

Available online at www.sciencedirect.com

Procedia Computer Science 3 (2011) 764–769

**Procedia
Computer
Science**

www.elsevier.com/locate/procedia

WCIT-2010

The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach

Alireza Pakgohar^{a*}, Reza Sigari Tabrizi^b, Mohadeseh Khalili^c, Alireza Esmaeili^d^aFaculty of Statistics, Payame Noor University(PNU), Branch of Ardakan, Iran. a_pakgohar@pnu.ac.ir^bMBA, Multimedia University, FOM, Cyberjaya, Malaysia. reza.tabrizi@live.com^cDepartement of Statistics, Payame Noor University(PNU), Branch of Meybod, Iran. mohadeseh_khalili@yahoo.com^dFaculty of Traffic, Oloome Entezami University(Police), Tehran, Iran. ar.esmaeili2007@gmail.com

Abstract

Accidents are one of the biggest public health problems in the world. As literature indicated, the traffic accidents were assessed to be most significant health problem in Iran. To date, no serious researches have analyzed high dimensional traffic data In Iran. This paper, therefore, aims to bridge the gap. In this study, the traffic data are analyzed by Data Mining techniques such as Logistic Regression, Classification and Regression Trees. In this paper the impact of such factors were investigated using these techniques. It is hoped that the current research findings will help governments in better road designs and traffic management.

© 2010 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](http://creativecommons.org/licenses/by-nc-nd/3.0/).

Selection and/or peer-review under responsibility of the Guest Editor.

Keywords: Data Mining, Classification, Regression Trees, CART, LR, Iran Road Accident

1. Introduction

Accidents are one of the biggest public health threats in the world [1]. According to statistics, 1.2 million people are died and 50 million people are injured annually from road crashes in the world [2]. Based on the statistics, 27567 people died and 276762 people were injured in vehicle collisions in 2007 [3]. The problem is worse because most victims were young people who had been in full health before the accident [1]. Almost 10 percent of accidents' victims die [2]. According to statistics, in average about 100 people are killed in road accidents in Iran every day, however; 3000 people are killed in road accidents in the world daily [1]. One percent of world's population belongs to Iran whereas country has 1 out of 40 of deaths from road accidents in the world. Hence, attempts to improve travels' safety and to reduce hazards of road accidents through development and application of traffic safety programs are essential tasks. What is still considered by the traffic experts is to identify the factors affecting the incidence or severity of an occurred crash. The traffic safety is subjected to vast and complex dimensions in which to be interacted together and consequently demands various knowledge and experiences. These knowledge and experiences are applied in three categories of factors including: Human factors, Road, and Vehicles [4]. Previous studies indicated that the role of the human factor influence on road safety and accidents close to 90 percent [4].

* Alireza Pakgohar. Tel.: +60-172455659
E-mail address: reza.tabrizi@live.com

2. Background

In many countries included the United States, many researchers studied the role of three factors namely “Human”, “Road and Environment” (In this article briefly called “Road”) and “Vehicle or Car” in severity of road crashes. According to the results and findings of these studies, 93% of all road crashes really have driver distraction (Human factor) as their root cause whereas 34% of all road accidents had been occurred because of the road factor. This was followed by role of vehicle factor with contribution of 12% [5]. According to national statistics, around 27000 people died and 276000 people were injured from road crashes in 2007 in Iran [3]. In average about 100 people are killed in road accidents in Iran every day, however; 3000 people are killed in road accidents in the world daily [1]. Reference [6] stated that the human factor has been found to contribute to crashes at a level of 97% in the world.

