THE USE OF BAYESIAN NETWORK IN ANALYSIS OF URBAN INTERSECTION CRASHES IN CHINA

Jinbao Zhao, Wei Deng

School of Transportation, Southeast University, China

Abstract. Traffic fatalities and injuries on urban roads especially at urban intersections constitute a growing problem in China. This study aims at researching urban intersection crashes in China and drawing conclusions by using hierarchical structured data with reference to Bayesian network (BN). On the basis of 3584 recorded crashes collected from the urban intersections of Changshu, China, a BN topological structure is developed to reflect the hierarchical characteristic of crash variables. The parameter learning process is completed with Dirichlet prior distribution. Junction tree engine is used to make inference on crash types at urban intersections with two respective given evidences, i.e. human factor and vehicle type. Parameter learning results suggest the efficacy of BN approach in the prediction accuracy. The average learned probability of illegal driving is 40.83%, which is much higher than other learned probabilities of human factors. The inferred probabilities of frontal collision at urban intersection crashes involving bicycles and electric bikes are 43.16% and 40.44% respectively, which is higher than the probabilities involving small cars and heavy vehicles. However, heavy vehicles have a higher inferred probability in side collision than light vehicles, whose inferred side collision probability is 41.02%. This study has a good potential in traffic safety discipline to reveal the correlation exists in traffic risk factors. By means of BN, researchers can make an intensive study on the hierarchical traffic crash data, determine the key risk factors and then propose corresponding and appropriate improvement measures.

Keywords: intersection crashes; Bayesian network; parameter learning; Bayesian inference; illegal driving.

Introduction

Traffic fatalities and injuries on urban roads especially at urban intersections constitute a growing problem in China. According to the statistics provided by the Ministry of Public Security of China (MPSC), 67759 people were killed and 275125 more were injured in the total of 238351 road traffic crashes in 2009 (The Annual Report… 2010). About 11.3% of these crashes occurred at urban intersections, which on average resulted in 7656 fatalities and 31089 injuries. Lots of urban traffic crash information has been collected by traffic police and large amount of recorded documents have been stored in government departments of China. However, these crash data were hardly systematically analysed, partly due to the serious illegal driving habits of Chinese drivers as well as the complicated organization of traffic flows, defectiveness of traffic safety facilities and other reasons, due to which, urban traffic crashes in China are much more serious and complicated than in many other countries.

In past years, several studies have dealt with urban traffic safety problem in China (Pei, Ma 2005; Liu *et al.* 2006; Chen, Wang 2003; Mao, Yu 2002), which all agree that this problem is so complicated and serious that detailed analysis and systematic explanations are needed. Most previous studies concentrated on using traditional accident prediction models, such as generalized linear regression model to make conceptual analyses and feature summary on urban traffic crashes, which are incapable of taking into account hierarchical traffic accident data structure. Previous researchers have considered several different variables, such as the intersection type, road grade, and human behaviour. Study results proved that these variables have notable impacts on urban traffic crashes. For example, Pei and Ma (2005) studied the relationship between the conditions and traffic crashes of urban road intersections in the cold region. Using traffic data collected from Harbin's intersections during five years, researchers stated that intersection types, the percentage of small cars and other factors could affect urban traffic crashes.

Corresponding author: Jinbao Zhao
E-mail: *jinbaozhao@aliyun.com; jinbaozhao@seu.edu.cn*

Taylor & Francis
Taylor & Francis Group

Recently, several advanced statistical techniques have been explored to more appropriately represent the nature of crash data. A number of studies have employed zero-inflated models to take into account the excess zero observations in crash data (Shankar *et al.* 2003), whereas Lord *et al.* (2007) questioned the validity of the basic zero-state assumption in these models. In this regard, Markov switching models and finite mixture models have been newly tested. Especially, the Markov switching models allowed specific road entities to switch between multiple crash states over time. More recently, a more flexible random parameter modelling approach, including random intercept and/or random slope, is emerging in the literature, in which model parameters are allowed to vary from item to site (Anastasopoulos, Mannering 2009). The surge of aggregate crash prediction models in response to the safety conscious planning have boosted the exploration of spatiotemporal models to account for the unmeasured confounders and spatiotemporal autocorrelations among adjacent geographic units.

Previous researchers generally suffered from a common limitation that they have to assume that the variables are independent of each other and the estimation procedure. However, this independence assumption may often not hold true since hierarchical data structures exist extensively because of the traffic data collection and clustering process. Disregarding the possible within-group correlations may lead to production of models with unreliable parameter estimates and statistical inferences.

Nowadays, Bayesian network (BN) constitutes a widely accepted formalism for representing uncertain knowledge and for efficiently reasoning to it. A BN is a graphical representation of a joint probability distribution, which consists of a qualitative part, a directed acyclic graph (DAG) representing conditional (in)dependence relationships, and a quantitative one, a collection of numerical parameters representing conditional probability distributions. One of the most important advantages of BNs is that the structure of the associated DAG determines the dependence and independence relationships among the variables, so that it is possible to find out, with no need of carrying out any numerical calculations, which variables are relevant or irrelevant for some other variable of interest. Early studies have shed light on the application of BNs in traffic crash analysis (Davis 2003; Trucco *et al.* 2008), but there are relatively few special researches carried on its application in urban intersection collisions. It would be worthwhile to explore the use of BN in analysis of urban intersection crashes, particularly to apply this advanced statistical technique analysing traffic crashes at urban intersections in China and improving their safety level.

