Noland and Quddus (2005) developed a disaggregate spatial analysis based on enumeration district area to examine the effect of congestion on traffic casualties (KSI and slight injuries). In this study, congestion was spatially controlled by proxy variables. Negative binomial models were used to analyze the factors affecting casualties during congested and uncongested periods. The study result showed that traffic casualties are likely to happen on higher speed roads and motorways but not during traffic congestion.

Aguero-Valverde and Jovanis (2006) developed Full Bayesian (FB) and negative binomial models to carry out spatial analysis of fatal and injury crashes in Pennsylvania. The study used counties as the spatial unit. The study concluded that counties with a higher percentage of the population under poverty level, higher percentage of their population in age groups 0-14, 15-24 and over 64, and increased road mileage and road density have significantly increased crash risk. The study also suggested that it was important to consider spatial correlation in road-segment and intersection-level accident models.

Wedagama and Dissanayake (2010) studied the influence of accident related factors on road fatalities considering Bali province in Indonesia as a case study. Logistic regression models were separately developed for fatal accidents considering motorcycles and all vehicles including motorcycles with data from Bali in Indonesia. Seven predictor variables were employed in the developed models. The study found the probabilities of female motorcycles and motorists were about 79% and 72% respectively contributing more on motorcycle and motor vehicle fatal accident than males. In addition, age was also significant to influence all vehicle fatalities. Age was accounted for about 50% to influence all vehicle fatalities.

Hauque et al. (2010) performed a detailed study of accidents and severity crashes involving motorcycles as vehicles. The main objective of their study was to evaluate how behavioural factors influence the crash risk and to identify the most vulnerable group of motorcyclists. A questionnaire containing 61 items of impulsive sensation seeking, aggression, and risk-taking behaviours was developed. By clustering the crash risk using the medoid portioning algorithm, a log linear model
relating rider behaviours to the crash score/number was developed by him. Aggressive and high risk-taking motorcyclists are more likely to fall under the high vulnerable group while impulsive sensation seeking behaviour is not found to be significant.

Seva et al. (2012) studied the motorcycle accidents in the Philippines considering personal and environmental factors. The variables considered by them for study were age, lighting conditions, traffic movement, road character, junction type, day, surface conditions, and driving behaviour. Logistic regression was used to predict the likelihood of an accident from the variables considered and a logit model was thus developed. According to their study, three variables were found to be significant predictors of motorcycle accidents. They were age, driving behaviour, and junction type. Wald's and Hosmer-Lemeshow test were used by them as logistic regression for goodness of fit. The main conclusion from their study was that the younger drivers are more likely to be involved in accidents.

Obaidat and Ramadan (2012) studied the traffic accidents at 28 hazardous locations of urban roads at Amman-Jordan roads. Their study found that the logarithmic and linear models were the most significant and realistic models that can be used to predict the relationship between the accident characteristics as a dependent variable and the other studied variables as independent variables. The following variables were found to be the most significant contributors to traffic accidents at hazardous locations: average running speed, posted speed, maximum and average degree of horizontal curves, number of vertical curves, median width, type of road surface, lighting (day or night), number of vehicles per hour, number of pedestrian crossing facilities and percentage of trucks. According to them, these factors form the contributions of different categories causing traffic accidents such as:

- Geometric factors: Number of lanes, width of one way of the road, median width, number and types of pedestrian crossings, number of horizontal and vertical curves, and maximum and average degrees of curvature.
- Drivers' behavioural factors: Posted and average running speeds.
- Environmental factors: Surface type, lighting condition and day/night.
- Traffic conditions factors: Traffic volume (ADT) and percentages of different vehicles.

Anowar et al., (2013) analyzed the accident patterns at selected intersections of an urban arterial in Dhaka. Data show that intersection accidents represent around 40 percent of total accidents occurring in the Metropolitan city of Dhaka. Based on the data analysis, the study also attempted to shed some light on the major causes, factors and types of accidents in order to identify the problem intersections and suggest appropriate counter-measures to reduce such accidents. The study researched 2 types of intersections – Urban and Suburban Intersections with 4 types of accident patterns – fatal, grievous, simple and just collision.

