Research Article

Causation Analysis of Hazardous Material Road Transportation Accidents by Bayesian Network Using Genie

Xiaoli Ma , Yingying Xing , and Jian Lu

The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China

Correspondence should be addressed to Jian Lu; jianjohnlu@tongji.edu.cn

Received 1 April 2018; Accepted 28 May 2018; Published 5 August 2018

Academic Editor: Xiaobo Qu

Copyright © 2018 Xiaoli Ma 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.

With the increase of hazardous materials (Hazmat) demand and transportation, frequent Hazmat road transportation accidents had arisen the widespread concern in the community. Thus, it is necessary to analyze the risk factors' implications, which would make the safety of Hazmat transportation evolve from "passive type" to "active type". In order to explore the influence of risk factors resulting in accidents and predict the occurrence of accidents under the combination of risk factors, 839 accidents that have occurred for the period 2015–2016 were collected and examined. The Bayesian network structure was established by experts' knowledge using Dempster-Shafer evidence theory. Parameter learning was conducted by the Expectation-Maximization (EM) algorithm in Genie 2.0. The two main results could be likely to obtain the following. (1) The Bayesian network model can explore the most probable factor or combination leading to the accident, which calculated the posterior probability of each risk factor. For example, the importance of three or more vehicles in an accident leading to the severe accident is higher than less vehicles, and in the absence of other evidences, the most probable reasons for "explosion accident" are vehicles carrying flammable liquids, larger quantity Hazmat, vehicle failure, and transporting in autumn. (2) The model can predict the occurrence of accident by setting the influence degrees of specific factor. Such that the probability of rear-end accidents caused by "speeding" is 0.42, and the probability could reach up to 0.97 when the driver is speeding at the low-class roads. Moreover, the complex logical relationship in Hazmat road transportation accidents could be obtained, and the uncertain relation among various risk factors could be expressed. These findings could provide theoretical support for transportation corporations and government department on taking effective measures to reduce the risk of Hazmat road transportation.

1. Introduction

In recent years, the demand for hazardous materials (Hazmat) has increased, resulting in increasing transportation requirement. More than 95% Hazmat require off-site transportation in China, and 63% are transported by road in Brazil, as well as 90% in the United States [1, 2]. However, Hazmat could provide the great convenience for people's life, but also significant risks to environment and human health exist. For instance, a total of 3744 heavy trucks were involved in severe accidents in the United States, of which 3% were carrying Hazmat [3]. On January 11, 2015, a tanker truck carrying gasoline collided with a bus in Pakistan, causing 57 deaths. And a tanker truck carrying liquid ammonia collided with a van, resulting in a large-scale spill of liquid ammonia, leading to 28 deaths and the number of poisons was up to 350 on

March 29, 2005, which arose on negative social impacts in China [4].

Hazmat transportation accidents would be able to produce catastrophic influence on human health, public safety, environment, and property due to the special characteristic of Hazmat, attracting more attention from general public and government on the management of Hazmat road transportation. Thus, how to improve the transportation condition and reduce the risk of transportation have become important and urgent problems for the industrial development. A growing amount studies about Hazmat transportation and production have been conducted [5–7]. Therefore, the need for investigating risk factors that contribute to Hazmat accident and the relationship of risk factors are highlighted to reduce the risk of Hazmat transportation. To that end, the effective method to describe and evaluate the accident process is causation

analysis, which could be used to determine government priorities related to the implementation of prevention measures [8]. And causation analysis also could provide the theoretical support for actionable information of controlling over the risk factors for the transportation corporations. In addition, exploring the most probable factor or combination leading to accidents and predicting accidents are the important research topics in the field of Hazmat safety, reducing the frequency and severity of accidents.

2. Literature Review

The purpose of this study is to explore risk factors to reduce the risk of Hazmat road transportation. Many studies have been conducted by using statistical methods. Haastrup and Brockhoff [9] statistically analyzed the cases of Hazmat accidents in Western Europe, and 39% of accidents occurred during transportation; in 682 accidents the consequence included fatality. A study about Hazmat transportation accidents divided risk factors into human, vehicle, packing, transportation facilities, road conditions, and environmental conditions [5]. Shen [10] studied 708 accidents with Hazmat in China from 2004 to 2011 and found that accidents easily occurred at expressways, and the higher probability of spill accident is associated with accident type. Fang et al. [11] concluded that speeding was the main reason for Hazmat transportation accidents through the analysis of accident data between 1999 and 2013.

