

Research Article

Prediction for Traffic Accident Severity: Comparing the Bayesian Network and Regression Models

Fang Zong, Hongguo Xu, and Huiyong Zhang

College of Transportation, Jilin University, 5988 Renmin Street, Changchun, Jilin 130022, China

Correspondence should be addressed to Huiyong Zhang; jlu.zf@163.com

Received 29 August 2013; Accepted 22 September 2013

Academic Editor: Baozhen Yao

Copyright © 2013 Fang Zong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper presents a comparison between two modeling techniques, Bayesian network and Regression models, by employing them in accident severity analysis. Three severity indicators, that is, number of fatalities, number of injuries and property damage, are investigated with the two methods, and the major contribution factors and their effects are identified. The results indicate that the goodness of fit of Bayesian network is higher than that of Regression models in accident severity modeling. This finding facilitates the improvement of accuracy for accident severity prediction. Study results can be applied to the prediction of accident severity, which is one of the essential steps in accident management process. By recognizing the key influences, this research also provides suggestions for government to take effective measures to reduce accident impacts and improve traffic safety.

1. Introduction

As a significant cause of deaths, injuries, and property loss, traffic accident is a major concern for public health and traffic safety. According to statistics from the Ministry of Public Security of China between 2009 and 2011, traffic crashes resulted in an average of 65 123 people dead and 255 540 cases injured annually in China (China Statistical Yearbook of Road Traffic Accidents, 2009–2011). It was reported that the cost of medical care and productivity losses associated with motor vehicle crash injuries was over \$99 billion, or nearly \$500, for each licensed driver in the United States (Centers for Disease Control and Prevention, 2010). Being one of the major steps of accident management, accident severity prediction can provide crucial information for emergency responders to evaluate the severity level of accidents, estimate the potential impacts, and implement efficient accident management procedures.

In recent years, increased attention has been directed at accident severity prediction, for which Bayesian network and Regression model are two widely used modeling techniques. However, to the authors' knowledge, there is no study that presents quantitative comparison of the two methods. Therefore, the present work focuses on conducting an accident severity modeling by employing both Bayesian network

and Regression model. The accuracies of the two methods will then be compared and a better one will be selected for accident severity prediction. By carrying out accident severity analysis, the risk factors and their effects will also be identified in the work.

The remainder of this paper is organized as follows. In Section 2, the literature review on predictions of severity is presented in general. The data are described in Section 3. This is followed by accident severity modeling with Bayesian network and Regression models in Sections 4 and 5, respectively. The paper concludes with a summary and directions for future research.

2. Literature Review

Regression analysis has been widely used to accident severity prediction and contributing factors determination. The most commonly used Regression models are Logistic Regression model and Ordered Probit model [1–6]. However, some researchers [7, 8] pointed out that most of the Regression models have their own assumptions and predefined underlying relationships between dependent and independent variables (i.e., linear relations between the variables).

If these assumptions are violated, the model could lead to erroneous estimations of the likelihood of severe injury.

Some researchers carried out traffic accident analysis by employing Bayesian network. For instance, de Oña et al. [8] applied Bayesian network to the identification of the factors affecting injury severity, which was classified into slightly injured and killed/severely injured. Based on Bayesian network, the factors associated with a killed/severely injured accident were identified to be accident type, driver age, lighting, and number of injuries. The results indicate that Bayesian network is capable of making predictions without the need for preassumptions and can make graphic representations of complex systems with interrelated components. Simoncic [9] constructed a Bayesian network for analysis of injury severity. The results showed that Bayesian network can be applied in road accident modeling. It also presented some advantages of Bayesian network, such as being able to involve more variables and larger data set than Regression model. Ozbay and Noyan [10]'s work constructed a Bayesian network and applied it to predicting incident duration and understanding factors associated with incident clearance time. The results indicated that Bayesian network can represent the stochastic nature of incidents. Gregoriades [11] highlighted the interest of using Bayesian network to model traffic accidents and discussed the need to not consider traffic accidents as a deterministic assessment problem, but model the impacts of the factors that could lead to traffic accidents.

Although previous works presented the advantages of adopting Bayesian network in accident severity modeling, there is no contribution that conducts a quantitative comparison of Bayesian network and Regression model. Therefore, both Bayesian network and Regression model will be applied to accident severity modeling in this work and the accuracy of the two models will be compared.

3. Data

The data set for this work contains police-reported traffic accident records for Jilin province, China, in 2010. With records containing missing values eliminated, our final data set consists of 2,246 cases, which are all motor-vehicle involved accidents. In addition to severity information, the data contains information regarding accident characteristics (accident occurrence time and accident location), vehicle characteristics (vehicle type involved and vehicle condition), environmental factors (weather condition and visibility distance), and road conditions (pavement condition, road geometrics and roadway surface condition, etc.).