The primary purpose of this study is to make preliminary hierarchical analysis and inference on urban intersection crashes, so as to raise the awareness of traffic safety analysts and stimulate the necessary concern on structured crash data. Specifically, an optimized structure of eight variables in BN is proposed to analyse the characteristics of human factors and crash types at urban intersections in China. To obtain the posterior probabilities of human factors, Bayesian method is used to conduct the parameter learning process with given parents of intersection type and environment condition as well as vehicle type. The researchers also make inference on crash types with two respective given evidences of human factor and vehicle type. Furthermore, the researchers also propose some appropriate measures to improve the safety status of urban intersections in China.

1. Methodology

1.1. Basic Description

The term Bayesian network (BN) was first stated by Pearl (1985), who formally characterized its expressive power and designed an algorithm for efficiently computing probabilities from a BN, but only for some particular network structures. Recently, BNs have become really popular tools for dealing with uncertain domains, mainly due to the formulation of efficient algorithms for computing probabilities from BNs (McCann *et al.* 2006; Malekmohammadi *et al.* 2009).

A BN is a statistical multivariate model for a set of variables $X = \{X_1, X_2, ..., X_n\}$, which is defined in terms of two components:

– qualitative component: a directed acyclic graph (DAG) where each vertex represents one of the variables in the model, and so that the presence of an edge linking two variables indicates the existence of statistical dependence between them;

– quantitative component: a conditional distribution $p\left(x_i \big| pa(x_i)\right)$ for each variable X_i, $i = 1, 2, ..., n$ given its parents in the graph, denoted as $pa(x_i)$.

The graph depicted in Fig. 1 could be the qualitative component of a BN for variables $X_1, ..., X_5$. According to the structure of the graph, it would be necessary to specify a conditional distribution for each variable given its parents. In this case, the distributions are $p(x_1)$, $p(x_2|x_1)$, $p(x_3|x_1)$, $p(x_4|x_2, x_3)$ and $p(x_5|x_2)$.

In practice, the analysis process of BN mainly involved three phases:

1. structure learning phase, that is to identify the quantitative topology of BN;
2. parameters learning phase, for calculating the numerical parameters (posterior probabilities) for a given network topology;
3. Bayesian inference phase, to obtain the marginal posterior distribution of all unknowns, and then integrate this distribution over the unknowns to obtain the desired marginal distribution.
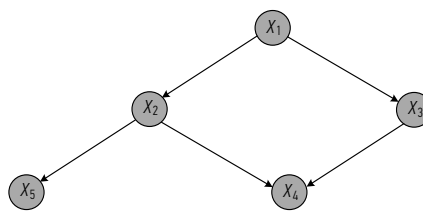

Fig. 1. A sample BN with five variables

1.2. Structure Learning

There is a lot of work devoted to the BN structure learning for classification purposes. However, it is known that this computation is NP-hard problem (Bouchaala *et al.* 2010). Searching for the best structure is difficult because the search space increases exponentially with the number of variables. In order to establish the BN structure, it can be the human expert who designs the network taking advantage of his/her knowledge about the relations among the variables; it is also possible to learn the structure by means of an automatic learning algorithm; and, of course, a combination of both systems can be applied, mixing the expert's knowledge and the learning mechanism. In our case, we firstly use the *K2* algorithm (Cooper, Herskovits 1992) to determine the initial structure of the BN from the database, and then improve the initial structure to obtain the optimized structure on the basic of expert's knowledge.

    K2 is an algorithm that creates and evaluates a BN from a dataset of cases. The *K2* algorithm assumes that an ordering on the variables is available and that, a priori, all structures are equally alike. Fig. 2 shows the algorithm in pseudo-code form. Given a dataset *D*, *K2* searches the BN structure B_{S*} with the maximal $P(B_S|D)$:

$$B_{S*} = \arg\max_S P(B_S|D), \qquad (1)$$

where: $P(B_S|D) = P(B_S)\prod_{i=1}^{n}\prod_{j=1}^{q_i}\dfrac{(r_i-1)!}{(N_{ij}+r_i-1)}\prod_{k=1}^{r_i}N_{ijk}$, referred the *K2* metric.

    As a greedy heuristic, *K2* algorithm starts by assuming that a node does not have parents, and then, in every step, it incrementally adds the parent whose addition most increases the probability of the resulting structure. However, only those nodes that correspond to variables that precede the node being processed are considered as candidates. *K2* stops adding parents to the nodes when the addition of a single parent cannot increase the probability. This approach does not guarantee the selection of a structure with the highest probability, so, mixing the expert's knowledge to optimize the learning result is necessary.

1.3. Parameter Learning

In a given structure of BN topology, two methods are commonly used for parameter learning: one is maximum likelihood estimation (MLE) method and another is Bayesian method. In this paper, Bayesian method is chosen to complete the parameter learning process. Suppose that all the parameters in the BN have prior distributions and there exists a complete collection of discrete dataset *D*. x_i is a random variable whose prior distribution is $p(x_i)$. $p(x_i|D)$ represents the posterior probability of x_i. The aim of parameter learning is to calculate the posterior probability of x_i as to seek for the biggest value for vector x_i based on the complete dataset *D*. $p(x_i)$ is usually assumed to be Dirichlet distribution (Maceachern 1994; Friedman, Koller 2003), for this distribution's conjugate feature can reduce the computing complexity, and its distribution form as shown as follows:

$$p(x_i) = Dir(x|\alpha_1,...,\alpha_n) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{n}\Gamma(\alpha_i)}\prod_{i=1}^{n}x_i^{\alpha_i-1}, \quad (2)$$

where: $\alpha = \sum_{i=1}^{n}\alpha_i$, $\alpha_i > 0$, $i = 1,2,...,n$; $\alpha_1,...,\alpha_n$ are Hyper parameters, and $\Gamma(\alpha)$ is the Gamma function.

