

Research Article

Bayesian Network-Based Knowledge Graph Inference for Highway Transportation Safety Risks

Luo Wenhui , Cai Fengtian, Wu Chuna, and Meng Xingkai

Research Institute of Highway Ministry of Transport, Beijing 100088, China

Correspondence should be addressed to Luo Wenhui; 815345591@163.com

Received 16 October 2020; Revised 20 January 2021; Accepted 18 February 2021; Published 5 March 2021

Academic Editor: Songtao Lv

Copyright © 2021 Luo Wenhui 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.

Accurate inference of knowledge about highway transportation safety risks forms a crucial aspect of building a knowledge graph. Based on the data related to highway transportation accidents, this study has developed a Bayesian network model. The initial identification of the network nodes is through expert scoring. The network structure is then constructed by utilizing the prior expert knowledge and K2 greedy search algorithm. Later, the network parameters are trained via the expectation-maximization (EM) algorithm. Finally, knowledge about highway transportation safety risks is inferred using the junction tree algorithm. A comparison is made between the trained conditional and actual probabilities during the network parameter training to verify the validity of the proposed model that accords with expert experience, thereby proving the model validity. Further, its main “causal chain” is inferred to be an improper emergency response-human failure-accident occurrence, where the probability of driver failure is 82%, and the probability of accident occurrence is 68% by taking “a certain road traffic accident” as an example. There is consistency between the inference results and the actual accident sequence that suggests the effectiveness of the proposed knowledge inference method.

1. Introduction

Knowledge graphs that contain abundant knowledge from various areas have become a hot topic across various research fields, with the rapid explosion of network information. As a reliable source of all-round information and underlying support, they have performed remarkably in the safety risk management realms, such as medicine and finance [1, 2]. An important area of mass casualty monitoring has always been highway transportation. The top priority of safe highway transportation is safety risk management, and, for this, building a complete knowledge graph of highway transportation safety risk management is necessary for achieving risk control. Crucial to the completion and expansion of knowledge graphs, the knowledge inference technique—among key techniques of building knowledge graphs, such as knowledge extraction, fusion, and inference—deduces new knowledge or identifies errors in a knowledge graph by using certain methods based on the existing knowledge in the graph. The ultimate aim for

highway transportation safety risk management is to predict the risks faced in transportation in advance that can be understood as a risk inference problem in knowledge graphs. Extensive research achievements concerning knowledge inference have been made to date:

- (1) Inferences based on logical rules: The logical rule-based reasoning process divides knowledge into rules and facts. There are forward, backward, and bidirectional reasoning approaches to inference methods. The Markov logic network (MLN), where the common inference tasks include the maximum posterior probability estimation and the conditional probability reasoning, is a particularly fruitful model. Chen et al. [3] applied the MLN to risk management in response to information security issues. First, they constructed an infrastructure dependency network by the expert method that was then converted into an MLN for the calculation of node weights via a learning algorithm. The proposed model's

innovativeness is in the introduction of new network threat factors and the calculation of their possible impacts on the entire system. Yang et al. [4] proposed an MLN-based method for joint sentiment analysis of sentences to address the inadequate utilization of contextual information with the existing knowledge inference methods, as well as the cross-domain connection problem of sentence information, in the context of knowledge graph inference. They found through experimentation that MLN-based knowledge inference achieved rather desirable results. Liu et al. [5] put forward an ensemble learning-based MLN model, as well as its learning algorithm, in response to the difficulty of MLN in inferring large-scale data, respecting the modification and improvement of MLN models. They conducted a knowledge extraction experiment in Google's large-scale corpus using this method. As experiments proved, their method had higher precision and recall than the pipelining approach.

- (2) Bayesian network-based inference models: Bayesian inference is a process of calculating posterior probability, based on conditional and prior probabilities. Relating to risk management, Lu et al. [6] proposed a Bayesian network-based model for flood risk inference aiming at the problem of flood risk misreporting. Here, expert scoring identified the Bayesian network, and parameter learning was performed with the Monte Carlo model. These helped in analyzing the changes in flood risk under one- and two-factor uncertainties. The Bayesian network built by them allowed bidirectional reasoning and probability distribution inference of arbitrary nodes. Experiments proved that their model was practically applicable to the risk assessment and control in reservoirs. Based on a qualitative weighted Bayesian network, Yin et al. [7] put forward a human factor inference model for coalmine accidents in order to address the analytical deficiency of human factors in coal mine accident analysis. The Bayesian network, based on typical accident cases, was constructed using the Human Factors Analysis and Classification System (HFACS) model, and the weight magnitudes between network nodes were obtained by the expert method. The mutual logical relationships between the human factors were reflected rather accurately by their model. With a view to combing the evolution process of such events, Xia et al. [8] proposed a dynamic Bayesian network-based unconventional scenario inference model for sudden disaster events. They constructed a dynamic Bayesian network with nodes such as the scenario condition, handling goals, handling measures, and intrinsic variables. The prior probabilities were designated for root nodes, while for variables without parent nodes, the expert method was employed to determine conditional probabilities. In the end, inference of accident scenario evolution was executed by introducing the "July 16 Oil Depot Explosion and Fire" that yielded correct inference results. Seyed Hassani et al. [9] developed a Bayesian network related to inference on knowledge graphs by taking into consideration some hidden or ignored information in complex social networks. They identified the nodes of the Bayesian network, such as comments, avatar information, ensemble photos, or interactive information, and using the collected data, trained its parameters. The effectiveness of their algorithm was tested on Facebook that found the model is highly accurate in finding information between users. Rajabi and Ataie-Ashtiani [10] proposed a fuzzy Bayesian inference method, in terms of model modification and improvement, to address the unavailability of parameter training data in conventional Bayesian inference. The method fused the information provided by experts with the Bayesian network model. They also developed an algorithm for solving the model. Computational results were compared with the Markov Chain Monte Carlo- (MCMC-) based algorithm that proved the effectiveness of their model and algorithm.
- (3) Machine learning-based models: Xie et al. [11] proposed a model that targeted the knowledge completion issue with knowledge graphs based on bag-of-words and convolutional neural network, where the bag-of-words model was used for the vector representation of texts, and the convolutional neural network was responsible for classifying and inferring word relationships. They confirmed the validity of their model through experimentation. A novel convolution-based model was proposed by Annervaz et al. [12] that extracted relevant prior knowledge from the graphs via an attention mechanism. Experiments on public datasets demonstrated that their method was effective in enhancing the performance of deep learning models. Godin et al. [13] developed a ternary reward framework to cope with the incorrect answering problem in the existing reinforcement learning-based question answering systems that established a new evaluation criterion by setting different rewards for wrong answers and no answers, thereby achieving better evaluation of model effectiveness. All of the above—completion, fast retrieval, and answering problems related to knowledge graphs—can be understood as the scenario applications of knowledge inference.
- (4) Hybrid models: Jiang et al. [14] developed a method for representing knowledge based on weighted knowledge graphs that was then combined with the probabilistic graphical model to establish a medical diagnosis knowledge network. A path sorting algorithm-based random walk model was proposed by Liu et al. [15] that performed knowledge inference on the semantics of sentences regarding the inverse relationship from object to subject. An exhaustive