Previous studies [7, 12] indicated that the factors associated with accident severity mainly include road characteristics, accidents characteristics, vehicle characteristics, driver characteristics, and environmental factors. However, driver characteristics related variables are not involved since suitable records for driver characteristics are not available in the data set. Based on a preliminary correlation test, 3 dependent variables and 14 candidate independent variables were selected from the data set, shown in Table 1. In the process of accident severity modeling, 50% of the total records are selected

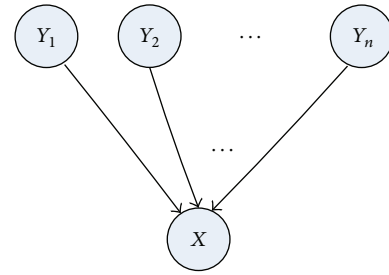


FIGURE 1: An example of Bayesian network.

as training data to calibrate the prediction models, and the left 50% are set aside as testing data.

4. Accident Severity Modeling

4.1. Accident Severity Modeling with Bayesian Network

4.1.1. A Brief Overview of the Bayesian Network. Over the last decade, Bayesian network has become a popular representation for encoding uncertain expert knowledge in expert systems. It has been applied to many fields, such as medicine, document classification, information retrieval, image processing, data fusion, and decision support systems [13].

Bayesian network is a graphical model representing random variables and their conditional dependencies. Figure 1 shows a simple Bayesian network, in which Y_1, Y_2, \dots, Y_n and X are random variables represented by nodes. Y_1, Y_2, \dots, Y_n are called parents of X and X is called the child of Y_1, Y_2, \dots, Y_n . The directed edge from Y_1, Y_2, \dots, Y_n to X indicates the dependence of X on its parent node.

4.1.2. Structure Learning. In most cases, the graphical structure of a Bayesian network needs to be automatically learnt from the data. This learning process can be described as follows. Let a random variable S be the structure of a Bayesian network and let $P(S)$ be its prior probability distribution. Given data set D , which consists of all the variables represented by nodes in the Bayesian network (e.g., X and Y_1, Y_2, \dots, Y_n in Figure 1), based on the Bayesian theorem, the posterior probability of S can be calculated as

$$P(S | D) = \frac{P(S, D)}{P(D)} = \frac{P(S)P(D | S)}{P(D)}, \quad (1)$$

where $P(S | D)$ is the posterior probability of structure S , $P(S)$ is the prior probability distribution of S , and $P(D)$ is the probability distribution of data set D .

The posterior probability $P(S | D)$ is also called Bayesian score, and (1) is called Bayesian score function. The structure that maximizes the Bayesian score will be chosen as the final structure of the Bayesian network.

4.1.3. Parameter Learning. In order to fully specify a Bayesian network, it is necessary to specify the conditional probability of each node upon its parent nodes in the network,

TABLE I: Variables and statistics based on survey data.

| Factors | Variables | Values | Percentage (%) |
|--------------------------|--------------------------------------|-------------------------|----------------|
| Accident severity | Number of fatalities: Nof | 0 : 1 | 89.59 |
| | | ≥ 1 : 2 | 10.41 |
| | Number of injuries: Noi | 0 : 1 | 9.86 |
| | | [1, 3) : 2 | 85.89 |
| | | [3, 11) : 3 | 4.14 |
| | | ≥ 11 : 4 | 0.11 |
| | Property damage (Yuan): Pd | <1000 : 1 | 61.18 |
| [1000, 30000) : 2 | | 37.19 | |
| | ≥ 30000 : 3 | 1.63 | |
| Accident characteristics | Time of day: Tod | day [6:00, 18:00) : 1 | 69.12 |
| | | night [18:00, 6:00) : 2 | 30.88 |
| | Location-Motor vehicle lanes: L-Mvl | Yes: 1 | 71.68 |
| | | No: 2 | 28.32 |
| | Location-Crosswalk: L-C | Yes: 1 | 3.42 |
| | | No: 2 | 96.58 |
| | Location-Regular road section: L-Rrs | Yes: 1 | 60.01 |
| | | No: 2 | 39.99 |
| | Location-Intersection: L-I | Yes: 1 | 38.90 |
| | | No: 2 | 61.10 |
| Vehicle characteristics | Motorcycle involved: Mi | Yes: 1 | 16.97 |
| | | No: 2 | 83.03 |
| | Bus or truck involved: Bti | Yes: 1 | 95.30 |
| | | No: 2 | 4.70 |
| | Vehicle condition: Vc | Good: 1 | 73.79 |
| Poor: 2 | | 26.21 | |
| Environmental factors | Weather condition: Wc | Sunny: 1 | 89.48 |
| | | Other: 2 | 10.52 |
| | Visibility distance (meter): Vd | <50 : 1 | 8.90 |
| | | [50, 100) : 2 | 22.70 |
| | | [100, 200) : 3 | 19.86 |
| | ≥ 200 : 4 | 48.54 | |
| Roadway characteristics | Pavement condition: Pc | Asphalt or cement: 1 | 99.80 |
| | | Other: 2 | 0.20 |
| | Roadway surface condition: Rsc | Dry: 1 | 85.16 |
| | | Other: 2 | 14.84 |
| | Road geometrics: Rg | Flat and straight: 1 | 98.57 |
| | | Hill or bend: 2 | 1.43 |
| | Traffic signal control: Tsc | Yes: 1 | 17.46 |
| No: 2 | | 82.54 | |