A logistic regression was employed by [7] to determine the factors influenced the road crashes. They focused on the environment factor to perform their study. Reference [8] employed data mining approaches such as logistic regression, regression tree, and neural network to model accident severity in Korea. They then compared the classification accuracy of their three classification techniques. Reference [9] combined GIS techniques, regression modeling, and cluster analysis to integrate all traffic data for having a comprehensive estimation of road accidents in Hong Kong. Reference [10] employed non-parametric classification tree technique to analyze traffic injury severity. They used classification and regression tree as a data mining technique to investigate the relationship between accident severity and driver’s attributes, and environmental characteristics. According to [10], pedestrians, motorcycle and bicycle riders had more serious injured in road crashes compared to vehicle passengers. The negative binomial regression models and classification and regression tree (CART) were used by [11] to analyze traffic road accident of the National Freeway in Taiwan. According to their study, the CART is much better method to analyze traffic accidents’ data.

3. Research Methodology

The focus of this research study is to identify the role of human factor in incidence and severity of road crashes. Unsurprisingly, so far there looks to be no other same research carried out with the similar purpose in the context of road crashes in Iran. Hence, attempts were made to explore and describe certain issues on role of human factor in severity of road crashes in Iran. Therefore, this study is classified as an applied research. For the purpose of this study mainly descriptive analysis, Logistic Regression, Classification and Regression Tree were employed. The data were used to carry out this research study were based on Database of Traffic Accidents of Iran’s Police. The size of the target population was 347285 road crashes that were occurred in Iran during 2006.

3.1. Research Objective

The research objective is:

- To find out the role of human factor in incidence and severity of road crashes in Iran

3.2. Description of key Variables

The accident record in Data Base of Police in Iran has more than thirty columns or attributes some of which are, light conditions, road direction, road surface, weather condition, type of driving license, age, gender, and safety belt. In this study, we only focused on human factor variables. The details of CART and LR outputs by using SPSS are not described for all variables because of space limitations, so only the impact of “Safety Belt” and “Driving License” on Severity of Road Crashes will be discussed in the following sections. Table 1 represents the variables, type of variable, and their descriptions.

Table 1. Description of key variables

Variable	Type	Description
Age	Independent	Age of the drivers
Gender	Independent	Gender of the drivers
Safety Belt	Independent	The situation of fasten safety belt
Driving License	Independent	Driver's Certificate
Accident Severity	Dependent	The severity of the crash

3.3. Experimentation

In order to evaluate accident severity, different classification models were made using CART and Logistic Regression. The data were assessed and then were added to a clean dataset included 347285 records. The dependent variable (Accident Severity) had three levels included "Fatal", "Injury", and "No Injury". During running CART and LR algorithms through SPSS, the software's defaults were adopted. After executing algorithms, the accuracies of 81% and 78.57% were achieved for CART and LR respectively. This is approved that CART has higher accuracy than LR method.

4. Results and Findings

In this paper, the CART and Multinomial Logistic Regression (MLR) were carried out using SPSS 16.0 to classify severity of road crashes in Iran.

4.1. Accident Classification using Multinomial Logistic Regression

Since, the dependent variable of this study has three different levels, thus the usual logistic regression cannot be used. Therefore, Multinomial Logistic Regression is a useful alternative for this situation in which the dependent variable is not limited to two categories. Table 2 represents the results of classification based on the Multinomial Logistic Regression. According to Table 2, 2567 of the 3426 deathly accidents were classified correctly. As shown in Table 2, 21543 of the 28195 people who injured in road crashes were classified correctly. Finally, it can be clearly seen that 248654 of the 315664 accidents that had no injury people were classified correctly. All in all, 78.57% of all accidents were classified correctly.

Table 2. The Classification Results for MLR Method

Observed	Predicted			Percent Correct
	Fatal	Injury	No Injury	
Fatal	2657	651	118	0.7755
Injury	5543	21543	1109	0.7641
No Injury	13856	53154	248654	0.7877
Overall Percentage	0.0635	0.2170	0.7195	0.7857

4.2. Accident Classification using CART

To continue the study of the role of human factor on severity of road crashes, the CART approach was selected because of the following reasons:

- Higher accuracy compared to MLR
- More simpler to understand and interpret
- Ability to use both numerical and categorical data.

