Contents lists available at Sjournals

Scientific Journal of
Pure and Applied Sciences

Journal homepage: www.Sjournals.com

Original article

Prediction of vehicle traffic accidents using bayesian networks

S.Sh. Alizadeh[a], S.B. Mortazavi[a],*, M.M. Sepehri[c]

[a]Ph D. Candidate of Occupational Health Engineering, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.
[b]Professor of Occupational Health Engineering, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.
[c]Associate Professor of Department of Industrial Engineering, Faculty of Engineering, Tarbiat Modares University,Tehran, Iran.

*Corresponding author; Professor of Occupational Health Engineering, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.

A R T I C L E   I N F O

A B S T R A C T

Every year, thousands of vehicle accidents occur in Iran and result thousands of deaths, injuries and material damage in country. Various factors such as driver characteristics, road characteristics, vehicle characteristics and atmospheric conditions affect the injuries severity of these accidents. In order to reduce the number and severity of these accidents, their analysis and prediction is essential. Currently, the accidents related data are collected which can be used to predict and prevent them. New technologies have enabled humans to collect the large volume of data in continuous and regular ways. One of these methods is to use Bayesian networks. Using the literature review, in this study a new method for analysis and prediction of vehicle traffic accidents is presented. These networks can be used for classification of traffic accidents, hazardous locations of roads and factors affecting accidents severity. Using of the results of the analysis of these networks will help to reduce the number of accidents and their severity. In addition, we can use the results of this analysis for developing of safety regulations.

1. Introduction

Traffic accidents are a major problem in developed and developing countries and driving safety and its management is one of the main concerns of each community. Annual worldwide 1.2 million people killed in traffic accidents, and 20 to 50 million people are injured and disabled (Peden M 2002, Peden M 2005). The number of traffic accidents and their effects reveals the importance of analysis the factors affecting the occurrence of these accidents. In order to reduce traffic accidents, many ways, including engineering measures, enforcement, education and culture exist. However, the engineering measures are the most effective in improving road safety (B.G. Heydecker 2001). Identify the factors affecting the severity of injuries caused by traffic accidents were the main target of many studies. Some examples of these factors include driver characteristics, road characteristics, vehicle characteristics, crash characteristics and climatic factors (Chang 2006, Kopelias 2007, Juan de O˜na 2011). One important aspect of the safety program is the ability to predict the occurrence of traffic accidents on roads (Sando 2005). Recently, researchers have used the data mining techniques such as neural networks, regression trees and Bayesian networks (Abdelwahab 2001, Randa Oqab Mujalli 2011). Researchers have focused on identifying the most important variables affecting the severity of accidents (Randa Oqab Mujalli 2011). Most previous studies have used regression analysis techniques such as logistic regression (Zarei 2008, Al-Ghamdi 2002, Bédard 2002, Kockelman 2002, Yamamoto 2004, Milton 2008, Randa Oqab Mujalli 2011). Regression analysis widely used to determine the factors affecting the injuries severity (Juan de O˜na 2011). These techniques have their limitations. Regression models that are often used to model the severity of injury have pre-assumption relationships between independent and dependent variables (Chang 2006, Juan de O˜na 2011). If these assumptions are not specified correctly, the model will be incorrect estimate of the likely severity of the injury. Some experts have suggested Bayesian networks for modeling accidents. Using Bayesian networks can generate the structure of relationships between the variables. The networks have three major advantages: 1. Mutual induction, 2. Integration missed variables and 3. Probabilistic interpretation. Using these networks, explore the main patterns of data, relationships between variables and forecast the accidents are relatively easily (Juan de O˜na 2011).

In Iran, descriptive and analytical studies were conducted on traffic accidents. Iran has the highest number of deaths in road accidents in the world and the majority of victims are between 20 to 30 years old. Table 1 shows the frequency distribution of the number of traffic accidents and deaths in the years 1997 to 2006. As it is observed with increasing population and number of vehicles, road traffic accidents and the number of injuries and deaths have increased. So effective measures to reduce the number of traffic accidents and severity of injuries on individuals should be planned and implemented (Zarei 2008).