    Occurrence probability of the sample data can be described as:

$$p(D) = \int p(x)p(D|x_i)dx_i =$$
$$\int \frac{\Gamma(\alpha)}{\prod_{i=1}^{n}\Gamma(\alpha_i)}\prod_{i=1}^{n}x_i^{\alpha_i-1}\prod_{i=1}^{n}x_i^{N_i}dx =$$
$$\frac{\Gamma(\alpha)}{\Gamma(\alpha+N)}\prod_{i=1}^{n}\frac{\Gamma(\alpha_i+N_i)}{\Gamma(\alpha_i)}, \qquad (3)$$

where: N_i is the number of times that father node has the *i*-th value, $N = \sum_{i=1}^{n}N_i$.

    Then the posterior probability of x_i is formulated as:

$$p(x_i|D_1+N_1,...,\alpha_n+N_n) =$$
$$\frac{p(x_i)p(D|x_i)}{p(D)} =$$
$$\frac{\Gamma(\alpha+N)}{\prod_{i=1}^{n}\Gamma(\alpha_i+N_i)}\prod_{i=1}^{n}x_i^{\alpha_i+N_i-1} =$$
$$Dir(x_i|\alpha_1+N_1,...,\alpha_n+N_n). \qquad (4)$$

```
Input: Ordering, Database+metric.
Output: net struture
for i=1 to n do
    π_i = ∅
    p_old = g(i, π_i)
    while ok_to_proceed and |π_i| < max_parents
    do
        z = arg max_k g(i, π_i ∪ {k}) where z ∈ pred(i)
        p_new = g(i, π_i ∪ {z})
        if p_new > p_old then
            p_old = p_new
            π_i = π_i ∪ z
        else
            ok_to_proceed = false
        end if
    end while
end for
```

Fig. 2. Pseudo-code of the *K2* structural learning algorithm

1.4. Bayesian Inference

As frequently used in theoretical studies and applications, the junction tree algorithm (Madsen, Jensen 1999; Huang, Darwiche 1996; Helman *et al.* 2004) is a machine learning method for exact marginalization in general graphs. In essence, it entails performing belief propagation on a modified graph. The basic premise is to eliminate cycles by clustering them into single node. In our case, junction tree algorithm is to complete the Bayesian inference process for crash types, which can be directly called in the Bayes net toolbox for *Matlab* (Murphy 2001), so we did not present the process in detail here.

2. Data

Database utilized for the presented study is collected from 3584 recorded crashes that occurred at urban intersections in Changshu, China. Before starting the process of BN learning and inference, we have to prepare the data in such way that the learning algorithm can deal with them. BN can deal with both discrete data and continuous data. However, if using continuous data, BN needs to calculate many integrals, which increases the computing complexity. Discrete data for categorical variables can build models capturing 'non-linear' relationships between variables. Furthermore, the discrete data also get rid of the restricting distributional assumptions (like multivariate normality assumptions prevalent in current statistical practice). Thus the advantage is making less assumptions and the possibility to find more complex relationships. Therefore, all the value of variables in the database would be discretized in our case.

Of the total 17 variables coded for each crash in the database, a number of them, like crash code, intersection code, width of crash location, etc., are excluded as they are irrelevant to the analytical purpose. A total of 8 covariates in the database are selected, i.e. weather condition, time of day, lighting condition, environment condition, intersection type, vehicle type, human factor and crash type. The descriptions of the selected variables, together with their data discretized results, are presented in Table 1.

3. Result Analysis

3.1. Structure Learning Results

As stated, one of the advantages of BN is the possibility to incorporate knowledge to the model, in such way that experts can incorporate their knowledge to the learned network. The researchers complete the program coding process in the Bayes net toolbox for *Matlab* (Murphy 2001) and utilize the mentioned $K2$ algorithm to obtain the initial structure of BN learned from the database. Fig. 3a shows the learned structure. As $K2$ algorithm does not guarantee the selection of a structure with the optimal result (Bouchaala *et al.* 2010), so our research team together with two experts in traffic safety filed checking this initial learned structure and discover some inaccuracies which conflicted with the practice:

Table 1. Utilized variables and their discretized results

Variables	Description	Discre-tized
Weather condition (WC)	Sunny day	1
	Cloudy day	2
	Rainy day	3
Time of day (TD)	06:00–24:00	1
	0:00–06:00	2
Lighting condition (LC)	Daytime	1
	Nighttime with lighting	2
	Nighttime without lighting	3
Environment condition (EC)	Appropriate	1
	Otherwise	2
Intersection type (IT)	*X* intersection	1
	T/Y intersection	2
	Round intersection	4
Vehicle type (VT)	Small car	1
	Motorcycle	2
	Heavy vehicles (bus, lorry, truck, etc.)	3
	Electric bike	4
	Bicycle	5
	Others (tricycle, tractor, etc.)	6
Human factor (HF)	Illegal driving	1
	Misjudgment	2
	Brake failure	3
	Improper steering	4
	Others (drunk, drug, etc.)	5
Crash type (CT)	Frontal collision	1
	Side collision	2
	Rear-end collision	3
	Scrape	4
	Others (rollover, explosion, etc.)	5