search was avoided by introducing a random sampling mechanism, and the effectiveness of their knowledge inference model was experimentally verified.

The advantage of logical rule-based reasoning is the intuitive inference procedure that can reflect prior knowledge, and its disadvantage is the difficulty in obtaining rules that lead to error accumulation. The advantage of deep learning-based inference models is a powerful reasoning capability, while their disadvantages are insufficient interpretability and data limitations. Predicting the probability of accident occurrence and estimating the “causal chain” are the foremost tasks of knowledge graph inference for highway transportation safety risks. On the other hand, we consider the task of inferring a knowledge graph about highway transportation safety risks from the Bayesian networking perspective, given the limited data collected in this study. Expert scoring identifies the nodes of the Bayesian network. The node parameters are learned via the EM algorithm, and knowledge about highway transportation safety risks is inferred with the junction tree model.

The contributions of this paper include the following:

- (1) A knowledge inference framework for road transportation risks is proposed, and annotated datasets are provided for the research field.
- (2) During the identification of network nodes, a network architecture identification method based on expert scoring combined with the K2 algorithm is proposed that ensures the incorporation of expert knowledge about road transportation risks by the network architecture.
- (3) Based on the Bayesian network created in this paper, the probability distributions of accident occurrence are inferred under multiple factors, including the driver, vehicle, road, environment, and management.

The remainder of this paper is organized as follows:

Section 2 puts forward a model framework for risk knowledge inference, Section 3 builds a Bayesian network-based knowledge inference model for road transportation risks, Section 4 infers knowledge about road transportation risks, and Section 5 concludes the study.

2. Creation of Risk Knowledge Inference Model

The Bayesian network may be summarized as a probabilistic inference network that is based on the Bayesian formula, where the nodes represent the random variables, and the directed edges between nodes represent the internode causality. Each node has a probability distribution. Assuming a given Bayesian network $G(S, P)$ consists of two parts, with S being a directed acyclic graph containing all nodes and P being a collection of conditional probability distribution tables. Construction of Bayesian network for inferring highway transportation safety risk knowledge comprises four steps, namely, (i) the network node identification, (ii) the network architecture identification, (iii)

the determination of node conditional probabilities, and (iv) the inference of safety risk knowledge for highway transportation. Overall, the modeling procedure incorporates the expert method and machine learning. Figure 1 presents the model framework proposed in this study.

2.1. Network Node Identification. At this stage, expert scoring identifies the network nodes. Prior expert knowledge is collected according to five categories of “drivers, vehicles, roads, environment, and management,” thereby constructing a matrix for node information acquisition as shown in the following expression:

$$E_f = (e_{ij})_{m \times n}, \quad (1)$$

where e_{ij} ($0 \leq e_{ij} \leq 1$) denotes the confidence that the i th expert has, regarding the causality between the j th risk and the f th category, $i \in [1, n]$, $j \in [1, m]$, $f \in [1, 5]$. Matrix E_f is traversed, and if $e_{ij} \geq a$, it indicates that the j th risk and the f th category are casually correlated.

2.2. Network Architecture Identification. The network architecture identification is divided into two steps, given the limited amount and quality of training data in this study, namely, the expert method based network architecture and the network architecture modification by data-based learning. For network architecture learning, the greedy search and conditional restriction algorithms are generally used, of which the methods based on greedy space search are adapted to minimum data size. K2 algorithm optimizes the search capability on the basis of ordinary greedy search by deleting the redundant edges. Hence, for network architecture learning, this study adopts the K2 algorithm, whose core idea is to find a network structure with high score functions. The K2 scoring method is described in the following formula[16]:

$$F\left(\frac{D}{G}\right) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)} \prod_{k=1}^{r_i} N_{ijk}!, \quad (2)$$

where G represents the Bayesian network architecture, D stands for the training dataset, i denotes the i th node, n denotes the number of nodes, j denotes the j th parent node of the current node, q_i is the number of parent nodes of the current node, k represents the k th value of the current node, r_i represents the number of possible values of the current node, and N_{ijk} denotes the number of examples in the training dataset D that corresponds to the k th value of current node and the j th value of parent node, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

For the K2 algorithm, its pseudocode is as in Algorithm 1 [16].