given the structure S and the data set D . Parameter learning refers to the process of identifying the parameters in the conditional distribution of any child nodes on the joint distribution of its parent nodes. Bayesian estimation is one of the methods for parameter learning, which assumes that parameter θ_i is a random variable with prior distribution $P(\theta_i)$. According to the Bayesian theorem, the posterior probability for the parameter ($P(\theta_i | D)$) given data set D is computed as

$$P(\theta_i | D) = \frac{P(\theta_i, D)}{P(D)} = \frac{P(\theta_i)P(D | \theta_i)}{P(D)}, \quad (2)$$

where $P(\theta_i)$ is the prior probability distribution of parameter θ_i and $P(D)$ is the probability distribution of data set D .

4.1.4. Learning Results. Since the number of possible structures grows exponentially as a function of the number of variables, it is computationally infeasible to find the most probable network structure, given the data, by exhaustively enumerating all possible network structures. Cooper and Herskovits [14] and Herskovits [15] proposed a greedy algorithm, called the K2 algorithm, which becomes one of the most popular methods for structure learning of Bayesian network. Besides the basic Bayesian theories, the K2

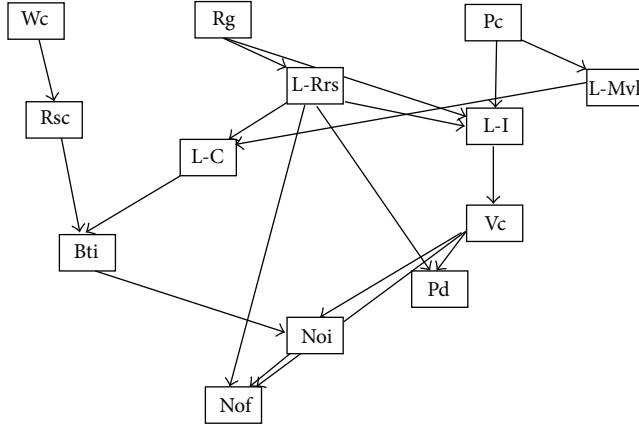


FIGURE 2: The structure of the Bayesian network.

algorithm uses two assumptions, namely, that there is an ordering available on the variables and that, a priori, all structure are equally likely. The K2 algorithm considers each node in the order given to it as input and determines its parents as follows. Initially, assume that each node has no parent. Then, add parents for some nodes when the score (computed by using (1)) of the resulting structure could be increased. Try to add parents for each node until no more parents can increase the score. Then the structure of the highest score is the final structure [14].

The structure of the severity prediction Bayesian network is learned by employing the K2 algorithm and the Full-BNT toolbox, which is an open-source Matlab package for directed graphical models [16]. The final network structure is shown in Figure 2. The results indicate that Bayesian network for accident severity analysis is composed of 13 nodes and the concerned edges, which represent the relationships between the nodes.

Based on the developed structure, the parameters are learned by employing the method of Bayesian estimation. The prior distributions of all the variables are assumed to be Dirichlet distribution, which is a kind of conjugate distribution allowing closed form for posterior distribution of parameters and closed-form solution for prediction. The Full-BNT toolbox of Matlab is employed to realize the algorithm of Bayesian estimation.

As the parent nodes gather the impacts of the indirect nodes and deliver them to the child nodes, the influence of parent nodes will be focused on. Under the impact of the parent nodes, that is, factors which have direct edge to the severity indicators in this structure, the parameter learning results of number of fatalities (Nof), number of injuries (Noi), and property damage (Pd) are shown in Tables 2, 3, and 4 respectively. The indicators of Mean Absolute Percentage Error (MAPE) and *Hit ratio* are adopted to examine the accuracy of the models.

MAPE, which looks at the average percentage difference between predicted values and observed ones, is calculated as

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - P_i}{A_i} \right|, \quad (3)$$

where A_i is the observed value and P_i is the predicted value for observation i .

The MAPE value of the fatality forecasting model is 0.0226, and the *Hit ratio* is 100%, which recommend that this model has a high accuracy [17, 18].

According to the developed structure, Nof's parent nodes are L-Rrs, Vc, and Noi. The estimation results indicate that the probability of occurrence of fatal accident increases when the condition of the involved vehicle gets worse. Moreover, higher number of deaths is associated with higher number of injuries. The accident that occurs at normal section of road, but not at abnormal section or intersection, tends to cause more fatalities. The reason may be that the involved vehicle usually speeds down when going through intersection or abnormal section of road.