The main aim of this study is to investigate the human factor affecting prediction and classification of the

accident severity. Accordingly, Driving License, Safety Belt, Age, and Gender were considered as attributes of human factor that impact crash severity in Iran’s roads. As shown in Fig 1, 91% of accidents have resulted in no injury. According to Fig 1, 8% of which have resulted in injury and only 1% of all accidents have ended to fatal.

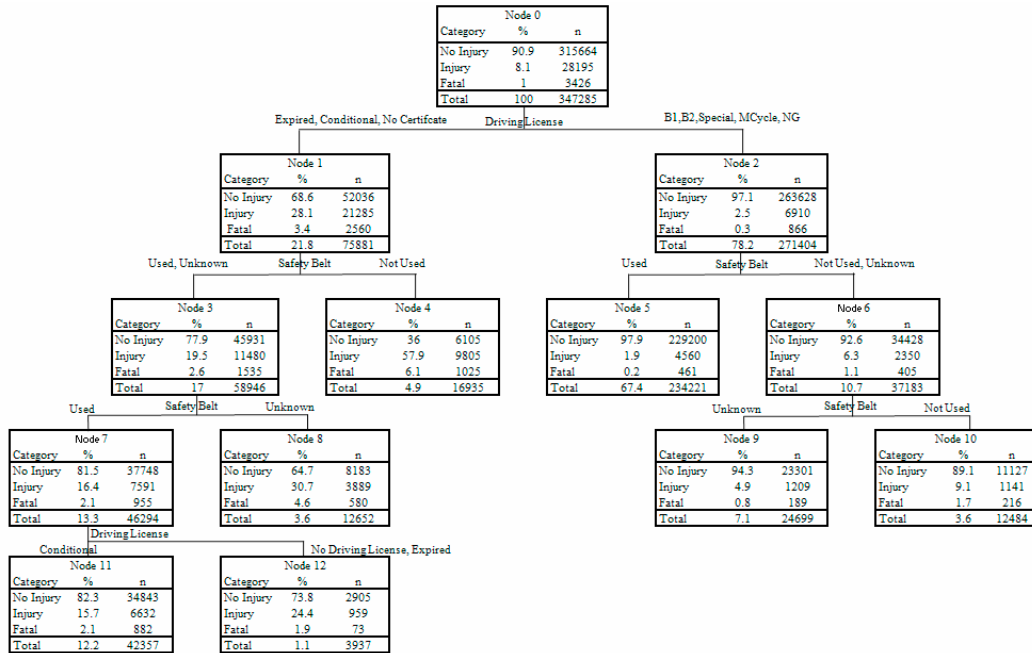


Fig 1. The CART Classification for the Impact of Human Factor on Accident Severity

4.3. The Role of Driving License and Safety Belt

The rates of “Injury” and “Fatal” for those who had no driving license or conditional driving license were 28 percent and 3 percent respectively. It was found that these rates increased to 52% (Injury) and 6% (Fatal) if car passengers did not fasten their safety belt. Moreover, approximately 22 percent of those hurtled were classified as “No Certificate”, “Conditional Driving License”, and “Expired Driving License”. The rates of “Injury” and “Fatal” for those who had driving license (1st Grade, 2nd Grade, Motorcycle Driving License, and Special) were 2.5% and 0.3% respectively. In addition, these rates decreased to 1.9% (Injury) and 0.2% (Fatal) for those who fastened their safety belt.

4.4. The Role of Gender and Age

According to the results of CART analysis, gender of drivers had no significant impact on severity of accidents. It was discovered that most people less than 19 years old (61%) had not fastened their safety belt. One percent of all drivers had less than 19 years old. Hence, 1 out of 10000 fatal accidents have happened for those who had lesser than 19 years old. According to CART analysis, there is a significant relationship between “Age” and “Commitment to Fasten Safety Belt”. Furthermore, there is a negative significant relationship between “Age” and rate of “Fatal” among accidents in Iran.