2.3. Network Parameter Learning. During network parameter learning, commonly used algorithms include the maximum likelihood estimation (MLE) and the Bayesian estimation and expectation-maximization (EM) algorithms, of which the MLE algorithm is generally suitable for scenarios with large data size. When the sample size is small and

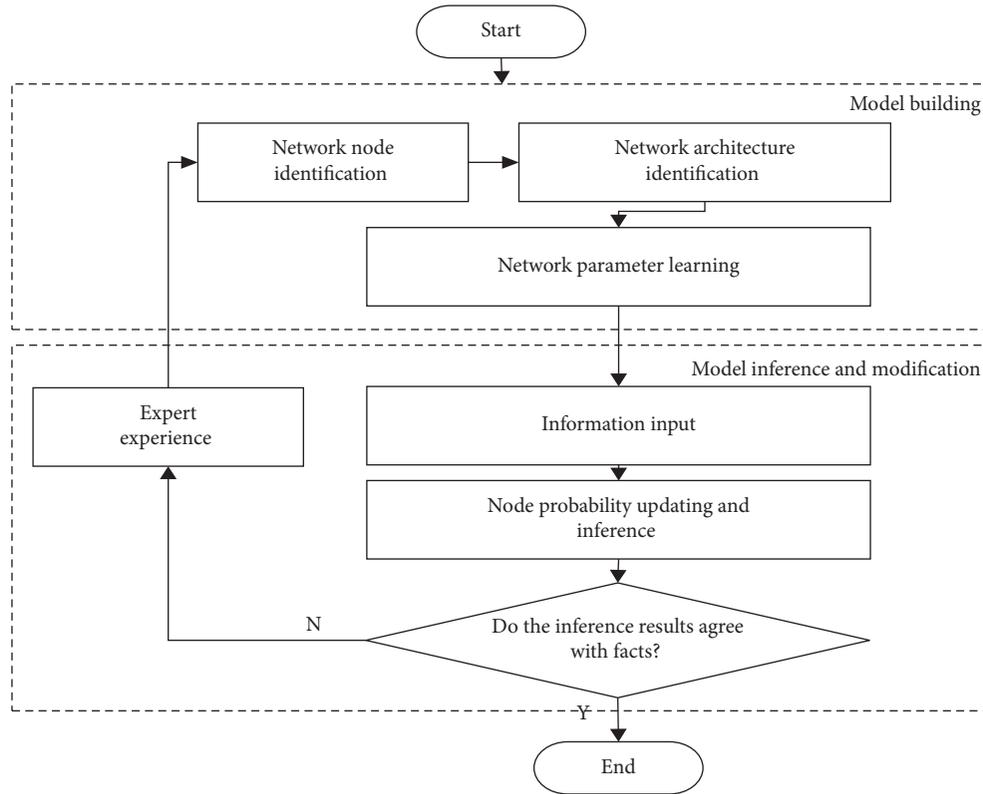
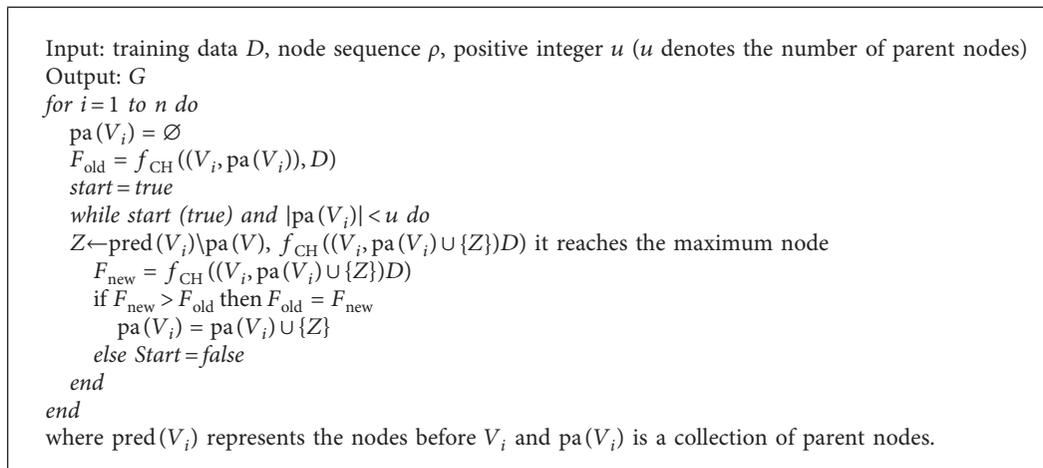


FIGURE 1: Flow chart of safety risk knowledge inference for highway transportation.



ALGORITHM 1

the prior probability is hardly attainable, the use of the EM algorithm yields a good learning effect. Hence, this study adopts the EM algorithm to learn network parameters. Its algorithmic procedure can be divided into E-step and M-step (see Algorithm 2) [17].

2.4. Network Inference. Network inference is the ultimate objective of this study. All joint probability distributions for nodes are obtained through the network parameter learning stage. The probability (result) distributions for a set of query

variables at the network inference stage are computed under exact values given for a set of evidence variables (causes).

Among common algorithms for network inference, such as junction tree and variable elimination, the former is adopted herein for network inference, owing to its easy-to-understand and accurate inference advantages. The junction tree algorithm is divided into four phases:

- (1) In the initial phase, the built Bayesian network is modularized, and the parent nodes of each node are connected with undirected edges

Input: observation variable Y , hidden variable Z , joint distribution $P(Y, Z|\theta)$, conditional probability distribution $P(Z|Y, \theta)$.