The MAPE value of the injury forecasting model is 0.0013, and the *Hit ratio* is 100%, which presents a high accuracy of the model.

Two parent nodes, namely, Bti and Vc, have direct impacts on number of injuries in the accident. The estimation results show that bus or truck involved accident tends to cause more injuries. In addition, the worse the vehicle condition is, the more injuries are in the accident.

The MAPE value of the property damage forecasting model is 0.0019, and the *Hit ratio* is 100%, which shows a high accuracy of the model.

Two parent nodes, that is, L-Rrs and Vc, have direct impact on property damage. The results indicate that, like the influences of Vc on Nof and Noi, poor vehicle condition is associated with large amount of property damage and vice versa. In addition, the accident that occurs at irregular section of road or intersection tends to cause large amount of property damage. Combining the effects of L-Rrs on Pd and Nof, it can be deduced that the accident that occurs at regular section of road tends to result in high number of deaths but small amount of property damage.

4.2. Accident Severity Modeling with Regression Models. The most commonly used Regression models in traffic injury analysis are the Logistic Regression model and the Ordered Probit model [1–6]. Since the alternatives of Noi and Pd are all ordered and the Logistic Regression model would fail to account for the ordinal nature of the dependent variable and have the problem of independence from irrelevant alternatives (IIA) [19], Ordered Probit model will be employed in forecasting of Noi and Pd. Besides, one of the Logistic Regression models-Binary Logit model, will be adopted in prediction of Nof, which has two discrete alternatives.

4.2.1. Binary Logit Model. As one of the Binomial choice models, Binary Logit model is commonly used in discrete choice modeling. According to the random utility theory [20], the utility of alternative i ($i = 1$ or 2 for Nof = 0 or Nof ≥ 1 , resp.) for accident n can be specified as

$$U_{in} = V_{in} + \varepsilon_{in}, \quad (4)$$

where V_{in} denotes the deterministic component of U_{in} , and ε_{in} is the random component of U_{in} .

TABLE 2: Parameter learning results of the fatality forecasting model.

| No. | Variables | | | Estimation results | | | | | | | | N |
|-----|-----------|----|-----|--------------------|--------|----------------|----------------|----------|--------|----------------|----------------|------|
| | L-Rrs | Vc | Noi | Nof ≥ 1 | | | | Nof = 0 | | | | |
| | | | | Bayesian | Test | Absolute error | Relative error | Bayesian | Test | Absolute error | Relative error | |
| 1 | 1 | 1 | 1 | 0.0052 | 0.0000 | 0.0052 | 1.0000 | 0.9948 | 1.0000 | 0.0052 | 0.0052 | 1123 |
| 2 | 1 | 1 | 2 | 0.9995 | 1.0000 | 0.0005 | 0.0005 | 0.0005 | 0.0000 | 0.0005 | 1.0000 | |
| 3 | 1 | 1 | 3 | 0.6622 | 0.6667 | 0.0045 | 0.0068 | 0.3378 | 0.3333 | 0.0045 | 0.0133 | |
| 4 | 1 | 2 | 1 | 0.2628 | 0.2626 | 0.0002 | 0.0008 | 0.7372 | 0.7374 | 0.0002 | 0.0003 | |
| 5 | 1 | 2 | 2 | 0.9855 | 0.9855 | 0.0000 | 0.0000 | 0.0145 | 0.0145 | 0.0000 | 0.0000 | |
| 6 | 1 | 2 | 3 | 0.9994 | 1.0000 | 0.0006 | 0.0006 | 0.0006 | 0.0000 | 0.0006 | 1.0000 | |
| 7 | 2 | 1 | 1 | 0.2012 | 0.2000 | 0.0012 | 0.0060 | 0.7988 | 0.8000 | 0.0012 | 0.0015 | |
| 8 | 2 | 1 | 2 | 0.9369 | 0.9375 | 0.0006 | 0.0006 | 0.0631 | 0.0625 | 0.0006 | 0.0095 | |
| 9 | 2 | 1 | 3 | 0.5000 | 0.0000 | 0.5000 | 1.0000 | 0.5000 | 0.0000 | 0.5000 | 1.0000 | |
| 10 | 2 | 2 | 1 | 0.2601 | 0.2601 | 0.0000 | 0.0000 | 0.7399 | 0.7399 | 0.0000 | 0.0000 | |
| 11 | 2 | 2 | 2 | 0.9575 | 0.9575 | 0.0000 | 0.0000 | 0.0425 | 0.0425 | 0.0000 | 0.0000 | |
| 12 | 2 | 2 | 3 | 0.9087 | 0.9091 | 0.0004 | 0.0004 | 0.0913 | 0.0909 | 0.0004 | 0.0044 | |

TABLE 3: Parameter learning results of the injury forecasting model.