Although statistical methods could analyze the relationships between accidents and the risk factors, they cannot account for the interplay among different factors and fail to reflect the fact that an accident is not usually the result of a single factor [12]. The use of causation analysis theory for accidents could extract the accident mechanism and accident models from a large number of typical accidents. For instance, Jason et al. [13] conducted the study about the influence of vehicle, occupant, driver, and environmental characteristics on accident injuries involved with heavy-duty trucks, and the conclusion was obtained by using the heteroskedastic ordered probit models, which showed that the likelihood of severe accident is estimated to rise with the more vehicles involved in accident. Uddin and Huynh [14] used an ordered probit model to explore the relationship among drivers, vehicles, roadways, environment, temporal characteristics, and the severity of accident. There was a study by using logit model to study the driver's behaviors effect on accidents, and the results indicated that the more significant risk factors were speeding, not using seatbelt, drivers' age, and drivers with no valid license [15]. In addition, the Bayesian network and tree-based methods were considered to explore deeper accident mechanisms, which is increasingly utilized in traffic accidents analysis. For instance, Oña et al. [16] classified traffic accidents based on the severity of injuries by using the Bayesian networks; the factors associated with fatal or severe accidents were identified by inference, such as accident type, the driver's age, and lighting. In order to simplify the model, Mujalli et al. [17] used Bayesian networks to reduce the number of variables in the study of analyzing the accidents severity on rural roads, and the result showed that the number

of variables could reduce up to 60% (the variables considered are accident type, age, atmospheric factors, gender, lighting, number of injured, and occupants involved), maintaining the good performance of models. Zhao et al. [18] pointed that the three most significant factors influencing Hazmat transportation by applying Bayesian networks were human factors, the transport vehicles and facilities, and the packaging and loading of Hazmat. Chen et al. [19] analyzed the between-accident variance and within-accident correlations by using Bayesian network and explored the risk factors influencing accidents and their heterogeneous impacts on accident severity in rural roads. And in order to improve the efficiency of emergency rescue of Hazmat transportation road accidents, a study was conducted to evaluate the time of accidents dealing based on the Bayesian network [20]. In addition, the Bayesian network model could also be used to describe the probability and risk of accidents [21–24].

However, despite many studies on the traffic accidents and Hazmat accidents, most of them are studied based on the analysis of specific, isolated, and single factor [25, 26]. Moreover, the characteristic of Hazmat was not taken into consideration during the analysis of accidents, limiting the studies of risk factors in Hazmat road transportation. In addition, the statistical methods could reveal the inherent rules on the occurrence of accidents, but the relationship of risk factors was not observed, which cannot reflect the accident mechanism. The application of causal analysis model (such as Bayesian network) can explain the correlation between risk factors and further explain the accident mechanism, but the Bayesian network structure may exist subjectivity due to the experts' knowledge, leading to incorrect description of relationships between nodes in the Bayesian network structure. Therefore, Hazmat road transportation accidents in China from 2015 to 2016 are considered as the research object to explore the potential risk factors of accidents based on experts' knowledge. The Bayesian network is used to explore the most probable factor or combination leading to accident and determine the correlation between the risk factors, providing the decision-making basis for Hazmat transportation corporations and government departments to reduce the risk of Hazmat transportation.

3. Database

The Hazmat transportation accident data was obtained from State Work Accident Briefing System, and Chemical Accidents Information Network for two years (2015–2016) in China, and the weather data was obtained from the China Meteorological Administration. The regional distribution of Hazmat transportation accidents is shown in Figure 1. The database considered in the study contains 839 records, and each record contains detailed information including the date, time, location, type of accidents, type and number of vehicles involved in accident, driver characteristic, the quantity and categories of Hazmat, accident consequence, causes of accident, and a detailed description of the accident. Sixteen variables extracted from the database were considered as the significant factors, which are shown in Table 1.

FIGURE 1: Regional distribution of accidents.

Accident information is accident type (rear-end, side-swipe, rollover, collision, and vehicle failure) and accident consequence (explosion, fire, spill, and nonspill). Previous studies [14, 27] divided injury severity into five categories; the accident severity in the paper is considered as no injury, severe injury, and fatality. Simplified classification of accident severity could ease the issue of potential relationship of related consequences of an accident and ensure the sufficient sample size for the Bayesian network model [28, 29]. In the paper, the simplified classification of accident severity would obtain the better results.

Hazmat information is Hazmat categories and quantity of Hazmat transportation.

Driver information is characteristics of the driver, such as age and behavior.

Location information is road surface condition and accident location (such as Group one, Group two, Group three, and Group four); the special road section including intersection, freeway service areas, toll stations, and gas stations are considered in the study.

Vehicle information is type and number of vehicles involved in accident.

Environment information is time distribution of accident (hour, day, and month), visibility (dawn: 5:00 to 6:59 am, day: 7:00 am to 4:59 pm, dusk: 5:00 to 6:59 pm, and dark: 7:00 pm to 4:59 am), and weather conditions (sunny, cloudy, rainy and snowy, and fog and haze) [18].