Output: model parameter θ

- (1) Assignment of the initial values for the model parameter;
- (2) E-step: θ_i is the value of the model parameter after the i th iteration, the calculated function of expectation on $i+1$ th iteration, $Q(\theta, \theta_i) = E_{Z \sim P(Z|Y, \theta_i)} [\log P(Y, Z|\theta)] = \sum_Z P(Z|Y, \theta_i) \log P(Y, Z|\theta)$
- (3) M-step: find $Q(\theta, \theta_i)$ that maximizes θ , determine the estimated value θ_{i+1} of the $i+1$ th iteration.
- (4) Repeat steps 2-3 until model converges.

ALGORITHM 2

- (2) The Bayesian network is converted into an undirected graph, where each arrow is replaced by an edge
- (3) The graph is triangulated, and an edge is added to the variables in the same loop
- (4) The triangulated graph is converted into a clustering tree, where each node represents factors in the variable subset

3. Bayesian Network-Based Knowledge Inference Model for Highway Transportation Safety Risks

The purpose of safety risk knowledge inference for road transportation is to achieve the “prior” control of unsafe risk factors. However, these risk factors have strong uncertainties, while Bayesian networks have provided preferable solutions to complex and uncertain problems [16]. Hence, following this idea, a Bayesian network-based knowledge inference model for road transportation risks, with which the inference problems of safety risk knowledge for road transportation are solved, has been proposed in this paper.

3.1. Network Node Identification. Since the focus of this research is on solving the problem of risk knowledge inference, our team has listed a total of 28 risk sources involving “drivers, vehicles, roads [18, 19], environment, and management” based on the description of safety risks in the accident reports over the years that have been accomplished by taking specific risks as the network nodes. We developed a quantitative questionnaire and distributed it among ten experts within the field. When $e_{ij} \geq 0.7$, a causality is considered to exist between risk and category. A risk has been chosen as an effective node if more than six experts confirmed its causality with the category. By employing expert scoring, the network nodes for road transportation risk knowledge have been constructed according to formula (1) as detailed in Table 1 that was done by selecting a total of 26 observable nodes and 5 virtual nodes.

3.2. Network Architecture Identification. A combination of the K2 algorithm and expert scoring has been used to identify the Bayesian network architecture. Initially, with the aid of GeNIe 2.0 software, learning of network structure was completed based on the database using the K2 algorithm.

Figure 2 displays the learned network architecture. Afterward, modification of the network structure was performed based on expert experience. Figure 3 illustrates the finalized network architecture.

As is clear from Figure 2, the network architecture presents a strong progressive causal relationship that was divided into three layers vertically. The relationship between nodes is clear, and there are no connecting lines between unrelated factors. The nodes are divided into two parts, including observable nodes and virtual nodes. The observable nodes are the nodes with actual data description, while the virtual nodes are the nodes added for the integrity of the network.

As is clear from Figure 3, the hierarchy of the network is not obvious, because through the learning of the algorithm, the implicit relationship between the nodes is further revealed, and the overall structure of the network satisfies the requirement of the cognition.

3.3. Network Parameter Learning. Learning of various network parameters was accomplished via the EM algorithm in GeNIe 2.0. The learning process can be summarized into five steps:

- (1) Assignment of initial values for various root nodes.
- (2) Establishment of correspondences between node names and IDs.
- (3) Import of datasets.
- (4) Matching of data with network nodes.
- (5) Parameter learning. The learning results are listed in Table 2, where N implies normal and F implies failure.

Table 2 clearly shows that, under identical conditions, improper emergency response and traffic violations exhibit the highest posterior probabilities among the driver factors, followed by fatigue driving, distracted driving, and drunk driving, while inexperienced driving shows the lowest posterior probability. These results are in line with expert experience and perception.

4. Inference of Safety Risk Knowledge for Highway Transportation

Using the junction tree algorithm, Bayesian network inference is implemented on highway transportation safety risks. According to the Bayesian network architecture,

TABLE 2: Driver parameter learning results.

Risk	Status							
Drink and driving	N	N	N	N	N	N	N	F
Fatigue driving	N	N	N	N	N	N	F	N
Distracted driving	N	N	N	N	N	F	N	N
Traffic violated	N	N	N	N	F	N	N	N
Improper emergency response	N	N	F	N	N	N	N	N
Unfamiliar driving	N	F	N	N	N	N	N	N
Driver factors	N	0.899	0.8125	0.3053	0.339638	0.558	0.506	0.5666
	F	0.101	0.1875	0.6946	0.66036	0.441	0.493	0.4333

impact analysis is carried out keeping in mind the following aspects:

- (1) Effects of the driver, vehicle, road, environment, and management factors on the occurrence of accidents.
- (2) Effects of drunk driving, fatigued driving, distracted driving, traffic violations, improper emergency response, and inexperienced driving on the driver risks.
- (3) Effects of the braking system, steering system, light signal system, tires, truck overload, and unbalanced load on the vehicle risks.
- (4) Effects of roadsides, bridges and tunnels, alignment, intersections, pavements, signs and markings, sight distance, and safety protection facilities on the road risks.
- (5) Effects of traffic accidents, severe weather, natural disasters, and night lighting on the traffic environment risks.
- (6) Effects of dynamic monitoring, regulations, and private contracting on the management risks.