| Variables | No. | Bti Vc | 1 | 2 | 3 | 4 |
|--------------------|-------------------------|----------------|--------|--------|--------|--------|
| | | | 1 | 1 | 2 | 2 |
| | | | 1 | 2 | 1 | 2 |
| Estimation results | Noi = 0 | Bayesian | 0.1299 | 0.0907 | 0.2444 | 0.1813 |
| | | Test | 0.1295 | 0.0906 | 0.2439 | 0.1812 |
| | | Absolute error | 0.0004 | 0.0001 | 0.0005 | 0.0001 |
| | | Relative error | 0.0031 | 0.0011 | 0.0020 | 0.0006 |
| | $1 \leq \text{Noi} < 3$ | Bayesian | 0.8552 | 0.8643 | 0.7293 | 0.7542 |
| | | Test | 0.8561 | 0.8644 | 0.7317 | 0.7544 |
| | | Absolute error | 0.0009 | 0.0001 | 0.0024 | 0.0002 |
| | | Relative error | 0.0011 | 0.0001 | 0.0033 | 0.0003 |
| | Noi ≥ 3 | Bayesian | 0.0150 | 0.0450 | 0.0263 | 0.0646 |
| | | Test | 0.0144 | 0.0450 | 0.0244 | 0.0645 |
| | | Absolute error | 0.0006 | 0.0000 | 0.0019 | 0.0001 |
| | | Relative error | 0.04 | 0.0000 | 0.0722 | 0.0015 |
| N | | | 1123 | | | |

TABLE 4: Parameter learning results of the property damage forecasting model.

| Variables | No. | L-Rrs Vc | 1 | 2 | 3 | 4 |
|--------------------|-------------------------------|----------------|--------|--------|--------|--------|
| | | | 1 | 1 | 2 | 2 |
| | | | 1 | 2 | 1 | 2 |
| Estimation results | Pd < 1000 | Bayesian | 0.7905 | 0.6802 | 0.4402 | 0.4641 |
| | | Test | 0.7917 | 0.6803 | 0.4405 | 0.4641 |
| | | Absolute error | 0.0012 | 0.0001 | 0.0003 | 0.0000 |
| | | Relative error | 0.0015 | 0.0001 | 0.0006 | 0.0000 |
| | $1000 \leq \text{Pd} < 30000$ | Bayesian | 0.1983 | 0.3072 | 0.5470 | 0.5093 |
| | | Test | 0.1979 | 0.3072 | 0.5476 | 0.5093 |
| | | Absolute error | 0.0004 | 0.0000 | 0.0006 | 0.0000 |
| | | Relative error | 0.0019 | 0.0000 | 0.0011 | 0.0000 |
| | Pd ≥ 30000 | Bayesian | 0.0113 | 0.0126 | 0.0129 | 0.0266 |
| | | Test | 0.0104 | 0.0125 | 0.0119 | 0.0266 |
| | | Absolute error | 0.0009 | 0.0001 | 0.0010 | 0.0000 |
| | | Relative error | 0.0782 | 0.0048 | 0.0772 | 0.0002 |
| N | | | 1123 | | | |

Here V_{in} can be written as

$$V_{in} = \sum_{k=1}^K \theta_k X_{kin}, \quad (5)$$

where X_{kin} is attribute k ($k = 1, \dots, K$) for accident n and alternative i , and θ_k is the estimable coefficients, which can be estimated by adopting the Maximum Likelihood method.

Assuming ε_{in} follows Gumbel distribution, the choice probability of alternative i ($i = 1$) for accident n is then

$$P_{1n} = \frac{\exp(V_1)}{\exp(V_1) + \exp(V_2)}, \quad (6)$$

and the choice probability of alternative i ($i=2$) for accident n is

$$P_{2n} = 1 - P_{1n}. \quad (7)$$

4.2.2. Ordered Probit Model. The Ordered multiple choice model assumes the relationship

$$\sum_{j=1}^J P_n(j) = F(\alpha_j - \beta_j X_n, \theta), \quad j = 1, \dots, J-1, \quad (8)$$

$$P_n(J) = 1 - \sum_{j=1}^J P_n(j),$$

where $P_n(j)$ is the probability that alternative j happens in accident n ($n = 1, \dots, N$), α_j is an alternative specific constant, X_n is a vector of the attributes of accident n , β_j is a vector of the estimated coefficients, and θ is a parameter that controls the shape of probability distribution F . Therefore, F can have various shapes of distribution based on different value of θ .

The Ordered Probit model, which assumes standard normal distribution for F , is the most commonly used ordered multiple choice model [21]. The Ordered Probit model has the following form:

$$\begin{aligned} P_n(1) &= \Phi(\alpha_1 - \beta_j X_n) \\ P_n(j) &= \Phi(\alpha_j - \beta_j X_n) - \Phi(\alpha_{j-1} - \beta_j X_n), \\ & \quad j = 2, \dots, J-1, \quad (9) \\ P_n(J) &= 1 - \sum_{j=1}^{J-1} P_n(j), \end{aligned}$$

where $P_n(j)$ is the cumulative standard normal distribution function. For all the probabilities to be positive, it must satisfy that $\alpha_1 < \alpha_2 < \dots < \alpha_{j-1}$.