4. Methodology

4.1. Definition of Bayesian Network. Bayesian network is considered as the effective method to describe the causality between the risk factors and the output in the system, also referred to as the belief network. The Bayesian network is

a Directed Acyclic Graph (DAG) and nodes represent variable status, while the directed edges represent dependencies between variables. The relationship or confidence coefficient between variables could be described by using Conditional Probability Table (CPT). The Bayesian formula is considered as the basis for the Bayesian network model, which could be expressed as

$$P(X | Y) = \frac{P(Y | X) \times P(X)}{P(Y)} \quad (1)$$

where $P(X | Y)$ is the probability of X under the condition of a known event Y . $P(Y | X)$ is the conditional probability of Y at the occurrence of X . And the joint distribution of two random variables X and Y can be expressed as

$$P(X, Y) = P(X) P(Y | X) \quad (2)$$

where $P(X)$ is called the prior probability and $P(Y | X)$ is the posterior probability. Combined with the chain rules, reducing the complexity of the probability model, the joint distribution of n variables is

$$\begin{aligned} P(X_1, X_2, \dots, X_n) \\ = P(X_1) P(X_2 | X_1) \dots P(X_n | X_1, X_2, \dots, X_{n-1}) \end{aligned} \quad (3)$$

and the joint distribution also could be expressed as

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Parent}(X_i)) \quad (4)$$

where $X = \{X_1, X_2, \dots, X_n\}$, setting S is a network structure, P is a set of local probability distributions associated with each variable, X_i denotes the variable node, and $\text{Parent}(X_i)$ denotes the father node of X_i in S .

TABLE 1: Variables of Hazmat road transportation accidents.

Factors	Variables	Variables description	Discretization	Frequency	Percentage
Hazmat factors	Hazmat categories	Explosives	1	27	3.20%
		Toxic gases	2	158	18.90%
		Flammable liquids	3	429	51.10%
		Corrosives	4	121	14.40%
		others	5	104	12.40%
Driver factors	Quantity of Hazmat	<10	1	127	15.10%
		10-24	2	284	33.80%
		25-39	3	358	42.70%
		≥40	4	70	8.40%
Driver factors	Age	24-35	1	144	17.20%
		36-45	2	644	76.70%
		46-60	3	51	6.10%
Location factors	Behaviors	Inappropriate driving	1	13	1.50%
		Speeding	2	36	4.30%
		Fatigue driving	3	20	2.40%
		Normal driving	4	770	91.80%
Location factors	Accident location	Group one	1	360	42.90%
		Group two	2	336	40.00%
		Group three	3	59	7.00%
		Group four	4	84	10.10%
		Intersection	1	18	2.10%
Environment factors	Special section	Freeway service areas	2	50	6.00%
		Toll stations	3	78	9.30%
		Gas stations	4	23	2.70%
		Normal	5	670	79.90%
Environment factors	Road surface	Dry	1	794	94.60%
		Wet	2	45	5.40%
		Spring	1	227	27.10%
		Summer	2	258	30.70%
		Autumn	3	186	22.20%
Environment factors	Season	Winter	4	168	20.00%
		Weekends	1	198	23.60%
		Weekdays	2	641	76.40%
		Sunny	1	202	24.10%
		Cloudy	2	347	41.40%
Environment factors	Weather	Rainy & snow	3	268	31.90%
		Fog & haze	4	22	2.60%
		dawn	1	94	11.20%
		day	2	409	48.70%
		dusk	3	60	7.20%
Environment factors	Visibility	dark	4	276	32.90%
		1	1	503	59.90%
		2	2	276	32.90%
		3	3	31	3.70%
		≥4	4	29	3.50%
Vehicle factors	Type of vehicle	Bus & Truck	1	13	1.55%
		Private cars & Truck	2	42	5.01%
		Non-motor & Truck	3	11	1.31%
		Bus&Private cars&Truck	4	10	1.19%
		Trucks	5	763	90.94%

TABLE 1: Continued.

Factors	Variables	Variables description	Discretization	Frequency	Percentage
Accidents factors	Accident type	Rear-end	1	189	22.50%
		Sideswipe	2	20	2.40%
		Rollover	3	340	40.50%
		Collision	4	145	17.30%
		Vehicle failure	5	145	17.30%
	Accident consequence	Explosion	1	25	3.00%
		Fire	2	96	11.40%
		Spill	3	682	81.30%
		Non-spill	4	36	4.30%
		No injury	1	656	78.19%
	Severity of accident	Severe injury	2	139	16.57%
		Fatality	3	44	5.24%

The construction of the Bayesian network model consists of following steps:

(1) Parameter determination: analyze the risk factors of Hazmat road transportation, and determine the variables needed for modeling (nodes of the Bayesian network), which could be shown in Table 1.

(2) Structure learning: determine the dependencies or independencies relationships between variables (nodes), so that a directed acyclic network structure was constructed.

(3) Parameter learning: based on the given Bayesian network structure, determine the CPT for each node, and the dependence relationship between random variables could be described quantitatively.

4.2. Structure Learning. The scientific network structure needs continuous iterations. At present, there are three methods to construct a Bayesian network structure [30]. (1) Construct the network structure subjectively through experts' knowledge. (2) Determine the network structure objectively via the analysis of data. (3) Construct the network structure based on experts' knowledge and data analysis. The method used in the paper for accident causation analysis is that establishing a preliminary Bayesian network structure based on the model assumption and then the network structure is adjusted with experts' knowledge and data analysis, avoiding the disadvantage of strong subjectivity and enormous amount of data computing. The Bayesian network structure is constructed as shown in Figure 2.

Steps for Building a Bayesian Network Structure

(1) Establish a preliminary Bayesian network structure based on the assumptions of model.