4.1. Data Description and Model Evaluation. Our team collected 600 reports on road transportation accidents from the safety management websites that occurred between 2012 and 2019. These accident reports have been sorted into three categories. The risk factors of each accident are marked either as 0 or as 1, where 1 indicates that the assessed risk is an accident-causing factor, and 0 indicates that the assessed risk is not an accident-causing factor. The model evaluation criteria are as follows:

- (a) Accordance with expert experience and knowledge of the analysis of the effects of relevant factors on accident occurrence.
- (b) Conformance to the accident state description of the accident reports in the case analysis.

4.2. Inference Analysis of Relevant Influencing Factors

4.2.1. Effects of Relevant Factors on the Occurrence of Accidents. The probability distributions of accident occurrence, based on the built Bayesian network model for highway transportation safety risks and the inference results,

are depicted in Figure 4 under human, vehicle, road, environment, and management factors.

From Figure 4(a), it is clear that, among the causes of accidents, the accident probability is the highest for the driver factors, with an inferential probability of 0.899, followed by the management factors, with an inferential probability of 0.485. In contrast, road factors constitute the least probable factors, primarily because they are indirect causes in general. According to Figure 4(b), the probability of accidents resulting from the failure of drivers, vehicles, and management is quite high, while that caused by the failure of environmental and road factors is rather low.

4.2.2. Effects of Relevant Factors on the Driver Risks. Figure 5 illustrates the influences of drunk driving, fatigue driving, distracted driving, traffic violations, improper emergency response, and inexperienced driving on the driver risks.

From Figure 5, it is clear that, regarding driver factors, the failure probabilities attributed to improper emergency response and traffic violations are comparatively higher at 0.694 and 0.660, respectively. In comparison, the failure probability caused by inexperienced driving is lower. The occurrence frequencies are the highest among causes of accidents for speeding, illegal lane changing, illegal overtaking, illegal parking, and illegal emergency lane parking, all of which can be classified as forms of a traffic violation. The probability of an accident is generally high in case of improper emergency response. The probability of accidents caused by inexperienced driving is rather low because of the strict management of drivers by road transportation companies.

4.2.3. Effects of Relevant Factors on the Vehicle Risks. The impacts of the braking system, steering system, light signal system, tires truck overload, and unbalanced load on the vehicle risks are illustrated in Figure 6.

As shown in Figure 6, the failure probabilities attributable to the braking system and tires are the highest among all vehicle factors—0.885 and 0.888, respectively. Braking system faults generally refer to braking failure or poor braking caused by improper operation, wear of parts, or inadequate vehicle clearance, while tire faults often refer to the severe wear or blowout of tires. The lowest failure probability is attributable to light signal system faults

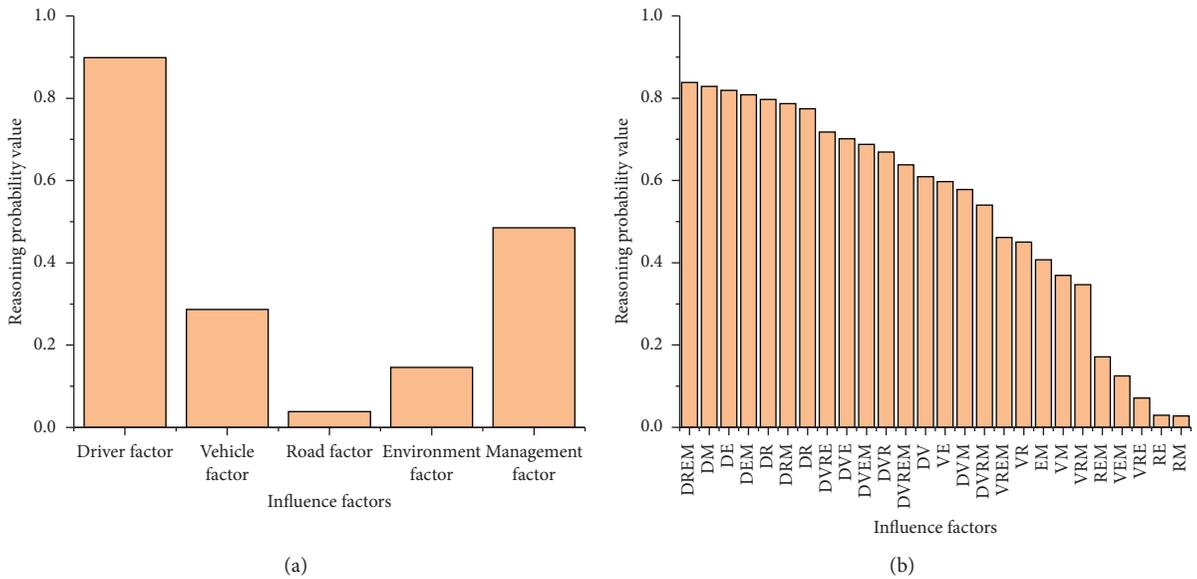


FIGURE 4: Effects of relevant factors on the occurrence of accidents.

