

# The Analysis and Prevent In Traffic Accidents

## Based on Bayesian Network

Zhu Xiang Xu<sup>1, a</sup> Yi Jiang<sup>2, b</sup> Fan Lin<sup>3, c</sup> Long Dai<sup>4, d</sup>

<sup>1</sup>Department of Computer Science, Xiamen University, Xiamen, 361000, China

<sup>2</sup>Department of Computer Science, Xiamen University, Xiamen, 361000, China

<sup>3</sup>Software school, Xiamen University, Xiamen, 361000, China

<sup>4</sup>Chief doctor, Xiamen Center for Disease Control and Prevention, Xiamen, 361021, China

<sup>a</sup>zxxu2006@163.com, <sup>b</sup>jiangyi@xmu.edu.cn, <sup>c</sup>iamafan@xmu.edu.cn, <sup>d</sup>xmdailong@163.com

**Keywords:** Data Mining, Bayesian Networks, K2 algorithm, Traffic Accident

**Abstract:** The development of the city has led to the frequent occurrence of traffic accidents. Whether we can analyze those accidents which had happened correctly will directly affect the avoidance of future ones of the similar kind. In this paper, we will establish Bayesian Networks traffic accident analysis model by K2 algorithm, which can make accident probability prediction and accident diagnosis. K2 algorithm is known to all with high efficiency and accuracy, but it requires to obtain order first, so to get the reasonable node order, first use clustering algorithm to divide the nodes into groups, in groups the similarity is high with each other. The probability of parent child relationship is larger, then reorder the nodes in every group by the expert experience finally determine the node sequence. Base on this, we can find the system weak links and adopt corresponding effective measures.

### 1. Introduction

The continuous development of city has cause urban traffic problems increasingly, traffic accidents occur frequently. Through analysis of the previous accidents, let's seek the effective ways to reduce the traffic accident probability. the current method is that using the association rules technical in the field of data mining to analyze the traffic accident records. We want to find the rules of the accidents by some frequent factors which caused traffic accidents, at last the obtained results is a set of rules  $\{A_i \rightarrow B_i \text{ (support, confidence), } A_i \rightarrow B_i, i = 1, 2, \dots\}$  [3-5]. through the set of rules we can know what kind of condition plays how much influence in the accident, but it's not enough to play a preventive role. In this paper, we put man, vehicles, roads and the environment as research objects. first, summarize the relationship between composition of factors and traffic accidents and analyze the modeling process and inference in traffic accident analysis applications on the basis of system engineering theory and Bayesian network theory of probability and statistics, in the end, based on before-mentioned study, we make use of accident records of the road traffic accidents information collection project in the ministry, and build the traffic accident analysis Bayesian network model for rational analysis[12].

## 2. Bayesian network theory

Bayesian network is a probability network, which is based on probabilistic reasoning graphical network and the mathematical model, and the Bayesian formula is the foundation of probability network [2]. The so-called probabilistic reasoning is a process that through some variables information to access other probability information, Bayesian network based on the probability reasoning is to solve undetermined and imperfect problems; it has a good advantage in the complicated equipment uncertainty and the fault caused by the relevance, so it obtains widely application in several areas.

Bayesian network consists of two parts [6-7]:

First, Directed acyclic graph with  $N$  nodes  $\mathcal{G}$ : the node in the chart means random variables, the edge connects nodes reflect the dependencies between the variables; the entire graphics structure contains a strong independent relationship among nodes.

Second. The related conditional probability of each node  $CPT$ : every node which has parents has a conditional probability table which expressed the dependencies of its father nodes. A Bayesian network can be expressed a joint probability distribution as (1):

$$P(X_1, \dots, X_N) = \prod_{i=1}^N P(X_i | X_{i-1}, \dots, X_N)$$

$$= \prod_{i=1}^N P(X_i | Pa_i)$$
(1)

The flow chart of bays nets modeling is shown in Fig.1:

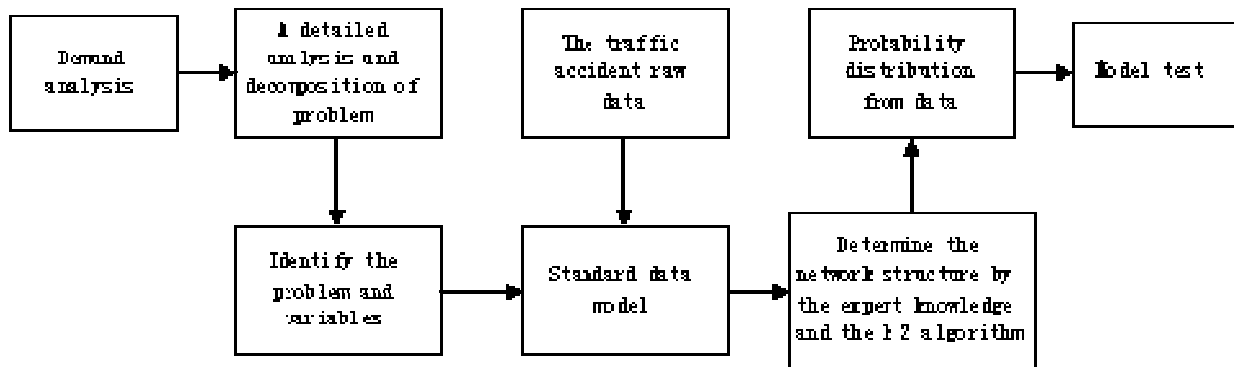


Fig.1 Bays nets modeling flow chart

K2 algorithm of Bayesian network structure ( $K2(X, order, u, data)$ )

Input:  $X = \{X_1, X_2, \dots, X_n\}$  -----a group of variables;

$order$  ----- Variable order;

$u$  ----- The upper bound of the number of parent nodes of the current node;

$data$  -----a set of complete data;

Output: a Bayesian network

```

1   $\mathcal{G} \leftarrow$  the boundless network composed by the nodes  $X_1, X_2, \dots, X_n$ 
2  for  $I=1$  to  $n$ 
3   $\pi_i \leftarrow \Phi$ ;
4   $P_{old} = g(i, \pi_i)$ 
5   $ok = true$ 
6  while  $ok$  and  $|\pi_i| < u$  do
7   $y = \arg \max_{x_j \in pre(x_i), x_j \notin \pi_i} [g(i, \pi_i \cup x_j)]$ 
8   $P_{new} = g(i, \pi_i \cup y)$ 
9  If  $P_{new} > P_{old}$ 
10  $P_{old} \leftarrow P_{new}$ ;
11  $\pi_i \leftarrow \pi_i \cup \{x_j\}$ ;
12 add  $x_j \rightarrow x_i$  in  $\mathcal{G}$ ;
13 Else
14 Breaks;
15 End if
16 End while
17 estimate the parameter  $\theta$  of  $\mathcal{G}$ 
18 Output ( $\mathcal{G}, \theta$ );
```

The essence of algorithm is to seek the father of each node, if we make it, the network structure is confirmed. The algorithm require the node is sorted in advance, assume to the node  $x_i$ ,  $pre(x_i)$  indicate the nodes in front of  $x_i$ , then the father node of  $x_i$  can only exist in  $pre(x_i)$ . After the expert knowledge, we have wiped off most impossible network structure that would quicken the running speed. Now enter a node and use the scoring function to find the father

node of  $x_i$  in  $pre(x_i)$ . K2 algorithm requires to obtain node order first, so to get the reasonable node order, first use clustering algorithm to divide the nodes into groups, in groups the similarity is high with each other. The probability of parent child relationship is larger, and then reorder the

nodes in every group by the expert experience. Once the network structure is confirmed, we would get the maximum likelihood or posteriori-probability by approximate learning algorithm, such as Monte carlo method, Gaussian approximation, gradient method and expectation maximization (EM) algorithm [8-9].