(2) Use Delphi method to determine the relationship between risk factors. In general, there are four possible relationships between variables:

(A) F_i directly lead to F_j , which could be represented as $F_i \longrightarrow F_j$.

(B) F_j directly lead to F_i , which could be represented as $F_j \longleftarrow F_i$.

(C) The relationship between variables cannot be determined, which could be represented as $F_i \longleftrightarrow F_j$.

(D) There is no relationship between variables, which could be represented as $F_i \perp F_j$.

(3) Synthesize results from multiple experts. D-S evidence theory is used to reduce the subjectivity of experts' knowledge, and the correlation between variables could be determined. The Dempster synthesis rule formula could be expressed as

$$M(A) = K \cdot \sum_{A_1 \cap A_2 \cap \dots \cap A_n} m_1(A_1) m_2(A_2) \dots m_n(A_n) \quad \forall A \subseteq \Theta, A \neq \emptyset, A_1, A_2, \dots, A_n \subset \Theta,$$

$$K = \left(\sum_{A_1 \cap A_2 \cap \dots \cap A_n \neq \emptyset} m_1(A_1) m_2(A_2) \dots m_n(A_n) \right)^{-1} \quad (5)$$

$$= \left(1 - \sum_{A_1 \cap A_2 \cap \dots \cap A_n = \emptyset} m_1(A_1) m_2(A_2) \dots m_n(A_n) \right)^{-1}$$

where A represents the possible relationship between variables, m_i represents the mass function, equaling to the expert opinions, and n represents the number of experts.

(4) As the relationship of variables cannot be obtained by Delphi and D-S evidence theory, the mutual information value of variables should be calculated. And the entropy can be expressed as

$$H(F_i) = \sum_{F_i} P(F_i) \log \frac{1}{P(F_i)} = - \sum_{F_i} P(F_i) \log P(F_i) \quad (6)$$

Conditional entropy is a measure of the uncertainty of a random variable F_i under the condition of giving F_j , which can be expressed as

FIGURE 2: The Bayesian network structure for Hazmat road transportation accidents.

$$\begin{aligned}
 H(F_i | F_j) &= \sum_{F_i} P(F_i | F_j) \log \frac{1}{P(F_i | F_j)} \\
 &= - \sum_{F_i} P(F_i | F_j) \log P(F_i | F_j)
 \end{aligned} \quad (7)$$

Before obtaining F_j , the uncertainty of F_i is $H(F_i)$, and after obtaining F_j , the uncertainty of F_i is $H(F_i | F_j)$, so that the difference of $H(F_i)$ and $H(F_i | F_j)$ is considered as the mutual information, which is expressed as

$$\begin{aligned}
 I(F_i, F_j) &= H(F_i) - H(F_i | F_j) \\
 &= \sum_{F_i} \sum_{F_j} P(F_i, F_j) \cdot \log_2 \frac{P(F_i, F_j)}{P(F_i) P(F_j)}
 \end{aligned} \quad (8)$$

4.3. Parameter Learning. There are missing data on Hazmat road transportation accidents; the Expectation- Maximization (EM) algorithm is considered as the effective method to perform the maximum likelihood estimation for a set of parameters θ from the incomplete dataset [31–33]. The EM algorithm starts with randomly assigning a configuration θ^0 for θ by the system. Suppose that θ^t is the outcome after t iterations. The calculation process mainly involved two steps: Expectation Step (E-Step) and Maximization Step (M-Step).

Consider that D_m is missing sample, and X_m is the set of all variables with missing value in the sample D_m . Set $X_m = x_m$, and the complete dataset would be obtained by adding

x_m to D_m . All of the possible result would be considered by EM algorithm due to that X_m may have more possibility, so the weight w_{x_m} is assigned for each possible result by EM algorithm, and the weighted sample could be given by

$$(D_m, X_m = x_m) [w_{x_m}] \quad (9)$$

where $w_{x_m} = P(X_m = x_m | D_m, \theta^t)$, and the weight ranges from 0 to 1.

E-Step: suppose the log-likelihood function of θ based on D^t .

$$\begin{aligned}
 m(\theta | D^t) &= \sum_{t=1}^m \sum_{x_m \in X_m} P(X_m = x_m | D_m, \theta^t) \\
 &\quad \cdot \log P(D_m, X_m = x_m | \theta)
 \end{aligned} \quad (10)$$

where $D = (D_1, D_2, \dots, D_m)$, and $m(\theta | D, \theta^t) = m(\theta | D^t)$ is referred to as the expected log-likelihood function. In the iteration, due to the characteristic of D , which is invariant, the formula could be expressed as

$$\begin{aligned}
 M(\theta | \theta^t) &= m(\theta | D, \theta^t) \\
 &= \sum_{t=1}^m \sum_{x_m \in X_m} P(X_m = x_m | D_m, \theta^t) \\
 &\quad \cdot \log P(D_m, X_m = x_m | \theta)
 \end{aligned} \quad (11)$$

FIGURE 3: The Bayesian network model after parameter learning in Genie 2.0.