2.2. Accident study in rural areas

Kloeden et al. (2001) studied the effect of travelling speed and the risk of crash involvement on rural roads. The relationship between free travelling speed and the risk of involvement in a casualty crash in 80 km/h or greater speed limit zones in rural South Australia was quantified using a case control study design. It was found that the risk of involvement in a casualty crash increased more than exponentially with increasing free travelling speed above the mean traffic speed and that travelling speeds below the mean traffic speed were associated with a lower risk of being involved in a casualty crash. The effect of hypothetical speed reductions on all of the 167 crashes investigated indicated large potential safety benefits from even small reductions in rural travelling speeds.

Shankar et al. (1995) explored the frequency of occurrence of highway accidents on the basis of a multivariate analysis of roadway geometries (e.g. horizontal and vertical alignments), weather and other seasonal effects. Based on accident data collected in the field, a negative binomial model of overall accident frequencies was estimated along with models of the frequency of specific accident types. Interactions between weather and geometric variables were proposed as part of the model specifications. The results of the analysis uncover important determinants of accident frequency.

Karlaftis and Golias (2002) studied the effects of road geometry and traffic volumes on rural roadway accident rates. The results showed that although the importance of isolated variables differs between two-lane and multilane roads, ‘geometric design’ variables and ‘pavement condition’ variables are the two most important factors affecting accident rates. Further, the methodology used in this paper allows for the explicit prediction of accident rates for given highway sections, as soon as the profile of a road section is given.

Taylor et al. (2002) also studied the relationship between speed and accidents on rural single-carriageway roads. The data collected for the study include Accident data (for a defined 5 year period), Traffic flow data, Vehicle speed data, Road characteristics and Geometric and layout data. Generalized linear modelling procedure was adopted for study of the data and their correlation. The main results of their studies was that the accident frequency in all categories increased rapidly with mean speed – the all accident frequency increased with speed to the power of approximately 2.5 - thus indicating that a 10% increase in mean speed results in a 26% increase in the frequency of all injury accidents. The effect of mean speed was found to be particularly large for junction accidents. A 10% increase in mean speed would be expected to result in a 30% increase in the frequency of fatal/serious accidents.

Hills et al. (2002) developed a safe and cost-efficient model for rural roads designing in developing countries considering the accidents occurred there. They selected 5 countries for their study which were Zimbabwe, Botswana, Malawi, Tanzania, India and Nepal and developed separate models for each of them. They used the Generalized Linear Modeling technique (GLIM) for modeling the data collected. It was found that a reasonable model fit could be made for all accident types but that the numbers of individual accident types were too small to produce reliable individual models. In the Papua New Guinea study, curvature and gradient proved to be significant explanatory variables. According to the model, the presence of a marked edge line in Nepal and India appears to be particularly beneficial in reducing accident rate.

Rengarasu et al. (2007) investigated the road geometry factors and the seasonal factors associated with head-on collisions and single vehicle collisions occurred in Hokkaido, Japan. Head-on collisions represent about 20% of all traffic collisions on the rural two lane national roads however; head-on collisions were responsible for about 40% of the fatal collisions. They developed a segmented accident database based on Traffic Accident Analysis System (TAAS) produced by Civil Engineering Research Institute for Cold Region Hokkaido. Analysis using Poisson-regression models showed that road
geometric factors and seasonal factors were important factors correlated with head-on collisions. The model proposed in this study is potentially capable of identifying the causal factors of head-on and single vehicle collisions.