5. Discussion on Results and Findings

According to the Classification algorithms, human factor contributed in 97.5% of all accidents. In contrast, the environmental factor had the second role and contributed in 70.5% of all road crashes. In addition, the vehicle factor contributed in 31.5% of all accidents. In 2006, more than 22 percent of those killed in road accidents were non-passengers such as pedestrians and cyclists. In addition, nearly 87% of non-passengers who killed in road crashes were pedestrian. Furthermore, 55% of all car passengers who died in accidents were driver. 61.6% of those who had an accident owned a second grade driving license and only 16.6% of whom had first grade driving license. In addition, 2.4% of whom had no driving license. According to the statistics, 67.9% of persons killed in road crashes had “Conditional” driving license whereas 18.3% of whom had second grade driving license. Furthermore, 6.7% of those had no driving license. 65.8% of those who have not injured had second grade driving license while 61.8% of those who have injured had “Conditional” driving license. According to data analysis, 81% of all drivers who had an accident in 2006 used their safety belt whereas 8.1% of those had not used and situation of 11% of whom was unknown. According to CART analysis, 95% of those who fastened their safety belt were not injured in road accidents whereas 37% of those who had not used their safety belt were injured.

6. Conclusion

A review through literature disclosed a research gap on the establishing the relationship between human factor and severity of road crashes using data mining techniques in Iran. Hence, attempts were made to discover the role of human factor in incidence and severity of road crashes based on the CART and LR approaches. The applicability of the CART and LR as data mining techniques on predicting severity of road crashes were investigated in this study. The results revealed that “Driving License” and “Safety Belt” represent distinct relationship with the degree of injury in road crashes in Iran. The investigation also represented that gender of drivers had no significant relationship with the degree of injury. In addition, the study provided sufficient witnesses that age of the drivers have significant negative relationship with rate of deadly accidents in Iran. The results of this study indicates the important role of human factor such as “Driving License” and “Safety Belt” in severity of accidents in Iran. Finally, it is hoped that the outcomes of this study can be used by the relevant managers to enhance road safety.

References

1. L., Evans. *Traffic Safety*. [book auth.] Mich Bloomfield. Science Serving Society. 2004.
2. (WHO)., World Health Organization. Burden of disease project. Global burden of disease estimates for 2001. WHO. [Online] 2001. <http://www3.who.int/whosis/menu.cfm?path=burden>.
3. The Effect of Human Factor on Occuring Road Accidents in Karaj-Ghazvin Highway in 2005:Control for Reduction. Ebrahim, Vahab Zadeh. 8, 2008, *Journal of Studies in Traffic Management*, Vol. 9.
4. The Preliminary Plan for Budgeting the Costs of Social Traning of Traffic. M.R., Behnood and A.R., PakGohar. Tehran : The Organization of Transportation and Traffic Management, 2008. The 8th Conference on Transportation Engineering and Traffic in Iran.
5. Mir Fazel, Nik Zad. *Country's Traffic Accident, Causes and Strategies to Solve the Problem*. Tehran : Bureau of Applied Research of Traffic Police, 2007.
6. A.R., Pakgohar. *A Study on the Critical Factors Influence the Reduction Rate in Traffic Accidents Using CART and LR Approaches*. s.l. : Bureau of Applied Research police traffic, 2007.
7. Roadway safety in rural and small urbanized areas. P. J., Ossenbruggen, J., Pendharkar and al., et. 4, 2001, *Accidents Analysis and Prevention*, Vol. 33, pp. 485-498.
8. Pattern recognition for a road traffic accident severity in Korea. S., Sohn and S., Hyungwon. 1, 2001,

Ergonomics, Vol. 44, pp. 101-117.

9. An algorithm for assessing the risk of traffic accidents. K. S., Ng, W. T., Hung and al., et. 2002, *Journal of Safety Research*, Vol. 33, pp. 387-410.

10. Analysis of traffic injury severity: An application of non-parametric classification tree techniques Accident analysis and prevention. L., Chang and Wang, H. 5, 2006, *Accident analysis and prevention*, Vol. 38, pp. 1019-1027.

11. Data mining of tree-based models to analyze freeway accident frequency. L., Chang and W., Chen. 2005, *Journal of Safety Research*, Vol. 36 , pp. 365-375.