M-Step: calculate the value of θ when $M(\theta | \theta^t)$ have reached the maximum:

$$\theta_{ijk}^{t+1} = \begin{cases} \frac{m_{ijk}^t}{\sum_{k=1}^{r_i} m_{ijk}^t} & \sum_{k=1}^{r_i} m_{ijk}^t > 0 \\ \frac{1}{r_i} & \sum_{k=1}^{r_i} m_{ijk}^t \leq 0 \end{cases} \quad (12)$$

where m_{ijk}^t is the sum of sample weights in the dataset D^t .

5. Results

The guidance for the variable selection and classification were followed by the analysis of accident data and previous studies [6, 34–36]. In the paper, sixteen variables are considered as the significant risk factors, as shown in Table 1. There are numerous types of software to establish the Bayesian network efficiently, such as Netica, Genie, Bayes Net Toolbox, and Analytica. In the paper, Genie2.0 (developed by the Decision Systems Laboratory, the University of Pittsburgh) was considered as the effective tool to finish the Bayesian network parameter learning by using EM algorithm, which would make the construction, analysis, and visualization of Bayesian network be performed efficiently, simplifying

the calculation. And the network parameters are repeatedly iterated by using the accident data; the conditions for the termination of calculation are as follows: (1) the variation of the posterior probability for single risk factor is less than 1%; (2) the cumulative variation of posterior probability for the entire network is less than 15%. The results were shown in Figure 3.

5.1. Causal Inference. The Bayesian network could be used to calculate the posterior probability of risk factors under conditions of an accident and obtain the most likely factors or combinations that caused accidents. Set the “explosion” in “accident consequence” as the example to explore the causal inference, and the evidence variable is “explosion”. As shown in Figure 4, the probabilities of risk factors are obtained through the update function of the Genie. And the probability of “autumn” in “season” increases from 22% to 35%; “vehicle failure” (referred as the tire blowout, spontaneous combustion, tanker damage) in “accident type” increases from 17% to 37%; the quantity of Hazmat increases from 8% to 20% for the category of more than 40 tons; “flammable liquids” in “Hazmat categories” increase from 51% to 65%; and the explosives increase from 3% to 8%. These findings mean that, in the absence of other evidences, the most probable reasons for “explosion” are vehicles carrying

FIGURE 4: Posterior probability when the variable is “explosion”.

flammable liquids, larger quantity of Hazmat, vehicle failure, and transporting in autumn.

In addition, if the “fatality” in the “severity of accident” is considered as the evidence variable, the probability change of “total vehicle involved accident” could be obtained. The probability of “three” increases from 4% to 11%, and “more than three” is increasing from 3% to 9%. This may be explained by the fact that the importance of 3 or more vehicles in an accident leads to the severe accident being higher than less vehicles. Moreover, as for the accident consequence, the probability of “spill” decreases; meanwhile the “explosion” (3% to 6%) and fire (11% to 18%) have increased. Due to the special characteristic of Hazmat, explosion and fire would cause a larger area affected and can easily result in casualties, especially in the urban road and higher population densities [26].

5.2. Accident Prediction. Based on the bidirectional reasoning, not only could the Bayesian network model obtain the risk factors or the combination caused accidents, but also the probability of accidents could be calculated under the risk factors or combination, for example, in Genie, setting the “speeding” in “driver behavior” as an evidence variable, meaning that the status of evidence variable is considered as 100%. As can be seen from Figure 5, the probability of

“rear-end” in “accident type” is found to increase from 27% to 42%, indicating that the drivers’ speeding could be more prone to lead to rear-end accidents. This is because the vehicle is difficult to control under the condition of speeding, and the braking time is longer. And previous studies have shown that driving behavior could significantly affect the severity of traffic accidents [37–39].

As shown in Figure 6, in addition to “speeding”, it is assumed that the transportation route is on low-class roads; that is, “Group four” in the “accident location” is considered as the evidence variable, and the probability of the entire network is automatically updated. It can be found that the probability of “rollover” in “accident type” further increases from 42% to 97%. This finding shows that “driver behavior” and “accident location” would affect the probability of “rollover” accident on different degrees. Therefore, when the driver is speeding on low-class roads, the more attention should be paid on the rollover accident.

6. Discussion and Conclusions

6.1. Hazmat Factors. Flammable liquids have the highest posterior probability (0.51) and would easily result in explosion. This could be explained by that increasing demand for the flammable liquid and decreasing reliability of transporting

FIGURE 5: Accident prediction when the evidence variable is “speeding”.

flammable liquids due to the single-mode packaging. The quantity of Hazmat transported would significantly affect the severity of accident. The larger the quantity of Hazmat transportation, making it not easy to control the emergency [40]. Moreover, the larger quantity of Hazmat transportation is prone to the serious consequences, such as explosion and spill, threatening people's health and environment [10].

6.2. Driver Factors. Previous studies have shown the relationship between driver's age and the severity of accidents [27, 41, 42]. According to the model results, the younger driver (less than 35) would be more prone to inappropriate driving behavior, which indicates the need for carrying out education programs and training for younger drivers. Tavris et al. [43] also found that younger drivers were much more likely to be involved in severe and fatal accidents. As for the driving behavior, speeding is more likely to lead to rollover accident, especially on the low-class road. This could ascribe the small amount of lanes and the road condition defects on low-class roads, and the speeding would make Hazmat slosh or move around inside the tank, which can constantly shift the vehicle weight, leading to vehicle to rollover due to the off balance [44, 45].