4.2.3. Estimation Results. By using logistic and probit procedure in SAS [22], the Binary Logit model and Ordered Probit models are estimated, and the results are shown in Table 5.

5. Discussions

By comparing the test results of MAPE and *Hit ratio* with respect to the predictions of the three severity indicators, it can be concluded that the goodness of fit of Bayesian network is higher than that of Regression models. This suggests that Bayesian network is more suitable in accident severity prediction than Regression models regarding modeling accuracy.

Besides goodness of fit, there is also difference between Bayesian network and Regression models regarding the interactions between the variables in the model. In Bayesian network, indirect nodes (or variables), which are related to the dependent variable, affect their own child nodes first, and then the impacts are delivered to the related edges and nodes until they arrive the dependent variable [23, 24]. As shown in Figure 3(1), every indirect node's change will cause dependent variable's change. The impact of indirect node on dependent variable can be obtained by inference based on the constructed Bayesian network. For instance, the impacts of L-C on the three accident severity variables are inferred according to the Bayesian network for accident severity analysis, and the results are shown in Table 6. Comparing with Bayesian network, all the independent variables, either indirect node, child node or direct node in the Bayesian network, affect the dependent variable directly in the Regression models [25, 26]. The interactions between variables in the Regression models are shown in Figure 3(2).

Moreover, for Regression models, two independent variables cannot exist in one model if they are related to each other. This causes the missing of some influences between variables. Also, Regression models will fail to present the impact between dependent variable and dependent variable as well as the interaction between independent variable and independent variable, such as the impact of Noi on Nof and the effect of Rsc on Bti in this study, respectively, which can be presented by the Bayesian work shown in Figure 2.

Furthermore, as mentioned above, most of the Regression models have their own assumptions and predefined underlying relationships between dependent and independent variables (i.e., linear relations between the variables or independence between variables) [7, 8]. The differences between Regression models and Bayesian network also reflect the methods of probabilistic reasoning [27–29]. That is, Bayesian network can reason under uncertainty, but Regression models cannot. In addition, for parameter estimation, Regression models need complete (without missing values) and quantitative data, while Bayesian network can be constructed with incomplete data or qualitative information [30, 31].

The above characteristics of Bayesian network prove that, compared with Regression models, Bayesian network is more suitable to be adopted in accident severity analysis.

6. Conclusions

In this paper, two modeling techniques, that is, Bayesian network and Regression models, are investigated in accident severity modeling. The goodness of fit of the two methods is compared according to the test results, and the differences

TABLE 5: Estimation results of the Regression models.

| Variables | Fatality forecasting model | | Injury forecasting model | | Property damage forecasting model | |
|---------------|----------------------------|---------|--------------------------|---------|-----------------------------------|---------|
| | Coef. | Z-stat. | Coef. | Z-stat. | Coef. | Z-stat. |
| Constant | -2.57 | -12.53 | | | | |
| Mi | 0.44 | 3.36 | -0.14 | -2.50 | -0.09 | -1.85 |
| Bi | 1.11 | 8.82 | -0.30 | -4.99 | | |
| Wc | -0.27 | -1.66 | 0.23 | 2.70 | -0.12 | -1.85 |
| Tod | -0.44 | -4.03 | | | -0.15 | -3.29 |
| Vd | | | | | 0.10 | 4.88 |
| Pc | | | -0.38 | -5.24 | | |
| Tsc | 0.22 | 2.06 | -0.07 | -1.57 | 0.11 | 2.76 |
| L-Mvl | | | -0.08 | -1.57 | | |
| L-C | -0.52 | -1.39 | | | -0.40 | -3.51 |
| L-Rrs | 0.73 | 5.98 | | | 0.72 | 4.07 |
| L-I | | | 0.23 | 4.94 | 0.20 | 1.10 |
| α_1 | | | -1.50 | | 0.80 | |
| α_2 | | | 1.47 | | 2.77 | |
| MAPE | 0.0530 | | 0.0415 | | 0.0698 | |
| Hit ratio (%) | | 84.65 | | 80.20 | | 60.23 |