– Intersection type is the parent node of lighting condition. It is easily to understand that intersection type affected human. However, lighting condition is hardly influenced by intersection type. Furthermore, intersection type can affect crash type as well as vehicle type.
– Vehicle type has nothing to do with human factor. Actually, drivers are affected obviously by vehicle types. Different vehicle types provide different driving environments for vehicle drivers and vehicle drivers reflect differently to these different environments, so vehicle type may be the parent node of human factor.
– Weather condition directly influences intersection type, which is inaccurate in practice. Weather condition can affect environment condition and then indirectly influence human through en-
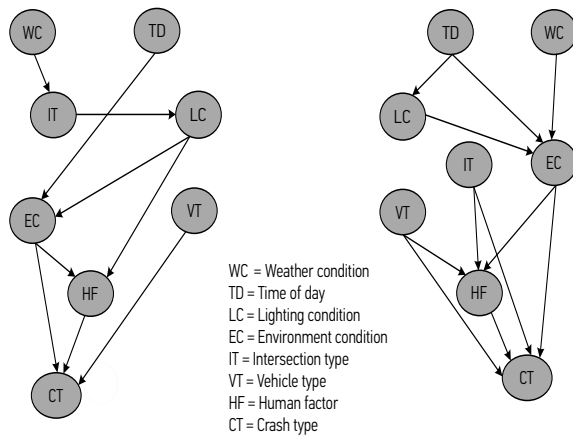
Fig. 3. Learned and optimized BN structures of urban intersection crash variables: a – learned structure; b – optimized structure

WC = Weather condition
TD = Time of day
LC = Lighting condition
EC = Environment condition
IT = Intersection type
VT = Vehicle type
HF = Human factor
CT = Crash type

vironment condition. So, in our case, the weather condition directly influences environment condition rather than intersection type.

By analysing the inaccurate points and mixing expert knowledge, the optimized structure obtained as shown in Fig. 3b. The optimized structure retains the reasonable parts of learned results from database and integrates rational improvement with expert analysis. Generally, traffic crashes are complex events comprehensively resulted from human, vehicle, road and environment contributing factors, this fact is reflected in the optimized structure. As a foremost contributing factor, human is influenced by many external conditions, such as vehicle type, intersection type, weather condition and environment condition, which affect human directly or indirectly.

The optimized result is a hierarchical structure, which can estimate the probability distributions, risk and uncertainty more accurately than the models which only take mean values into account. Once the structure is learned, the probability distribution of a node given its parents is obtained, and even the other way round, the probability distribution of a parent node given its child nodes can also be obtained, which allows us to know the effects given the causes and the causes given the effects. BN is able to incorporate expert knowledge via a participatory modelling procedure, since the relations between the variables can be visualized easily through the graphical representation of the network, and so they can be modified by the experts or stakeholders just by adding or removing variables and links in the graph. This advantage makes them also easier to understand and visualize by final users.

3.2. Parameter Learning Results

Once the graph structure is learned, the probability distributions are obtained by estimating the corresponding parameters. Table 2 shows the posterior probabilities (parameter learning results) of human factor with three given parents, i.e. intersection type, environment condition and vehicle type, with the intersection type is *X* intersection.

Fig. 4 shows the absolute error distributions of posterior probabilities. The maximum absolute error is $2.67 \cdot 10^{-4}$, occurring at coding 29, i.e. intersection type = '1', environment condition = '1', vehicle type = '5' and human factor = '3'. The total absolute error of these 60 posterior probabilities is $1.91 \cdot 10^{-3}$, while the mean absolute error is less than $3.2 \cdot 10^{-5}$. Absolute error analysis suggested the efficacy of BN approach in the network prediction accuracy. BN models can readily represent complex interrelated stochastic processes, scale well, and are not dependent on limit arguments for their justification. Furthermore, use of BN permits us to draw direct inferences regarding posterior probabilities, and grounds our inferential framework on an axiomatic basis.

Table 2. Posterior probabilities of human factors

Intersection type	Environment condition	Vehicle type	Human Factor				
			1	2	3	4	5
1	1	1	0.4278 (1)	0.2252 (13)	0.1652 (25)	0.1296 (37)	0.0522 (49)
1	1	2	0.4494 (2)	0.2595 (14)	0.1519 (26)	0.0823 (38)	0.0570 (50)
1	1	3	0.4153 (3)	0.2146 (15)	0.1772 (27)	0.1299 (39)	0.0630 (51)
1	1	4	0.4486 (4)	0.2288 (16)	0.1347 (28)	0.1241 (40)	0.0638 (52)
1	1	5	0.3948 (5)	0.2106 (17)	0.2105 (29)	0.1053 (41)	0.0789 (53)
1	1	6	0.3571 (6)	0.2619 (18)	0.1905 (30)	0.1429 (42)	0.0476 (54)
1	2	1	0.4648 (7)	0.2278 (19)	0.1489 (31)	0.1037 (43)	0.0549 (55)
1	2	2	0.3942 (8)	0.2211 (20)	0.1410 (32)	0.1602 (44)	0.0833 (56)
1	2	3	0.4043 (9)	0.3522 (21)	0.1247 (33)	0.0638 (45)	0.0551 (57)
1	2	4	0.4362 (10)	0.2093 (22)	0.1655 (34)	0.1230 (46)	0.0662 (58)
1	2	5	0.3506 (11)	0.3402 (23)	0.1092 (35)	0.1003 (47)	0.0999 (59)
1	2	6	0.3572 (12)	0.3568 (24)	0.0480 (36)	0.1905 (48)	0.0476 (60)

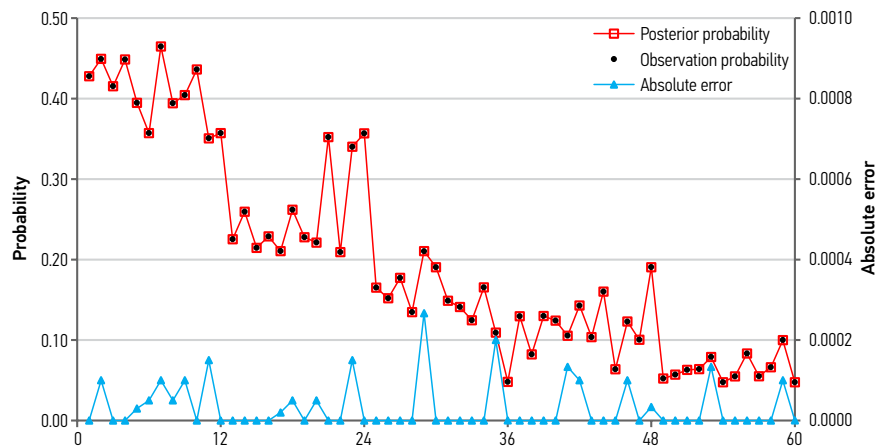*Note*: The numbers in brackets represented the probability coding in Fig. 4.