6.3. Location Factors. The model results show that “Group one” (the posterior probability is 0.43) and “Group two” (the posterior probability is 0.40) in “accident location” are likely to be associated with severe accidents, which could be attributed by the combination of higher average speed and larger speed dispersion. More importantly, “Group one” and “Group two” roads are considered as the major transport corridors for Hazmat [10, 46]. In addition, some special sections would also considered as the significant risk factors; this could be explained by the fact that there are more interference factors (such as line of sight, pedestrians, and signal lights) at intersections and the greater potential explosion risk around the gas stations [47].

6.4. Environment Factors. Hazmat road transportation accidents would easily occur at summer (the posterior probability is 0.31), which is attributed to the characteristic of Hazmat, such as flammable and explosive. And the posterior probability of accidents occurring at weekdays is 0.76, which could be explained by that freeway could be toll-free on important holidays, resulting in significant increase of traffic volume, which could decrease the speed of vehicles. Moreover, Hazmat transportation vehicles were not allowed to drive on freeway (Pan, 2013). Weather is a significant factor for the Hazmat transportation, with cloudy having the

FIGURE 6: Accident prediction when the evidence variable are "speeding" and "Group four".

highest posterior probability (0.41) followed by rainy (0.32). This could be ascribed that the driver's mood and visual would be decreased in cloudy and rainy, and the rainy would lower the friction coefficient of roads due to the thin film of water existing between the road surface and tires, which could make the road slippery, increasing the braking distance effectively [48, 49]. Regarding visibility, daytime has the highest posterior probability (0.49), and the dark is 0.33. This is because most transportation corporations are more likely to transport Hazmat at daytime in China [50]. In addition, poor visibility at night would make drivers tired, resulting in driver fatigue, especially from 11:00 pm to 3:00 am [51]. In the sample of accident data, drivers are more prone to fatigue status accounting for 62% of total accidents from 7:00 pm to 4:59 am.

6.5. Vehicle Factors. As for the total vehicles involved in accident, "more than three" would easily result in higher severity of accidents. And the private car involved in accident would cause the severe accident. Two reasons could explain these findings: one is that more vehicles would cause more people involved in accidents, resulting in more people injured; another one is the disparity in mass and speed of trucks compared to other vehicles. In case of an accident,

lighter vehicles (such as private cars) usually absorb the greatest part of the kinetic energy and suffer from more severe injury.

6.6. Accident Factors. Many studies have shown the significant relationship of accidents type and severity, indicating that the rollover accident is associated with the higher severity of accident [16, 44]. The Bayesian network results show that rollover accident has the highest posterior probability (0.41). The reason could be that Hazmat sloshing or moving around inside the tank can constantly shift the vehicle weight, making the vehicle off balance, causing the transportation vehicle to roll over, especially during abrupt evasive maneuvers or turning the vehicle [10]. In addition, as for the consequence of accident, the posterior probability of spill could reach up to 0.81, threatening human health and environment. The result could be explained by that Hazmat releasing could immediately result in poisoning and suffocation, which is difficult for people on-site to escape quickly, resulting in severe and fatality accident [50].

In summary, the occurrence of Hazmat road transportation accidents is unexpected, random, dangerous, and potential. Frequent accidents imply that it is necessary to explore risk factors by using accident mechanism. Bayesian network

is the effective method to deal with uncertainties, which exhibit the potential hierarchical relation by the Directed Acyclic Graph. In the paper, the Bayesian network was developed based on experts' knowledge and modified based on the Hazmat road transportation accident data (N=839) in China. The Bayesian network structure was established by using Genie 2.0, and the results of network structure model reveal the influence of risk factors resulting in accidents and the relationship among risk factors. The study shows that the posterior probability of the Bayesian network could provide effective method for finding the important factors and the factors combination of accidents. These findings could provide theoretical guidance, which could help transportation corporations and government departments take necessary measures to reduce the frequency of Hazmat accidents. More importantly, it must be noted that the aforementioned results were obtained by analyzing the data sample collected from State Work Accident Briefing System and Hazardous Chemical Accidents Communications, which could be existing limitations. As for the further studies, the conclusions should be more generalizable if the dataset had larger size of sample and accidents from multiple states.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study has been supported by projects of the National Natural Science Foundation of China (no. 71671127).