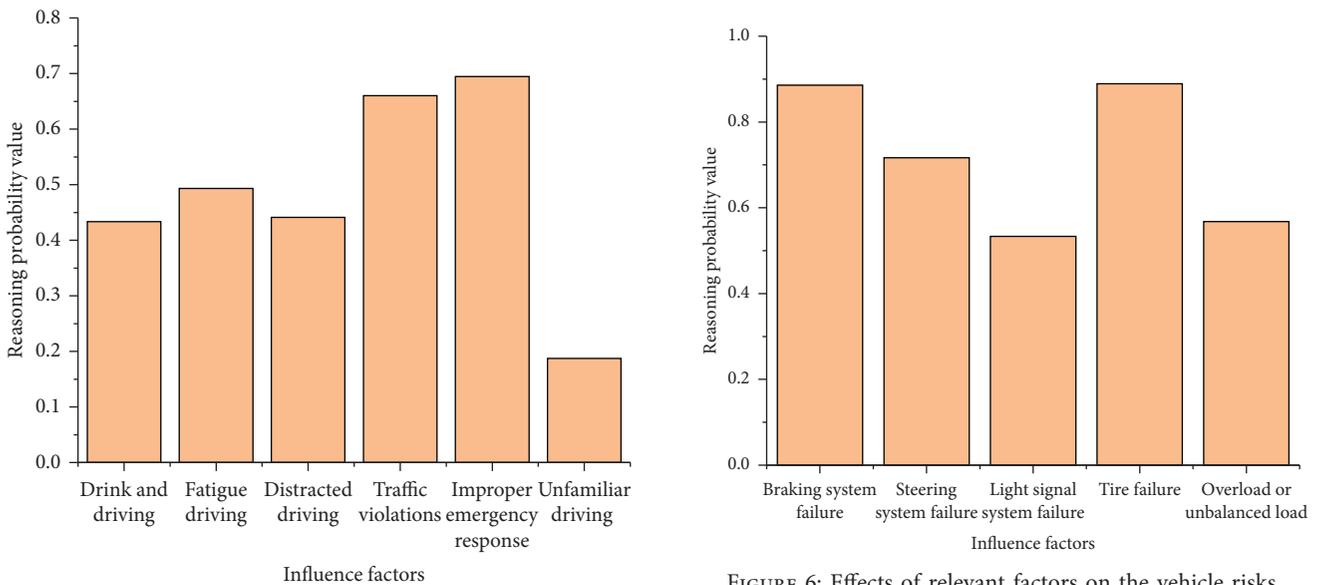


FIGURE 5: Effects of relevant driver factors.

FIGURE 6: Effects of relevant factors on the vehicle risks.

because accidents can often be avoided in this case as long as the drivers take appropriate emergency measures.

4.2.4. *Effects of Relevant Factors on the Road Risks.* In Figure 7, the influences of roadsides, bridges and tunnels, alignment, intersections, pavements, signs and markings, sight distance, and safety protection facilities on the road risks are detailed.

From Figure 7, it is clear that the failure probabilities attributable to pavement and alignment problems are the highest among all road factors—valued at 0.852 and 0.782, respectively. Pavement problems generally refer to slippery surfaces, potholes, and so forth, while alignment problems include sharp bends, steep slopes, and long descents. As these are rather common

causes of accidents, the relevant probabilities are also high. In contrast, sight distance is the least likely cause among all road factors because such problems generally involve roadside obstructions or intersections, whose probabilities are rather low according to accident statistics.

4.2.5. *Effects of Relevant Factors on Management Failures.* Figure 8 presents how dynamic monitoring, regulations, and private contracting affect the management risks.

According to Figure 8, the probability of management problems caused by improper dynamic monitoring is the highest at 0.699. At present, information about fatigued driving, distracted driving, speeding, vehicle trajectories, and so forth can be tracked and forewarned by the dynamic

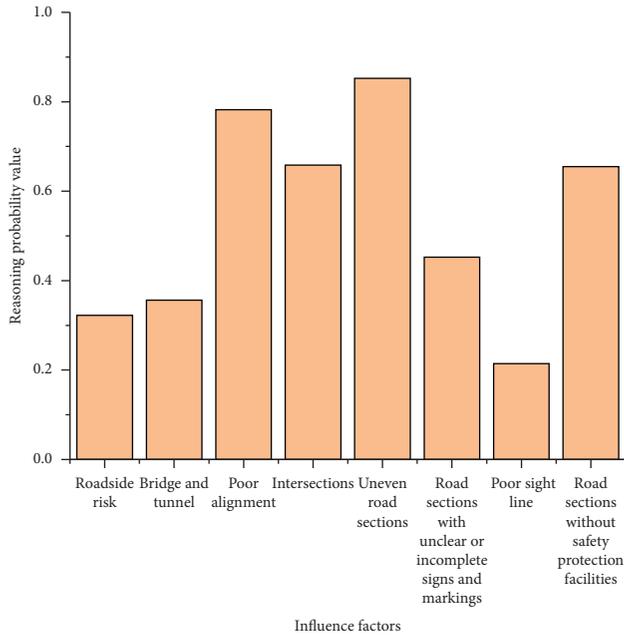


FIGURE 7: Effects of relevant factors on the road risks.

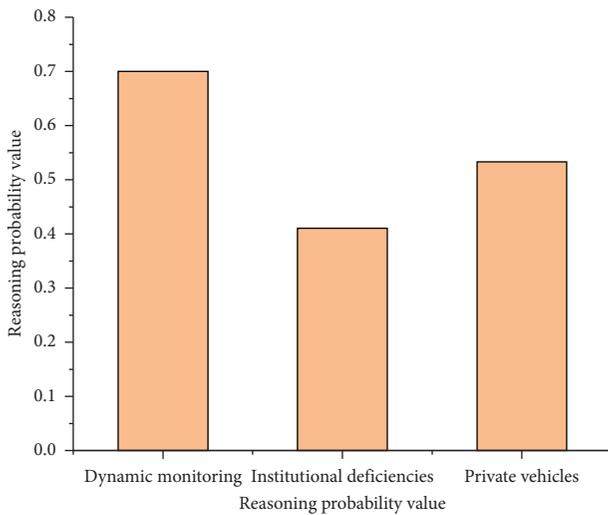


FIGURE 8: Effects of relevant factors on management failures.

monitoring system for road transport vehicles. A substantial number of accidents can be avoided by the proper functioning of the vehicle monitoring system. In contrast, the occurrence of an accident is generally attributed to the improper functioning of the dynamic monitoring system. Regulations are the least likely factors causing management failures. Overall, regulatory problems include driving management and inadequate vehicle management that is indirectly responsible for the causes of accidents in general.