Chiou et al. (2010) studied the contributory factors to crash severity in Taiwan’s freeways which is considered to be rural roads using genetic mining approach. They considered a very large number of variables. They were surface condition, signal control, driver gender, weather, obstacles on road, lighting conditions, speed limit, road status, marking, license, occupation and age of driver, travel period and purpose, location, vehicle type, action of driver, collision type and severity, etc. According to the model developed by them, travel period, major cause, collision type and purpose of the journey are the 4 key factors to the severity of accidents and hence attention must be given to these 4 factors to ameliorate the traffic safety.

Mustakim and Fujita (2011) developed an accident predictive model for rural roadway based on the data collected at rural roadway, Malaysia. They carried out black spot study to develop accident predictive models. Multiple non-linear regression method was used to relate the discrete accident data with the road and traffic flow explanatory variable. Their results showed that the existing number of major access points, without traffic light, rise in speed, increasing number of Annual Average Daily Traffic (AADT), growing number of motorcycle and motorcar and reducing the time gap are the potential contributors of increment accident rates on multiple rural roadway. The final accident prediction model for the route was found out to be:

\[ \ln(\text{APW}) = 0.350 \times (\text{AADT}) + 0.00409 \times (\text{AS})^{1.75} + 0.00002 \times (\text{AAP})^{0.01898} \times (\text{GP})^{0.035} \]

\[ \ln(\text{APW})^{55} = 0.0402 \times (\text{AP}) + 0.0150 \times (\text{AS}) + 0.0001 \times (\text{MC+C}) + 0.1092 \times (\text{GP}) + 0.0830 \times (\text{APW}) \]

The above models have R² of 0.9973 and 0.9979 respectively. Hence, it is noted that the factors which contribute to accidents at four lane two way undivided rural roadway are, number of access points (AP), vehicle speed (AS), Annual Average Daily Traffic (AADT), motorcycle (MC), motorcar (C), gap (GP) and total length of the Accident section (TL).

3. Discussion

3.1. Urban Areas

The various accident models discussed here shows that regression models are most commonly used in the field of traffic safety by the researchers, though it should also be marked that some new models have also been in the study like the Multinomial Logit i.e. MNL, Bayesian method and negative binomial distribution. Almost all the factors have been studied by different authors, which seem to affect the accidents in urban areas. Except Seva et al. (2012) and Obaidat and Ramadan (2012), others have tried to investigate single factor causing accidents in detail, not taking all factors at a time. The model by Seva et al. (2012) was a very good model considering all the general factors and preparing MNL after testing goodness of fit by logistic regression, but it could have added the road geometry factors as well. After going through all the above mentioned models, it seems that the study by Obaidat and Ramadan (2012) is the most accurate as it has considered almost all factors responsible for accidents. Although it seems to be the most accurate but they could have prepared much better model like logit or logistic regression for more accuracy though. Even the study conducted by Fridstorm et al. (1995) to model the effect of weather, randomness, exposure, etc on accidents was a good study considering the year of conducting experiment and even poisons distribution has been used by the researcher to model their data and randomness was found to be a significant factor for small accidents. Similarly the study conducted on age and gender factors affecting accidents in Bali province, Indonesia by Wedagama and Dissanayake (2010), was a very narrow model in terms of factors but in terms of analysis it was a very high end model with great accuracy. The study by Hauque et al. (2010) was quite common but the variables taken were new. Overall, it was a good motorcycle accident survey and modeling. Graham and Glaiser (2003) did a full-fledged urban study where the urban density of population, land use pattern have been taken into consideration which are believed to be the important factors in urban areas. Negative binomial provides result with good accuracy when the probability of occurring is very less. Noland and Quddus (2005) in their study added a factor of traffic congestion to the above discussed studies which was significantly a new addition to their model, but the study only takes a single factor into account. The model is based on that same as previous i.e. the negative binomial distribution. Study of personal/human behavior, road geometry and traffic conditions for occurrence of accidents was done by Agueiro-Valverede and Jovannis (2006). It was again a good study with good accuracy considering the methods used for analysis i.e. the negative binomial and Bayesian methods. The study by Anowar et al. (2013) was a good study in terms of research in urban conditions in Asian continent. It was not possible to draw the collision diagrams for the high accident locations due to lack of sketches of accident locations in the accident report forms and other factual information. Also the exposure data (traffic volume, vehicle travel), an important piece of information for safety related studies, was not available for the study.