References

- [1] K. G. Zografos and K. N. Androustopoulos, "A decision support system for integrated hazardous materials routing and emergency response decisions," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 6, pp. 684–703, 2008.
- [2] F. G. Cordeiro, B. S. Bezerra, A. S. P. Peixoto, and R. A. R. Ramos, "Methodological aspects for modeling the environmental risk of transporting hazardous materials by road," *Transportation Research Part D: Transport and Environment*, vol. 44, pp. 105–121, 2016.
- [3] Federal Motor Carrier Safety Administration (FMCSA), "Large Truck and Bus Crash Facts 2014," <https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/docs/Large-Truck-and-Bus-Crash-Facts-2014%28April%202016%29.pdf>, 2016.
- [4] J. Zhao L, "Risk Analysis of Dangerous Chemicals Transportation," *Systems Engineering-Theory Practice*, vol. 27, no. 12, pp. 117–122, 2007.
- [5] W. Y. Hua and A. Tong P, "Risk Analysis on Road Transport System of Dangerous Chemicals," *China Safety Science Journal*, vol. 15, no. 2, pp. 8–12, 2005.
- [6] L. Zhao J, P. Wu, and K. Xu, "Statistic analysis and countermeasures on dangerous chemical accidents in China," *China Safety Science Journal*, 2009.
- [7] J. Yang, F. Li, J. Zhou, L. Zhang, L. Huang, and J. Bi, "A survey on hazardous materials accidents during road transport in China from 2000 to 2008," *Journal of Hazardous Materials*, vol. 184, no. 1-3, pp. 647–653, 2010.
- [8] T. Kauppinen and J. Rantanen, "Work and Health Country Profiles and National Surveillance Indicators in Occupational Health and Safety," *Applied Occupational & Environmental Hygiene*, vol. 17, no. 9, p. 603, 2002.
- [9] P. Haastrup and L. Brockhoff, "Severity of accidents with hazardous materials. A comparison between transportation and fixed installations," *Journal of Loss Prevention in the Process Industries*, vol. 3, no. 4, pp. 395–405, 1990.
- [10] X. Shen, Y. Yan, X. Li, C. Xie, and L. Wang, "Analysis on Tank Truck Accidents Involved in Road Hazardous Materials Transportation in China," *Traffic Injury Prevention*, vol. 15, no. 7, pp. 762–768, 2014.
- [11] K. Fang, G. Y. Ke, and M. Verma, "A routing and scheduling approach to rail transportation of hazardous materials with demand due dates," *European Journal of Operational Research*, vol. 261, no. 1, pp. 154–168, 2017.
- [12] F. Bird and G. Germain, *Practical Loss Control Leadership*, International Loss Control Institute, Duluth, GA, USA, Revised edition, 1990.
- [13] J. D. Lemp, K. M. Kockelman, and A. Unnikrishnan, "Analysis of large truck crash severity using heteroskedastic ordered probit models," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 370–380, 2011.
- [14] M. Uddin and N. Huynh, "Factors influencing injury severity of crashes involving HAZMAT trucks," *International Journal of Transportation Science and Technology*, vol. 7, no. 1, pp. 1–9, 2018.
- [15] E. K. Adanu and S. Jones, "Effects of Human-Centered Factors on Crash Injury Severities," *Journal of Advanced Transportation*, vol. 2017, no. 1528, pp. 1–11, 2017.
- [16] 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 & Prevention*, vol. 43, no. 1, pp. 402–411, 2011.
- [17] R. O. Mujalli and J. De Oña, "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks," *Journal of Safety Research*, vol. 42, no. 5, pp. 317–326, 2011.
- [18] L. J. Zhao, X. L. Wang, and Y. Qian, "Analysis of factors that influence hazardous material transportation accidents based on Bayesian networks: a case study in China," *Safety Science*, vol. 50, no. 4, pp. 1049–1055, 2012.
- [19] C. Chen, G. Zhang, X. C. Liu et al., "Driver injury severity outcome analysis in rural interstate highway crashes: a two-level Bayesian logistic regression interpretation," *Accident Analysis & Prevention*, vol. 97, pp. 69–78, 2016.
- [20] J. Chen, M. Zhang, S. Yu, and J. Wang, "A Bayesian Network for the Transportation Accidents of Hazardous Materials Handling Time Assessment," *Procedia Engineering*, vol. 211, pp. 63–69, 2018.
- [21] M. Deublein, M. Schubert, B. T. Adey, J. Köhler, and M. H. Faber, "Prediction of road accidents: a Bayesian hierarchical approach," *Accident Analysis & Prevention*, vol. 51, pp. 274–291, 2013.