The aim of this research is to demonstrate that by using Bayesian networks can predict the accidents and the factors influencing them, determine the dangerous parts of the roads and then plan intelligently and effectively to control these accidents.

Table1
Distribution of the number of traffic accidents and deaths in the years 1997 to 2006.

Year	Population	Vehicle number	Traffic accidents number	Injury	Fatality
1997	61897000	6494004	160284	67884	13679
1998	62640000	6804939	200284	79289	14981
1999	63392000	7115874	245284	91048	15482
2000	64153000	7426809	292284	108100	17059
2001	64922000	7737744	340047	117566	19727
2002	65701000	8485285	448304	167372	21873
2003	66490000	10364637	554849	222309	25722
2004	67478000	12323989	625030	245754	26089
2005	68467000	14283341	658256	274257	27746
2006	70473000	16242693	642656	276762	27565

2. Bayesian networks

In the past decade, Bayesian networks have been used as a general method for modeling expert knowledge in expert systems and their theoretical and computational advances are developing in many areas (Mittal 2007, Randa Oqab Mujalli 2011). Some of these areas are modeling in medicine, documents classification, recall information, image processing, decision-making systems, engineering, computer games and law (Randa Oqab Mujalli 2011).

Assume U={x1,....., xn}, n≥1 is a set of variables. Bayesian network of set of variables U includes a graphical structure and probabilistic tables Bp={p(xi| pa(xi), xi∈U} which p(xi) is set of parents and children xi in Bayesian network and i=(1,2,3,......, n). In Bayesian networks to compute the probability distribution of the combined set of variables, the following equation is used:

$$P(X_1, X_2,..., X_n) = \prod_{}^{n} P(X_i | Parents(X_i))$$

Based on the theory of Bayesian networks, arcs between variables show the causal and a direct dependence between variables (Randa Oqab Mujalli 2011).

A Bayesian network is a graphical model to represent the combined probability distribution of the set of variables. Knowledge obtained for a problem is modeled in the form of qualitative and quantitative information. In Bayesian networks, two types of information are presented for each variable:

- Network curves that are used to represent conditional independence relations.
- A table that determines the probability distribution of each node to its immediate parent.

As an innovation in the field of accident investigation, in Bayesian networks each node is connected directly to its parents. Bayesian networks are used to predict or identify the causes that have the greatest impact on traffic accidents (Kim 2006, Xiaoping Zheng 2009).

Compared with neural networks, each node in the network Bayesian has a special meaning. In these networks it is not require any pre-assumptions about the distribution of data. Accidents that have very dependent variables can be examined using these networks. Variables relationships and structures can be studied using these networks (Xiaoping Zheng 2009).

3. Accidents data

Accidents data can be obtain from police traffic centers. Then these data are evaluated and untrue or incomplete data are removed. Variables recorded in the police report traffic that affect accidents and severity of injuries are selected. Some of the most important variables are (Juan de O˜na 2011):

- Injury severity variables: Injury number (such as driver, travelers and pedestrians), injury severity rate (such as fatality, sever, negligible).
- Road information: Path characteristics that the accident occurred (such as pavement width, road width, shoulder type, or lack of paved roads).
- Weather information: The weather conditions at the time of accident (for example good weather, rain, fog, snow, and wind).
- Accident information: Accident type, accident occurrence time (year, month, day, hour), vehicle type.
- Driver information: driver characteristics such as age and sex (Randa Oqab Mujalli 2011).

Table 2 shows some of variables that are needed to investigate accidents (Juan de O˜na 2011, Randa Oqab Mujalli 2011).

Table 2
Some of variables that are needed to investigate accidents.