### 3. Modeling

The data model of this paper is built by the Road traffic accidents information collection project list of the ministry [1]. Through the data filtering, we remove excess information and summarize the following key attributes to analyze;

Transportation <sup>$t_0$</sup>  : 0-truck 1-car

Visibility <sup>$t_1$</sup>  : 0-within fifty meters 1-between fifty meters and one hundred meters  
2-between one hundred meters and two hundred meters 3-over two hundred meters

Lighting condition <sup>$t_2$</sup>  : 0-daytime 1-night with street lamp 2-night without street lamp

Road condition <sup>$t_3$</sup>  : 0-flat road 1-other

Weather <sup>$t_4$</sup>  : 0-fine 1-rain 2-cloudy

Traffic condition <sup>$t_5$</sup>  : 0-crowded 1-uncrowded

Horizontal curve radius <sup>$t_6$</sup>  : 0-within five hundred meters 1-between five hundred meters and six hundred meters 2-over six hundred meters

Gradient <sup>$t_7$</sup>  : 0-between zero percent and fifty percent 1-between fifty percent and a hundred percent 2-between one hundred percent and two hundred percent

Accident type A: 0-unhappen 1-happen

In addition, there are two hidden variables [10], one is pilot tensity <sup>$p^0$</sup> , and the other is road alignment reasonable degree <sup>$p^1$</sup> . Both  $p^0$  and  $p^1$  have three different values: 0-low 1-middle 2-high. According to the above analysis, the traffic accident Bayesian network topology is shown in Fig.2:

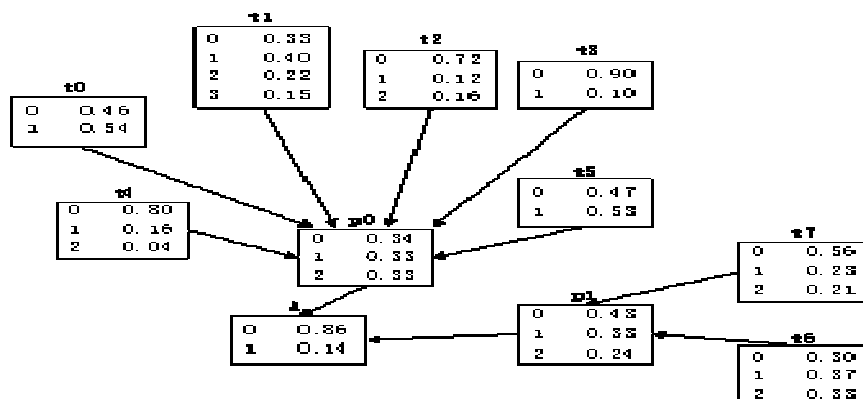


Fig.2 Bayesian network

After studying the traffic accident data, the probability of the accident is fourteen percent. As we know, the Bayesian network can perform causal inference, if the ratio of the truck rises from forty-five percent to sixty-eight percent, the probability of the accident will rise from fourteen percent to seventeen percent.

### 3. Evaluate

This paper has established the Bayesian network traffic accident analysis model, the result shows: bays nets can clearly express the accident factors, and the importance of different combination of factors. Based on this model, we not only can perform the probability prediction [11], and can carry on the accident diagnosis, find out the weakest link. As is known to all, the traffic accident is caused by various factors, the paper only gives several main factors, if other factors to join, and it will increase the precision of the model.

### Reference

- [1] Ke-jun LONG, Yong LIU, Traffic accident mechanism analysis based on Bayesian network 2010.10(1)
- [2] ShuangCheng Wang, the Bayesian network learning, reasoning and application, lixin accountant press, 2010
- [3] Jiawei Han, Micheline Kamber, Data mining Concepts and Techniques, Morgan Kaufmann Publishers, 2006
- [4] Shang Wei, The Analysis Of Multidimensional Association Rule In Traffic accident, J. 2006
- [5] Tang, Z.H., MaccLennan. The principle and application of data mining : SQL Server 2005 database, Tsinghua University Press, 2007.
- [6] Xun Liang, Data mining algorithm and the application, Peking University press, 2006
- [7] Yu Jinhua, Probability and mathematical statistics, higher education press, Zhongshan university mathematics, 2000
- [8] Pear J. Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Mateo: Morgan Kaufmann Publishers, 1988
- [9] Shu-ha WANG, Gang LIU, Zhaohui QI, BIC Scoring Bayesian Network Model and Its Application, J. 2008, 34(15)
- [10] Xiao-hu QIN, Li LIU, Ying ZHANG, A Traffic Accident Prediction Method Based on Bayesian Network Model, J. 2005, 22(11)
- [11] Yu-sheng HU, Xu-yan TU, Xiao-yu CUI, A Inferential Method of Uncertain Knowledge Based on Bayes-Network, J. 2001, 7(12)