4.2.6. *Effects of Relevant Factors on the Traffic Environment.* Figure 9 illustrates how traffic accidents, severe weather, natural disasters, and night lighting affect the traffic environment.

From the figure, it is clear that the failure probability attributable to severe weather is the highest among all environmental factors, which is at 0.666. Severe weather mainly refers to conditions such as heavy rain, fog, snow, or wind warnings, which are the most common causes of traffic environment failures. Meanwhile, given the relatively low incidence of natural disasters, the probability of environmental failures resulting from them is the lowest.

All the above analysis results accord with expert experience and knowledge, suggesting the effectiveness of the created model.

4.3. *Case-Based Analysis.* Introducing “a certain major road traffic accident” helps in carrying out an empirical analysis of knowledge inference. On February 20, 2019, at 19:07 h, an ordinary passenger bus collided with the sidewall of a tunnel on the 2441 km + 100 m section of the G65 Bao-tou–Maoming Expressway (within a tunnel in Hezuo Village, Lingui District, Guilin), which caused 4 deaths, 3 serious injuries, 16 minor injuries, and severe vehicle damage. The causes of this major road traffic accident were analyzed as follows:

- (1) Driver factors: after the bus entered the tunnel, the driver oversped, ignoring the signs and markings. In response to the deceleration of the preceding vehicle, the driver applied emergency brakes, which was an improper emergency response, thereby losing control of his vehicle.
- (2) Vehicle factor: during steering, the bus suddenly went out of control, so the possibility of vehicle failure cannot be ruled out.
- (3) Environment factor: there was a heavy shower at the time of the accident that made the road surface slippery.
- (4) Management factors: due to the long-term, nonlocal, out-of-scope operation of the tourist bus, its dynamic vehicle monitoring terminal had been offline.

Based on the above description, the node elements for this accident are detailed in Table 3.

Based on the Bayesian network built in Figure 3, the nodes relating to accident occurrence are mapped into a Bayesian network, followed by knowledge inference in GeNIe 2.0. As shown in Figure 10, the probabilities are updated for each node.

Figure 10 clearly shows that the foremost cause of this accident was the human failure attributed to improper emergency response and traffic violations, whose probability was 82%; the second important cause was the vehicle failure

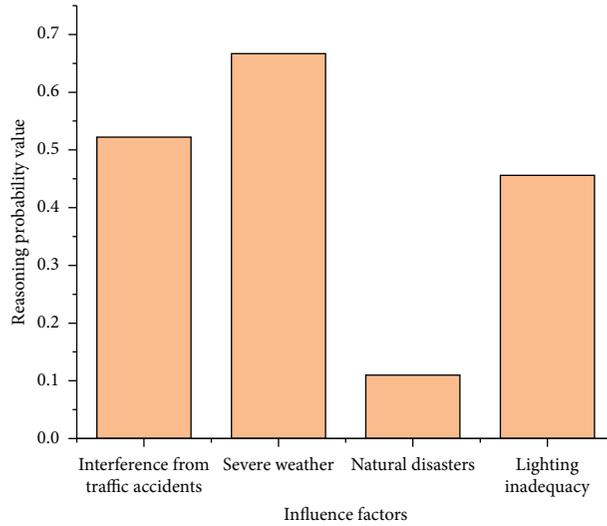


FIGURE 9: Effects of relevant factors on the environmental risks.

TABLE 3: Bayesian network nodes for the accident.

Risk factor	Network node
Driver	Improper emergency response
Vehicle	Traffic violation
	Vehicle fault
Road	Pavement
Environment	Bridge and tunnel
	Severe weather
Management	Dynamic monitoring
	Private contracting

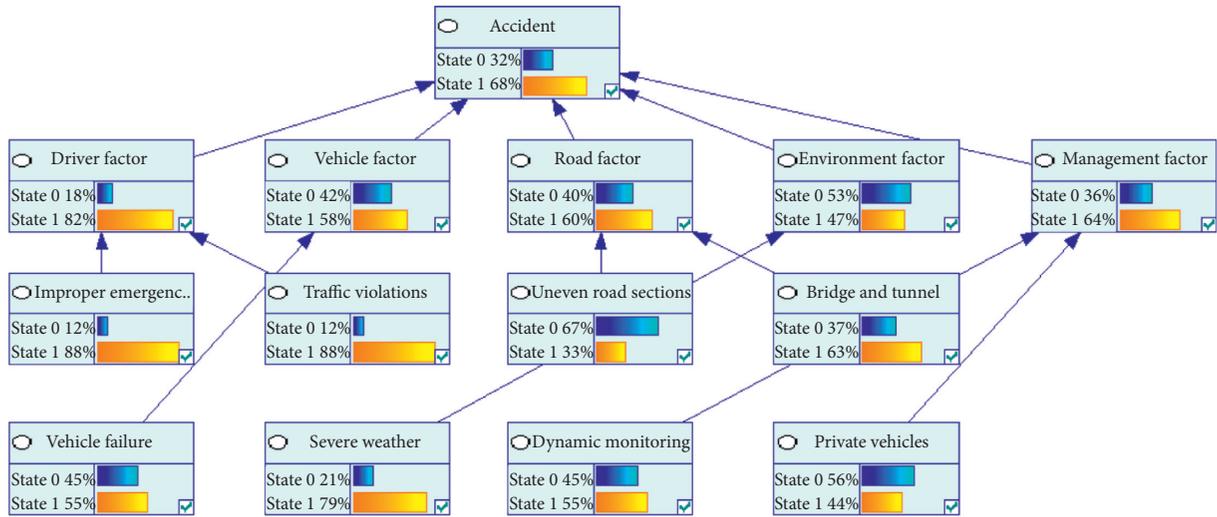


FIGURE 10: Probability updates and inference diagram of the accident.