No.	Variable	Value	No.	Variable	Value
1	Number of victims	1	8	Paved shoulders	Yes
		>1			No
2	Shoulders type	No	9	Time	0-6
		< 1.5 m			6-12
		1.5-2.5 m			12-18
		> 2.5 m			18-24
3	Accident type	Vehicle left	10	Age	18-25
		Collisions with fixed objects			25-64
		Horn crash			> 64
		The multi-car accident	11	Cause	The driver features
4	Weather condition	Good			Way features
		Heavy Rain			Vehicle features
		Moderate rain			Other
		Other			
5	The number of vehicles involved	1	12	Sex	Male
		2			Female
		> 2			
6	Day	Start of the week	13	Season	Spring
		Weekend			Summer
		Holidays			Fall
					Winter
7	Road width	< 3.25 m	14	Month	12 months
		3.25-3.75 m	15	Lighting	Dark
		> 3.75 m			Day light
					Insufficient
					Sufficient

4. Bayesian networks

To construct the Bayesian networks and use them to classify, predict and evaluate the factors influencing the accidents can use the following process(Chang 2006):

4.1. Data gathering

Data gathering is one of the most important steps in creating a Bayesian network. Depending on the type of study there is variety of ways of data gathering such as archival data, tests, reviews, interviews and observations. Commonly in data mining, an existing database is used. These data are provided by police officers at the scene of a traffic accident (Chang 2006).

4.2. Variables selection

Accurate prediction of the accident occurrence depends on data collection and variables used in the construction of Bayesian network. In machine learning, variable selection is a process to select a set of variables and eliminate non-effective variables (Chang 2006).

The variables are selected based on knowledge of the potential causes of accidents, accident reports, as well as variables derived from the literature review (Table 2 are examples of variables) (Chang 2006).

4.3. Date pre-processing

The defined variables using traffic database and literature review evaluate and then a final list of variables and their categories will be prepared.

Injuries caused by accidents are divided into three categories: minor injuries, serious injuries and deaths. Missing or incomplete data or insufficient data variables have been investigated.

4.4. Construction of Bayesian Networks

Bayesian networks construction process and find an appropriate network structure is very important. Bayesian network shows the relationship between the accident and the various factors affecting its occurrence (Fig.1). The network includes two parts: Nodes and arcs. The nodes are variables and the arcs are relationships between variables. Figure 1 displays the relationship between different factors with accident type, severity of injury and error. Various factors such as traffic volume, weather conditions, day of week, road type, age, sex, impact the type of accident and severity of damage. There is interdependence between different factors. For example, the severity of damage is depended to the belt and using the belt is depended to the driver's gender (Sando 2005).

4.5. Network validation

Traffic data obtained from a database are divided into two main categories: Training data set that are two-thirds of the total data and test data set that are a third of total data (Sando 2005). Bayesian network build based on the training set and test using the test data set (Chang 2006, Juan de O˜na 2011).

4.6. Bayesian network evaluation indices

For evaluation of Bayesian networks performance, the following five sub-indices are used (Peden M 2002, Chang 2006, Milton 2008, Juan de O˜na 2011):

$$Accuracy = \frac{tSI + tKSI}{tSI + tKSI + fSI + fKSI} 100\%$$

$$Sensitivity = \frac{tSI}{tSI + fKSI} 100\%$$

$$Specificity = \frac{tKSI}{tKSI + fSI} 100\%$$

$$HMSS = \frac{2 \times Sensitivity \times Specificity}{Sensitivity + Specificity}$$