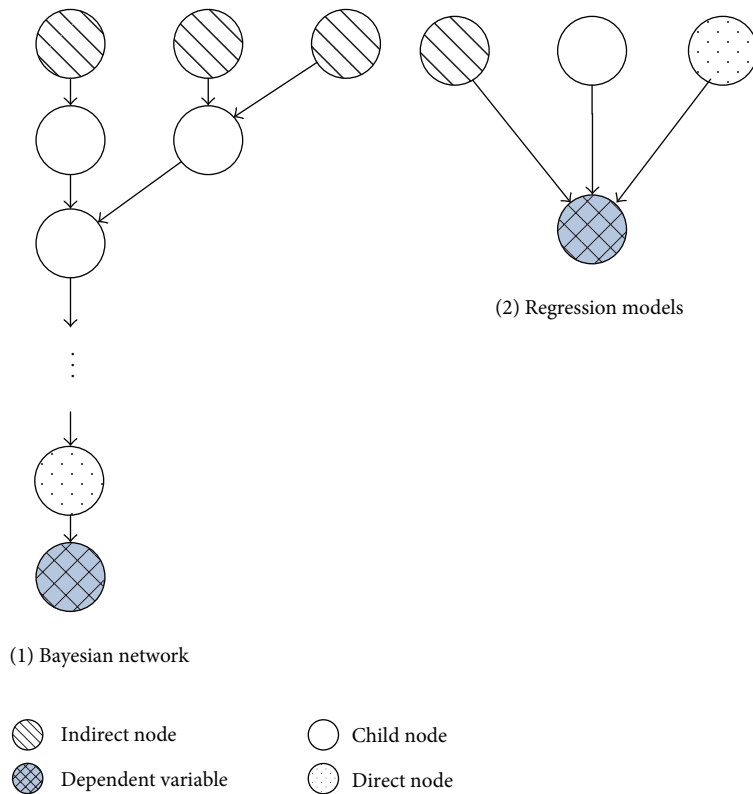


FIGURE 3: Comparison of Bayesian network and Regression models with respect to the interactions between variables.

between the two methods are analyzed. The results suggest that, comparing with Regression models, Bayesian network is more suitable for accident severity prediction.

Study results can be applied to predicting traffic accident severity and identifying the key effects of contributed factors on accident severity. By comparing Bayesian network and

Regression models, it also makes a methodological contribution in enhancing prediction accuracy of severity estimation.

It should be pointed out that both the structure and the parameter of the proposed Bayesian network will change when there are specific numbers of new reported cases added into the data set. According to the study by Zhang [32],

TABLE 6: Impact of L-C on accident severity.

| Accident severity | | At crosswalk | Not at crosswalk |
|-------------------|-------------------------------|--------------|------------------|
| Noi | Noi = 0 | 0.0578 | 0.0881 |
| | $1 \leq \text{Noi} < 3$ | 0.8864 | 0.8883 |
| | $\text{Noi} \geq 3$ | 0.0558 | 0.0237 |
| Nof | Nof > 0 | 0.9080 | 0.8990 |
| | Nof = 0 | 0.0920 | 0.1010 |
| Pd | $\text{Pd} < 1000$ | 0.7484 | 0.7313 |
| | $1000 \leq \text{Pd} < 30000$ | 0.2439 | 0.2614 |
| | $\text{Pd} \geq 30000$ | 0.0077 | 0.0074 |

the structure and the parameter of the Bayesian network will change when the amount of new records reaches 10% and 5% of the number of the original cases, respectively.

One limitation of current work is that some factors, such as driver characteristics and traffic condition, which have potential effects on accident severity, are not considered because of the lack of suitable data. Further study should be conducted to examine the impacts of these factors on accident severity.

Acknowledgments

The research is funded by the National Natural Science Foundation of China (50908099 and 51078167) and the Doctoral Program of Higher Education of China (201104493).