Fig. 4. Absolute error distributions of posterior probabilities at *X* intersection

As shown in Fig. 5, with the parents of *X* intersection type and environment condition as well as vehicle type, the average posterior probability of illegal driving is 40.83%, much higher than other human factors'. Running a red light, light vehicles' reverse drive, poor lane discipline, bad tailgating, not indicating and undertaking are just a few of the bad habits occurring frequently at China's urban intersections. These phenomena are even more accustomed in the medium and small cities in China, where the regulatory on traffic operation is frail. Aside from the inconvenience to the drivers themselves and other road users, this kind of inconsiderate driving is also very dangerous and easily results in crashes.

3.3. Bayesian Inference Results

Bayesian analysis is a process of fitting a probability model to the database and summarizing the posterior probability distribution on model parameters and on unobserved quantities. Instead of producing MLE for unknowns totally based on the sample data, Bayesian methods explicitly use the probability for quantifying uncertainty in inferences based on the statistical data analysis. In Bayesian models, given model assumptions and parameters, the likelihood of the observed data is used to modify the prior beliefs of the unknowns, resulting in the updated knowledge in the form of posterior distributions. This process is also called Bayesian inference.

The inference results of crash types with given evidence of human factor indicate that illegal driving is the most possible reason results for frontal collision. As shown in Fig. 6, its inferred probability reaches up to 42.48%. The inferred probability of side collision caused by improper steering is 45.20%, which indicates that steer properly at intersections can efficaciously reduce the occurrences of side collisions. Brake failure is the most possible reason resulting in rear-end collisions, whose inferred probability is 40.06%, obviously higher than other human factors', this fact indicated that braking timely to a danger could efficaciously avoid the occurrence of rear-end collisions at the intersections.

As shown in Fig. 7, to any vehicle types, frontal collision and side collision are two main forms of traffic crashes at urban intersections. The probabilities of frontal collision in traffic crashes participated respectively by bicycle and electric bike are 43.16% and 40.44%, higher than the probabilities participated respectively by small cars and heavy vehicles, which are 31.35% and 30.42%.
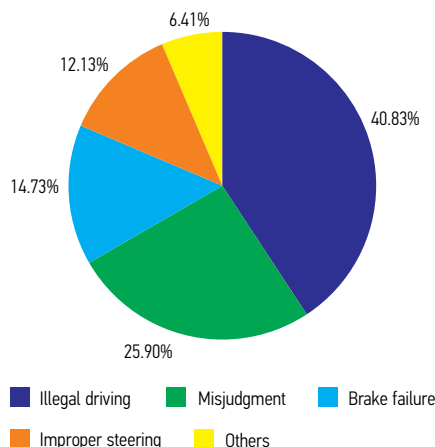

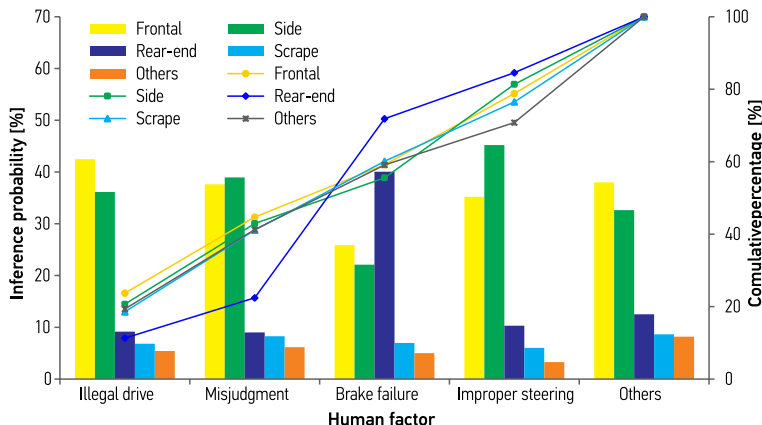Fig. 5. Average posterior probabilities of human factors

Fig. 6. Inference results of crash types (evidence: human factor)
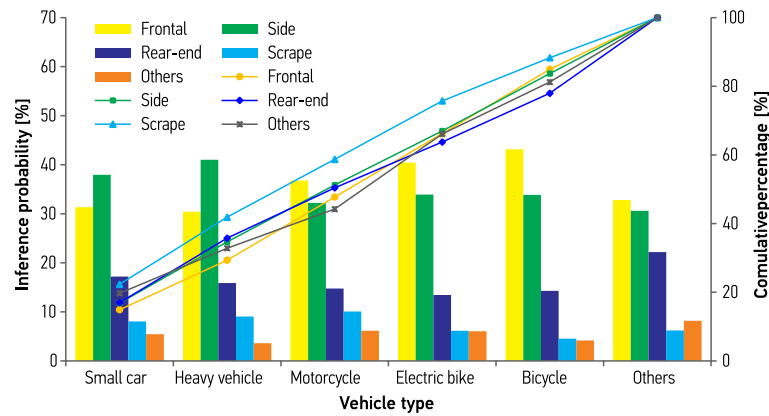
Fig. 7. Inference results of crash types (evidence: vehicle type)

Many light vehicle (bicycle and electric bike) drivers drive reversely at urban intersections in China, which easily lead to the occurrences of frontal collisions. Furthermore, light vehicles have relatively light weights and their momentum and kinetic energy are relatively small, frontal collision is the main collision form that can result in these vehicles' crashes. The inferred probability of side collision occurred on heavy vehicles is 41.02%, higher than other vehicle types, partly due to the fact that heavy vehicle drivers usually have an inaccurate aesthesis to their side vision as the huge body type of these heavy vehicle makes the side aesthesis difficult for the drivers.