3.2. Rural Areas

The studies conducted by Shankar et al. (1995) was a good study seeing the factors that he undertook as research for his study. He used the negative binomial model for his study. The results of his analysis uncover important determinants of accident frequency. By studying the relationship between weather and geometric elements, this paper offers insight into potential measures to counter the adverse effects of weather on highway sections with challenging geometries. Researchers like Karlaftis and Golas (2002) also studied the impact of Traffic volume on accidents in rural roads. Mustakim and Fujita (2011) did a nice full-fledged study on all aspects of traffic factors related to occurrence of accidents. Even the model used gives good results when goodness of fit is considered. Kloeden et al. (2001) used a new method used for drawing the speed curves to know its effect on accidents called the hypothetical crash outcome method. Study by Taylor et al. (2002) was another good study with traffic factors considered for study in detail and road geometric data to some extent. The study by Chiu et al. (2010) was itself very good and the findings were of much importance. Software algorithm has been used for modeling. The most important finding of this model is that almost every cause of happening accident is considered for the study and almost all factors were modeled. One of the best studies so far in the field of accident study and also accurate as it uses genetic mining approach for its modeling. Rengaraju et al. (2007) studied road geometric factors in detail in their paper and also regression models were developed with better R² values. It’s a good study considering one aspect of traffic safety has been researched deeply. Among the reviewed models, the most accurate study in this field seems to be done by Hills et al. (2002). It’s a good model considering the research was spread over 5 developing countries including India, and
comparison of models were also done. Also the factors considered were appropriate considering that the models were developed for different countries.

Conclusion

The whole study revolved around the modelling of road crashes and it can be concluded from the literature of various researchers that though there has been ample amount of research in this field of road safety management, still many developing countries have not been able to decrease their share of deaths in road crashes. Hence, much study is needed in the field of traffic safety and planning.

Statistical methodologies have been used to model the data and findings obtained from survey for a better and easy understanding. The most common models used are the regression techniques (linear, logistic, multiple) and few authors use regression techniques for finding goodness of fit and then model the equations and coefficients into multinomial logit models. Road crashes account for nearly one-half of all teenage deaths. In case of accidents in urban roads, many variables like age of drivers, gender, running speed, road conditions, lighting conditions, etc. are found to be the causative agents of accidents. It can also be seen that researchers usually try to focus on one variable that cause accidents and study it thoroughly rather than considering all factor at a time. In urban road accidents study, some models developed were very accurate considering the used of all forms of regression i.e. linear, nonlinear and multiple linear regressions. In rural road accidents, it is observed that mostly researchers consider speed as a major cause for accidents to occur. Few studies also considered almost every possible factor affecting accidents in rural roads and also a new software based algorithm and approach was used known as the genetic mining approach for modeling the data. Results of the research conducted by few researchers have showed that the major cause of traffic accidents was careless driving (71%) in developing countries. These are few conclusions that can be noted down from the literatures and discussions mentioned in the article.

Many future studies open up in this area of research. Though traffic congestion is a wide spread problem for all developing countries and when we talk about India, it is ranked as top in number of accidents and accidental deaths per year in whole world, still studies of traffic safety have been limited to developed countries mainly. The models mainly used the regression techniques. They are good but very old and conventional. New approaches like genetic mining fuzzy logics have coming and also all the old approaches to the new approaches are more accurate and software oriented so more user friendly. More research on integration of traffic safety with systems and software should be done. Better planning strategies with good management system should be employed for averting the risks posing accidents occurrence.

References


MORTH (2010) Road Accidents in India.