Which:
tSI: Number of true minor injuries
tKSI: Number of true sever injuries or fatalities
fSI: Number of false minor injuries
fKSI: Number of false sever injuries or fatalities
Accuracy: Proportions of cases are classified correctly by the network. Accuracy only provides information about the overall performance of the network.
Sensitivity: Proportion of minor injuries that are predicted correctly among all minor injuries.
Specificity: Proportion of severe injuries or deaths that are predicted correctly among all serious injuries or all deaths.
Harmonic Mean of Sensitivity and Specificity (HMSS): Presents a weight equals sensitivity and specificity.
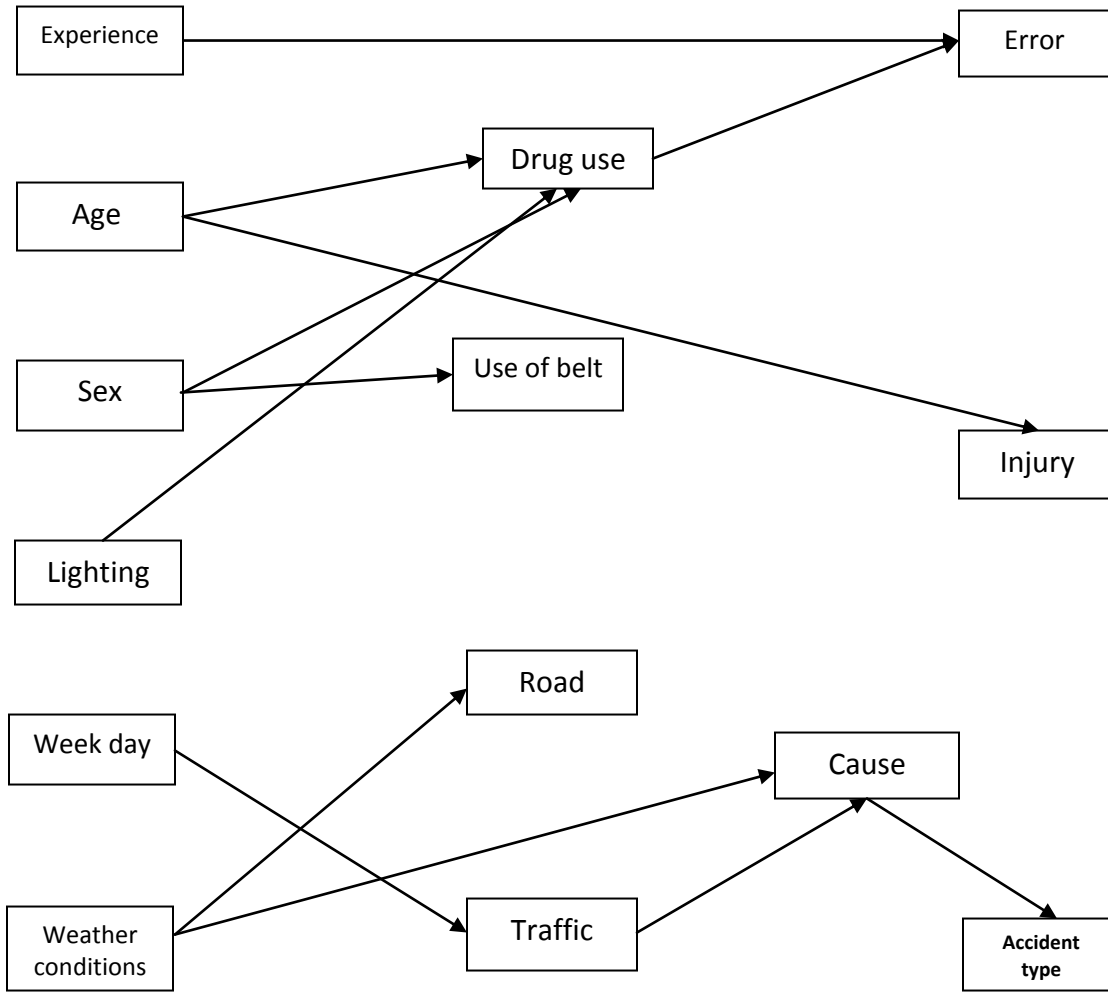
Fig.1. An example of traffic accident Bayesian network.