4. Further Discussions

Usually, Chinese traffic appears to have no rules or, if there are rules, it appears they are neither followed nor enforced. There is absolutely no point getting angry if someone cuts you off or drives against the red light or on the wrong side of the road. Fig. 3 has shown the learned probability of illegal driving at intersections (e.g. running a red light, light vehicles' reverse drive, etc.) is much higher than other human factors' learned probabilities in China. As statistics from our research observation at 12 intersections in Chuangshu and Nanjing reports, there is an average of 21.5 illegal driving events per hour per intersection.

A questionnaire composed by two parts carried out by our research team to study drivers' attitudes to illegal driving and some of their diver behaviours during past one year in Changshu. In the first part of the questionnaire, subjects are asked to indicate on a five-point scale (strongly disagree = 1, disagree = 2, indifferent = 3, agree = 4, strongly agree = 5). Subjects of the second part are also based on a five-point scale (never = 1, seldom = 2, sometimes = 3, often = 4, very often = 5). A sample of total 296 respondents is average 35.5 years old with 33.9% between 25 and 45 years old, and there are 19.6% females and 80.4% males.

Table 3 shows the diver attitudes to illegal driving violations at intersections with five questions whose mean values are 'the smaller the better'. Of the five violations, diver attitudes about 'average' violations (means ranging from 3.25 to 3.98; average mean = 3.82), the

highest is 'If there are no punishments, I can drive illegally'. 34.8% of total 296 respondents held 'agree' attitude to pass though intersections illegally if there are no punishments, only 12.5% of them held 'strongly disagree' attitude. Many cities in China have any road safety cameras or traffic monitors at intersections, to say nothing of being covered with a comprehensive monitoring system. With consideration of complexities compounded by the slow progress of passing road safety rules and regulations, diver attitudes to illegal driving at intersections and the bureaucracy in China's governments, appropriate punishment measures and monitoring systems at urban main intersections should be implemented. There are 32.4% of the respondents hold 'agree' attitudes on the wrong point of view that if it is safety, they can run a red light or do other illegal driving behaviours, which indicates that it is urgent to spread traffic safety education among China's divers, so as to improve their awareness on treasuring their own as well as other people's life and property.

The car-pedestrian interactions are very complicated in China. To car drivers, the general rule appears to be 'keep moving no matter what'. Cutting people off, swerving into the oncoming lane, driving on the shoul-

Table 3. Diver attitudes to diver attitudes to illegal driving violations

Drive violations	Total sample (n = 296)			The bigger the better
	Range	Mean	SD	
If there are no punishments, I can drive illegally	1…5	3.98	1.32	No
Motor vehicle drivers needn't to evade pedestrians	1…5	3.25	1.41	No
If it is safe, I can drive illegally	1…5	3.91	0.96	No
To rush through, I can drive illegally	1…5	3.78	1.27	No
If it is urgent, I can drive illegally	1…5	3.80	1.13	No

der, or in a fenced-off bicycle lane, or the wrong way down a divided urban road are all fine as long as they keep you moving in the right general direction and do not cause an immediate crash. Table 4 shows the respondents' driving behaviours during past one year. Of the five behaviours, the mean values of the first three behaviours are assessed by 'the bigger the better', while the last two mean values are assessed by 'the smaller the better'. Of the first three behaviours, evaded pedestrians when there are no signal no zebra crossing had the worst performance. Only 10.5% of all respondents 'often' evaded pedestrians when there are no signal and no zebra crossing. There are many unsignalized intersections in both mega-city and medium-city in China. Generally, unsignalized intersections should be operated under basic rules of the road (no control other than warning and guide signs) or under STOP or YIELD control. However, many unsignalized intersections in China have any STOP or YIELD control signs, which encourage drivers' illegal diving behaviours. There are 46.3% of all respondents 'sometimes' run red light at signalized intersections, only 8.9% of them 'never' run a red light. In general, bad alignments, slow improvement in guardrail design and drawback in intersection signs have resulted in divers' serious illegal driving behaviours in China, which further result in China's urban intersection crashes much more serious and complicated than many other countries'.

Another serious problem is the reverse drive of light vehicles at intersections in China, which easily results in light vehicles' frontal collisions. China is a famous bicycle country and many urban residents choose bicycle as their travel model, however, bicycles and sometimes cars ignore one-way signs. At intersections on divided urban roads, seeing pedestrians, bicycles and electric bikes going the wrong way down the shoulder is entirely normal, and a few go the wrong way beside the centre fence. At traffic circles (roundabouts), drivers hate going around the island in the middle if they can avoid it; they will often just swing left instead. Lane markings are also routinely ignored. A typical area in

Table 4. Driver behaviors during past one year

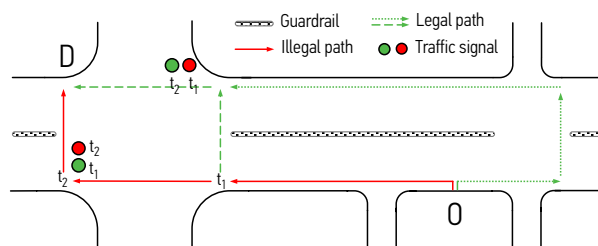| Driver behaviors	Total Sample (n = 296)			The bigger the better
	Range	Mean	SD	
Evaded pedestrians when there are no signal no zebra crossing	1…5	1.62	1.37	Yes
Slowed down when there are no signal no pedestrians	1…5	2.75	1.26	Yes
Evaded major road vehicles while on minor road	1…5	1.97	1.08	Yes
Ran a red light	1…5	2.84	0.92	No
Cut in queue in traffic jam	1…5	2.69	1.14	No