References

- [1] A. S. Al-Ghamdi, "Using logistic regression to estimate the influence of accident factors on accident severity," *Accident Analysis and Prevention*, vol. 34, no. 6, pp. 729–741, 2002.
- [2] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: an exploratory empirical analysis," *Accident Analysis and Prevention*, vol. 40, no. 1, pp. 260–266, 2008.
- [3] M. Bédard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," *Accident Analysis and Prevention*, vol. 34, no. 6, pp. 717–727, 2002.
- [4] K. K. W. Yau, H. P. Lo, and S. H. H. Fung, "Multiple-vehicle traffic accidents in Hong Kong," *Accident Analysis and Prevention*, vol. 38, no. 6, pp. 1157–1161, 2006.
- [5] T. Yamamoto and V. N. Shankar, "Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects," *Accident Analysis and Prevention*, vol. 36, no. 5, pp. 869–876, 2004.
- [6] K. M. Kockelman and Y. Kweon, "Driver injury severity: an application of ordered probit models," *Accident Analysis and Prevention*, vol. 34, no. 3, pp. 313–321, 2002.
- [7] L. Chang and H. Wang, "Analysis of traffic injury severity: an application of non-parametric classification tree techniques," *Accident Analysis and Prevention*, vol. 38, no. 5, pp. 1019–1027, 2006.
- [8] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Analysis and Prevention*, vol. 43, no. 1, pp. 402–411, 2011.
- [9] M. Simoncic, "A Bayesian network model of two-car accidents," *Journal of Transportation and Statistics*, vol. 7, no. 2-3, pp. 13–25, 2004.
- [10] K. Ozbay and N. Noyan, "Estimation of incident clearance times using Bayesian Networks approach," *Accident Analysis and Prevention*, vol. 38, no. 3, pp. 542–555, 2006.
- [11] A. Gregoriades, "Towards a user-centred road safety management method based on road traffic simulation," in *Proceedings of the 39th Conference on Winter Simulation: 40 years! The Best is Yet to come*, pp. 1905–1914, Washington, DC, USA, December 2007.
- [12] P. Kopelias, F. Papadimitriou, K. Papandreou, and P. Prevedouros, "Urban freeway crash analysis geometric, operational, and weather effects on crash number and severity," *Transportation Research Record*, vol. 2015, pp. 123–131, 2007.
- [13] A. Mittal, A. Kassim, and T. Tan, *Bayesian Network Technologies: Applications and Graphical Models*, IGI Publishing, New York, NY, USA, 2007.
- [14] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Machine Learning*, vol. 9, no. 4, pp. 309–347, 1992.
- [15] E. Herskovits, *Computer-based probabilistic-network construction [Ph.D. dissertation]*, Medical Information Sciences, Stanford University, Stanford, Calif, USA, 1991.
- [16] D. Eaton and K. Murphy, "Bayesian structure learning using dynamic programming and MCMC," in *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI '07)*, pp. 101–108, July 2007.
- [17] F. Zong, H. Y. Lin, B. Yu, and X. Pan, "Daily commute time prediction based on Genetic algorithm," *Mathematical Problems in Engineering*, vol. 2012, Article ID 321574, 20 pages, 2012.
- [18] F. Zong, Z. C. Juan, and H. F. Jia, "Examination of staggered shifts impacts on travel behavior: a case study of Beijing," *Transport*, vol. 28, no. 2, pp. 175–185, 2013.
- [19] P. Ray, "Independence of irrelevant alternatives," *Econometrica*, vol. 41, pp. 987–991, 1973.
- [20] J. L. Bowman and M. E. Ben-Akiva, "Activity-based disaggregate travel demand model system with activity schedules," *Transportation Research A*, vol. 35, no. 1, pp. 1–28, 2000.
- [21] M. A. Quddus, R. B. Noland, and H. C. Chin, "An analysis of motorcycle injury and vehicle damage severity using ordered probit models," *Journal of Safety Research*, vol. 33, no. 4, pp. 445–462, 2002.
- [22] R. A. Hanneman, 2013, Ordered Logit and Probit Models with PROC LOGIST and PROC PROBIT. The course in Generalized Linear Models. University of California, <http://faculty.ucr.edu/~hanneman/soc271/ologit.html>.

- [23] B. Z. Yao, C. Y. Yang, and J. B. Yao, "Tunnel surrounding rock displacement prediction using support vector machine," *International Journal of Computational Intelligence Systems*, vol. 3, no. 6, pp. 843–852, 2010.
- [24] J. B. Yao, B. Z. Yao, L. Li, and Y. L. Jiang, "Hybrid model for displacement prediction of tunnel surrounding rock," *Neural Network World*, vol. 22, pp. 263–275, 2012.
- [25] Y. Bin, Y. Zhongzhen, and Y. Baozhen, "Bus arrival time prediction using support vector machines," *Journal of Intelligent Transportation Systems*, vol. 10, no. 4, pp. 151–158, 2006.
- [26] B. Yu, W. H. K. Lam, and M. L. Tam, "Bus arrival time prediction at bus stop with multiple routes," *Transportation Research C*, vol. 19, no. 6, pp. 1157–1170, 2011.
- [27] B. Yu, Z. Yang, and B. Yao, "An improved ant colony optimization for vehicle routing problem," *European Journal of Operational Research*, vol. 196, no. 1, pp. 171–176, 2009.
- [28] B. Yu, Z.-Z. Yang, and J.-X. Xie, "A parallel improved ant colony optimization for multi-depot vehicle routing problem," *Journal of the Operational Research Society*, vol. 62, no. 1, pp. 183–188, 2011.
- [29] B. Yu, Z. Z. Yang, and S. Li, "Real-time partway deadheading strategy based on transit service reliability assessment," *Transportation Research A*, vol. 46, no. 8, pp. 1265–1279, 2012.
- [30] B. Y. Chen, W. H. K. Lam, A. Sumalee, Q. Q. Li, H. Shao, and Z. X. Fang, "Finding reliable shortest paths in road networks under uncertainty," *Networks and Spatial Economics*, vol. 13, no. 2, pp. 123–148, 2013.
- [31] B. Y. Chen, W. H. K. Lam, A. Sumalee, and H. Shao, "An efficient solution algorithm for solving multi-class reliability-based traffic assignment problem," *Mathematical and Computer Modelling*, vol. 54, no. 5-6, pp. 1428–1439, 2011.
- [32] H. Y. Zhang, *Analyzing traffic accident situation with Bayesian network [Doctoral dissertation]*, Jilin University, Changchun, China, 2013, (Chinese).



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