- [22] C. Xu, W. Wang, P. Liu, and Z. Li, "Calibration of crash risk models on freeways with limited real-time traffic data using Bayesian meta-analysis and Bayesian inference approach," *Accident Analysis & Prevention*, vol. 85, pp. 207–218, 2015.
- [23] C. Tang, Y. Yi, Z. Yang, and J. Sun, "Risk analysis of emergent water pollution accidents based on a Bayesian Network," *Journal of Environmental Management*, vol. 165, pp. 199–205, 2016.
- [24] X. Zou and W. L. Yue, "A Bayesian Network Approach to Causation Analysis of Road Accidents Using Netica," *Journal of Advanced Transportation*, 2017.
- [25] R. Bubbico, S. Di Cave, B. Mazzarotta, and B. Silveti, "Preliminary study on the transport of hazardous materials through tunnels," *Accident Analysis & Prevention*, vol. 41, no. 6, pp. 1199–1205, 2009.
- [26] R. A. Garrido and A. C. Bronfman, "Equity and social acceptability in multiple hazardous materials routing through urban areas," *Transportation Research Part A: Policy and Practice*, vol. 102, pp. 244–260, 2016.
- [27] T. Usman, L. Fu, and L. F. Miranda-Moreno, "Analysis of factors affecting winter collision severity," in *Meeting of the Transportation Research Board*, Washington, DC, USA, 2013.
- [28] F. Chen and S. Chen, "Injury severities of truck drivers in single- and multi-vehicle accidents on rural highways," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1677–1688, 2011.
- [29] S. Islam, S. L. Jones, and D. Dye, "Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama," *Accident Analysis & Prevention*, vol. 67, pp. 148–158, 2014.
- [30] Q. Xiaohu, L. Li, and Z. Ying, "A traffic accident prediction method based on Bayesian network model," *Computer Simulation*, vol. 22, no. 11, pp. 230–232, 2005.
- [31] S. L. Lauritzen, "The EM algorithm for graphical association models with missing data," *Computational Statistics & Data Analysis*, vol. 19, no. 2, pp. 191–201, 1995.
- [32] F. V. Jensen and T. D. Nielsen, "Bayesian Networks and Decision Graphs," *Technometrics*, vol. 50, no. 1, p. 362, 2012.
- [33] J. Zhou, W. Xu, X. Guo, and J. Ding, "A method for modeling and analysis of directed weighted accident causation network (DWACN)," *Physica A: Statistical Mechanics and its Applications*, vol. 437, pp. 263–277, 2015.
- [34] J. Abellán, G. López, and J. de Oña, "Analysis of traffic accident severity using decision rules via decision trees," *Expert Systems with Applications*, vol. 40, no. 15, pp. 6047–6054, 2013.
- [35] R. O. Mujalli, G. López, and L. Garach, "Bayes classifiers for imbalanced traffic accidents datasets," *Accident Analysis & Prevention*, vol. 88, pp. 37–51, 2016.
- [36] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Analysis & Prevention*, vol. 108, pp. 27–36, 2017.
- [37] L. Fridström and S. Ingebrigtsen, "An aggregate accident model based on pooled, regional time-series data*," *Accident Analysis & Prevention*, vol. 23, no. 5, pp. 363–378, 1991.
- [38] G. F. Ulfarsson, S. Kim, and E. T. Lentz, "Factors affecting common vehicle-to-vehicle collision types: Road safety priorities in an aging society," *Transportation Research Board*, vol. 1980, no. 1, pp. 70–78, 2006.
- [39] G. Zhang, K. K. W. Yau, X. Zhang, and Y. Li, "Traffic accidents involving fatigue driving and their extent of casualties," *Accident Analysis & Prevention*, vol. 87, pp. 34–42, 2016.
- [40] A. Vorster, "Transporting dangerous goods worldwide: materials handling logistics," *South African Pharmaceutical Cosmetic Review*, 2015.
- [41] A. P. Jones and S. H. Jørgensen, "The use of multilevel models for the prediction of road accident outcomes," *Accident Analysis & Prevention*, vol. 35, no. 1, pp. 59–69, 2003.
- [42] S. Kaplan and C. G. Prato, "Risk factors associated with bus accident severity in the United States: a generalized ordered logit model," *Journal of Safety Research*, vol. 43, no. 3, pp. 171–180, 2012.
- [43] D. R. Tavis, E. M. Kuhn, and P. M. Layde, "Age and gender patterns in motor vehicle crash injuries: Importance of type of crash and occupant role," *Accident Analysis & Prevention*, vol. 33, no. 2, pp. 167–172, 2001.
- [44] C. S. Duncan, A. J. Khattak, and F. M. Council, "Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions," *Transportation Research Record*, no. 1635, pp. 63–71, 1998.
- [45] A. Montella, L. Imbriani, and F. Mauriello, "Factors Contributing to Run-off-the-Road Severe Crashes," in *Proceedings of the Transportation Research Board 94th Annual Meeting*, Washington, DC, USA, 2015.
- [46] A. Oggero, R. M. Darbra, M. Muñoz, E. Planas, and J. Casal, "A survey of accidents occurring during the transport of hazardous substances by road and rail," *Journal of Hazardous Materials*, vol. 133, no. 1–3, pp. 1–7, 2006.
- [47] R. Tay and S. M. Rifaat, "Factors contributing to the severity of intersection crashes," *Journal of Advanced Transportation*, vol. 41, no. 3, pp. 245–265, 2010.
- [48] Q. Lin and W. A. Nixon, "Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2055, no. 2055, pp. 139–146, 2008.
- [49] P. Konstantopoulos, P. Chapman, and D. Crundall, "Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 827–834, 2010.
- [50] H.-D. Zhang and X.-P. Zheng, "Characteristics of hazardous chemical accidents in China: A statistical investigation," *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 4, pp. 686–693, 2012.
- [51] J. Pahukula, S. Hernandez, and A. Unnikrishnan, "A time of day analysis of crashes involving large trucks in urban areas," *Accident Analysis & Prevention*, vol. 75, pp. 155–163, 2015.