Fig. 8. Schematic drawing of a study area
(note: in mainland China, traffic drives should be on the right-hand side of the road)

Changshu is selected to explain the seriousness of this problem, which is illustrated in Fig. 8. The main intersection in this area did not have left-turn waiting area for light vehicles. Starting from the origin point O to the destination point D, the legal path are the green dotted lines, in reality, however, many light vehicle drivers tend to choose the red solid lines as their driving paths. An investigation on light vehicle drivers (total sample is 350, with 38% females and 62% males) carried out by our research team shows that up to 85.7% of them would like to select the red solid lines as their driving path under the illustrated condition. 73.3% of the illegal path choosers select this illegal path on the basis of saving time and 18.0% of them select this illegal path because they would not like to detour so as to save their physical strength, while the others just follow the crowd psychology of 'others can do it , so can I'.

5. Recommendations

During the repaid urbanization process, China still faces a serious urban traffic safety situation and experiences a high incidence of urban traffic crashes. Urban intersections are usually the black-points of road traffic crashes. Enhancing urban intersection safety level is essential to realizing, maintaining and developing the fundamental interests of urban public. To improve the safety level of China's urban intersections, some suggestions were recommended here on the basic of this study and our understanding.

Illegal driving is very common in China and urban road users in China are insufficiently aware of the importance of traffic safety and the governance by law. It may take a significant amount of time to improve their awareness and behaviour. There is no doubt that urban road users need to improve their behaviour in certain areas, such as running red lights, illegal line change and other conduct which severely affects urban intersection safety. China's government should also make efforts on establishing a society-based education and publicity system on urban traffic safety to enhance and improve the nation's awareness of urban traffic safety.

Except paying more attention on urban traffic safety education, efforts should also be made to build up an effective urban road monitoring system to reduce illegal driving behaviours on urban roads especially at urban intersections.
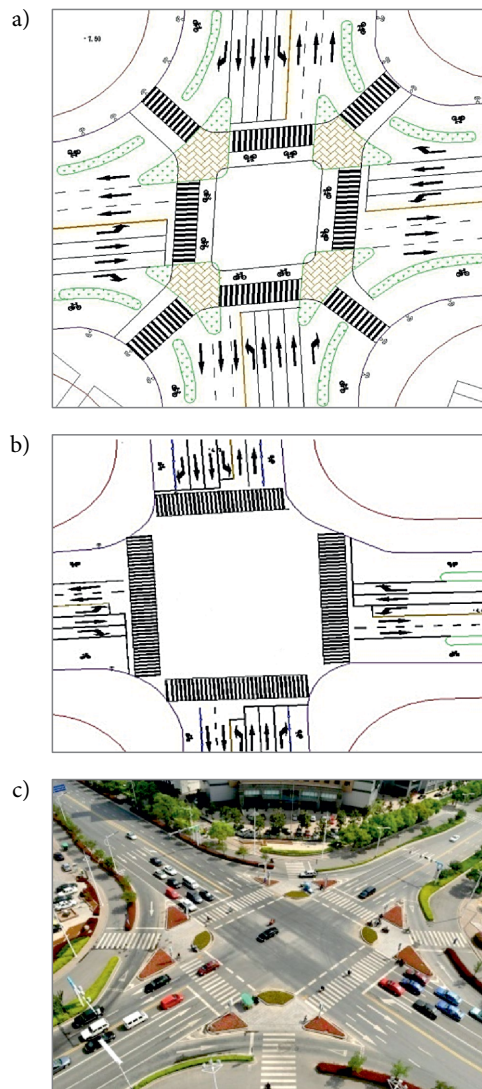
Fig. 9. The before and after of optimization of *X* intersection:
a – CAD drawing before optimization; b – CAD drawing
after optimization; c – actual optimization result

Optimize the urban road infrastructural facilities to be friendlier to bicycles and pedestrians. In China, light vehicle drivers and pedestrians are usually from low and middle income families, and they are usually considered to be vulnerable road users. To improve their safety status, second-crossing safety island (as shown in Fig. 9) for bicycles and pedestrians should be constructed at urban intersections. Left-turn waiting area for light vehicle drivers should be framed and pedestrian footbridge or tunnel should be constructed at the intersections where there are heavy volumes of light vehicles and pedestrians.

Conclusions

The study presented in this paper develops a BN structure to analyse the hierarchical characteristic of urban intersection crash variables in China. Based on the results of structure learning, parameter learning, Bayesian inference and related discusses, the following conclusions can be made:

– The risk factors in traffic safety represent a characteristic of hierarchical structure, that is, one factor in a structure may influence other factors and be influenced by the others. One of the most important advantages of BN is that the structure of the associated DAG determines the dependence and independence relationships among the variables, so it is possible to reflect the hierarchical characteristic of risk factors in traffic safety problems.
– Absolute error analysis shows that if the training data is complete, parameter learning results will be highly precise. BN models can readily represent complex interrelated stochastic processes, scale well, and are not dependent on limit arguments for their justification. Furthermore, use of BN permits us to draw direct inferences regarding posterior probabilities, and grounds our inferential framework on an axiomatic basis.
– Bayesian inference results indicate that illegal driving is the most possible reason resulted in frontal collision at the urban intersections in China and its inferred probability reached up to 42.48%. The inferred probability of side collision caused by improper steering is 45.20%, while brake failure is the most possible reason resulted in rear-end collision. The probabilities of frontal collision in traffic crashes participated respectively by bicycle and electric bike are 43.16% and 40.44%, higher than the probabilities participated respectively by small cars and heavy vehicles, but heavy vehicles had a higher inferred probability in side collision than light vehicles, whose inferred side collision probability is 41.02%.
– Illegal driving has become the most risk human factor leading to traffic fatalities and injuries at urban intersections in China. Since 50,063 people are killed and 185,785 more are injured in road traffic crashes in 1986 (MPSC, 1987), China have holed the world's first status in road traffic death and injure for more than twenty consecutive years. Urban intersections are usually the black-points of road traffic crashes. If China would like to have a record-breaking in reducing the road traffic fatalities and injuries, the local authorities should spare no efforts to improve the safety level of urban intersections in the near future.