5. Discussion and conclusion

Bayesian networks have shown their effectiveness in various fields of research and they can also be used in the study and analysis of traffic accidents. Compared with other data mining techniques, the advantage of these networks is their complex vision if multiple dependent variables are exists (Juan de O˜na 2011). Using Bayesian networks can predict injury severity of accident, type of accident and dangerous roads.

Many modeling techniques ranging from simple regression methods to complex knowledge-based methods have been used for modeling data traffic safety. Bayesian networks have several advantages, including a graphical structure that will facilitate understanding of the accident and its influencing factors. The use of conditional and unconditional probabilities makes it closer to reality. Regardless of the number of variables in the network, usually Bayesian networks have good prediction accuracy. It is expected that the use of a comprehensive network can be modeled more effectively to road traffic accidents (Sando 2005).

Using established networks decision makers understand the factors affecting road traffic accidents such as location, road type, vehicle type, driver age, driver gender, weather conditions. Considering each of these cases, they can plan necessary actions. One of the most important advantages of Bayesian networks is its two-way inductive feature. In other words, if a car crash occurred, we can construct its network as a model for analyzing the causes of the accident. Even using this network, we can determine the weight of each factor affecting the accident. On the other hand, if the authorities want with existing status predict an accident, they can use a particular network that built as a template that accident.

References

Abdelwahab, H.T., Abdel-Aty, M.A., 2001. Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. Transport. Res. Rec., (1746), p. 6-13.

Al-Ghamdi, A.S., 2002. Using logistic regression to estimate the influence of accident factors on accident severity Accident. Analys. Prevent., 34, p. 729-741.

Bédard, M., Guyatt, G.H., Stones, M.J., Hirdes, J.P., 2002. The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accident. Analys. Prevent., 34, p. 717-727.

Chang, L.Y., Wang, H.W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid. Anal. Prevent., 38, p. 1019-1027.

Heydecker, B.G., J.W., 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference. Advanc. Eng. Software., 32, p. 859-869.

Juan de O~na, R.O.M., Francisco, J., 2011. Calvo, Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accid. Anal. Prevent., 43 p. 402-411.

Kim, M.C., Seong, P.H., 2006. A computational method for probabilistic safety assessment of I&C systems and human operators in nuclear power plants Reliability Engineering and System Safety., 91(5), p. 580-593.

Kockelman, K.M., Kweon, Y.J., 2002. Driver injury severity: an application of ordered probit models. Accident. Analys. Prevent., 34, p. 313-321.

Kopelias, P., Papadimitriou, F., Papandreou, K., Prevedouros, P., 2007. Urban freeway crash analysis. Transport. Res. Rec., p. 123-131.

Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: An exploratory empirical analysis. . Accident. Analys. Prevent. 40, p. 260-266.

Mittal, A., Kassim, A., Tan, T., 2007. Bayesian network technologies: Applications and graphical models. Hershey, PA: IGI Publishing.

Peden M, H.A., 2002. Road traffic injuries are a global public health problem. BMJ, 324.

Peden M, T.T., 2005. Counting road traffic deaths and injuries: poor data should not detract from doing something!. Annl. Emer. Med., 46, p. 158- 60.

Randa Oqab Mujalli, J.d.O., 2011. A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks. J. Safety Res., 42 p. 317-326.

Sando, T., 2005. Modeling highway crashes using bayesian belief networks techniques and GIS in Dep. Civil. Env. Eng., The Florida State University college of engineering

Xiaoping Zheng, M.L., 2009. An overview of accident forecasting methodologies. J. Loss Prevent. Proc. Industr., 22, p. 484-491.

Yamamoto, T., Shankar, V.N., 2004. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects Accident Analys. Prevent., 36, p. 869-876.

Zarei, M., 2008. Iran traffic accident fatality from 1997 to 2006, Hakim Res. J., 3 (11): p. 42-46.