References

Anastasopoulos, P. C.; Mannering, F. 2009. A note on modeling vehicle accident frequencies with random-parameters count models, *Accident Analysis & Prevention* 41(1): 153–159. http://dx.doi.org/10.1016/j.aap.2008.10.005

Bouchaala, L.; Masmoudi, A.; Gargouri, F.; Rebai, A. 2010. Improving algorithms for structure learning in Bayesian networks using a new implicit score, *Expert Systems with Applications* 37(7): 5470–5475.
http://dx.doi.org/10.1016/j.eswa.2010.02.065

Chen, K.-M.; Wang, Y.-P. 2003. Distribution characteristics and countermeasures of urban traffic accidents, *Journal of Traffic and Transportation Engineering* 3(1): 84–87.

Cooper, G. F.; Herskovits, E. 1992. A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* 9(4): 309–347.
http://dx.doi.org/10.1007/BF00994110

Davis, G. A. 2003. Bayesian reconstruction of traffic accidents, *Law, Probability and Risk* 2(2): 69–89.
http://dx.doi.org/10.1093/lpr/2.2.69

Friedman N.; Koller, D. 2003. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks, *Machine Learning* 50(1–2): 95–125.
http://dx.doi.org/10.1023/A:1020249912095

Helman, P.; Veroff, R.; Atlas, S. R.; Willman, C. 2004. A Bayesian network classification methodology for gene expression data, *Journal of Computational Biology* 11(4): 581–615.
http://dx.doi.org/10.1089/cmb.2004.11.581

Huang, C.; Darwiche, A. 1996. Inference in belief networks: a procedural guide, *International Journal of Approximate Reasoning* 15(3): 225–263.
http://dx.doi.org/10.1016/S0888-613X(96)00069-2

Liu, Q.; Lu, H.-P.; Zhang, Y.-B.; Zuo, B. 2006. Characteristic analysis and countermeasure study on road traffic accidents in China, *China Safety Science Journal* 16(6): 123–128.

Lord, D.; Washington, S.; Ivan, J. N. 2007. Further notes on the application of zero-inflated models in highway safety, *Accident Analysis & Prevention* 39(1): 53–57.
http://dx.doi.org/10.1016/j.aap.2006.06.004

Maceachern, S. N. 1994. Estimating normal means with a conjugate style Dirichlet process prior, *Communications in Statistics – Simulation and Computation* 23(3): 727–741.
http://dx.doi.org/10.1080/03610919408813196

Madsen, A. L.; Jensen, F. V. 1999. Lazy propagation: a junction tree inference algorithm based on lazy evaluation, *Artificial Intelligence* 113(1–2): 203–245.
http://dx.doi.org/10.1016/S0004-3702(99)00062-4

Malekmohammadi, B.; Kerachian, R.; Zahraie, B. 2009. Developing monthly operating rules for a cascade system of reservoirs: application of Bayesian networks, *Environmental Modelling & Software* 24(12): 1420–1432.
http://dx.doi.org/10.1016/j.envsoft.2009.06.008

Mao, M.; Yu, X. 2002. Analysis of traffic accident causation, *Journal of Highway and Transportation Research and Development* 19(5): 125–127.

McCann, R. K.; Marcot, B. G.; Ellis, R. 2006. Bayesian belief networks: applications in ecology and natural resource management, *Canadian Journal of Forest Research* 36(12): 3053–3062. http://dx.doi.org/10.1139/x06-238

Murphy, K. P. 2001. The Bayes net toolbox for Matlab, in *Computing Science and Statistics Vol. 33: Proceedings of the 33rd Symposium on the Interface*, 13–16 June 2001. Costa Mesa, California, 20 p. Available from Internet: http://www.interfacesymposia.org/I01/I2001Proceedings/KMurphy/KMurphy.pdf

Pearl, J. 1985. A constraint propagation approach to probabilistic reasoning, in *Proceedings of the First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-85)*, 14–16 August 1985, Corvallis, Oregon, 31–42.

Pei, Y.-L.; Ma, Y.-L. 2005. The relationship between condition and traffic accidents of urban road intersections in the cold region, *Journal of Harbin Institute of Technology* 37(l): 130–133.

Shankar, V. N.; Ulfarsson, G. F.; Pendyala, R. M.; Neberagall, M. L. B. 2003. Modeling crashes involving pedestrians and motorized traffic, *Safety Science* 41(7): 627–640.
http://dx.doi.org/10.1016/S0925-7535(02)00017-6

The Annual Report for Road Traffic Accident. 1987. The Ministry of Public Security of the People's Republic of China, Beijing, China.

The Annual Report for Road Traffic Accident. 2010. The Ministry of Public Security of the People's Republic of China, Beijing, China.

Trucco, P.; Cagno, E.; Ruggeri, F.; Grande, O. 2008. A Bayesian Belief Network modelling of organisational factors in risk analysis: a case study in maritime transportation, *Reliability Engineering & System Safety* 93(6): 845–856.
http://dx.doi.org/10.1016/j.ress.2007.03.035