attributable to vehicle fault, with a probability of 58%; besides, the probability of traffic environment problems resulting from severe weather that was 47%; and the probability of management failure resulting from inadequate dynamic monitoring and

private contracting was 68%. Ultimately, the probability of occurrence of the accident was 0.68, which is rather high. These descriptions conform to the actual accident inference process that proves the feasibility of the proposed method.

5. Conclusions

Highway transportation safety risks involve multiple factors, such as driver, vehicle, road, environment, and management. According to the results of the Bayesian network inference, the primary cause of accidents is driver-related. Meanwhile, management factors are nonnegligible as well. The foremost human factors are improper emergency responses and traffic violations, of which improper emergency responses are often ignored by drivers or managers. Additionally, the foremost vehicle factor is faulty vehicles, which is attributed to the failure of the braking system or tires. Severe weather serves as the foremost factor for traffic environment. Interference from existing accidents is another factor that is often ignored by drivers or managers.

By using the Bayesian network for inference of knowledge graphs oriented to highway transportation safety risks, the rule acquisition problem with rule-based reasoning, as well as the limitations in data size and quality can be avoided. With powerful reasoning, the above method can preferably be integrated with the transportation safety risk knowledge graphs to achieve flexible inference of accident probability in the dynamic risk coupling scenarios. In the next phase, in view of the dynamic nature of such knowledge graphs, we will probe deeper into the collaborative fusion between the Bayesian network-based knowledge inference and the dynamic knowledge graphs.

Data Availability

The data are available through contacting the author via e-mail: 815345591@163.com; hyperlink (<https://pan.baidu.com/s/1Y9d3zGtXtowCmWKHrQ678g>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Special Fund 2019-C507 of Science and Technology Innovation of Research Institute of Highway Ministry of Transport, and the Special Fund Project of Basic Research Expenses of Central Public Welfare Research Institutes 2020-9066.

References

- [1] T. Yu, J. Li, Q. Yu et al., "Knowledge graph for TCM health preservation: design, construction, and applications," *Artificial Intelligence in Medicine*, vol. 77, pp. 48–52, 2017.
- [2] B. Ribeiro, N. Chen, and A. Kovacec, "Shaping graph pattern mining for financial risk," *Neurocomputing*, vol. 326–327, pp. 123–131, 2019.
- [3] Y. Chen, Y. D. Wang, J. D. Wang et al., "Method for information system risk management based on Markov logic networks," *Computer Engineering and Applications*, vol. 52, no. 18, pp. 104–110, 2016.
- [4] L. G. Yang, S. P. Tang, and J. Zhu, "A Markov logic network based sentence sentimental analysis method," *Transactions of Beijing Institute of Technology*, vol. 33, no. 6, pp. 600–604, 2013.
- [5] Y. Liu, C. Ouyang, and J. Li, "Ensemble method to joint inference for knowledge extraction," *Expert Systems with Applications*, vol. 83, pp. 114–121, 2017.
- [6] Q. Lu, P.-a. Zhong, B. Xu et al., "Risk analysis for reservoir flood control operation considering two-dimensional uncertainties based on Bayesian network," *Journal of Hydrology*, vol. 589, Article ID 125353, 2020.
- [7] D. L. Yin, Z. B. Chen, J. C. Zeng et al., "Human factors inference of safety accidents in coal mine based on qualitative Bayesian network with weight," *Journal of WUT (Information & Management Engineering)*, vol. 39, no. 1, pp. 514–518, 2017.
- [8] D. Y. Xia, X. M. Qian, and Z. P. Duan, "Scenario deduction model of unconventional emergency based on dynamic Bayesian network," *Journal of Northeastern University (Natural Science)*, vol. 36, no. 6, pp. 897–902, 2015.
- [9] A. Seyed Hassani, M. S. Haghghi, and A. Khonsari, "Bayesian inference of private social network links using prior information and propagated data," *Journal of Parallel and Distributed Computing*, vol. 125, pp. 72–80, 2019.
- [10] M. M. Rajabi and B. Ataie-Ashtiani, "Efficient fuzzy Bayesian inference algorithms for incorporating expert knowledge in parameter estimation," *Journal of Hydrology*, vol. 536, pp. 255–272, 2016.
- [11] R. Xie, Z. Liu, J. Jia et al., "Representation learning of knowledge graphs with entity descriptions," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, Phoenix, AZ, USA, 2016.
- [12] K. M. Annervaz, S. B. R. Chowdhury, and A. Dukkupati, "Learning beyond datasets: knowledge graph augmented neural networks for natural language processing," 2018, <https://arxiv.org/abs/1802.05930>.
- [13] F. Godin, A. Kumar, and A. Mittal, "Using ternary rewards to reason over knowledge graphs with deep reinforcement learning," 2019, <https://arxiv.org/abs/1902.10236>.
- [14] J. Jiang, X. Li, C. Zhao, Y. Guan, and Q. Yu, "Learning and inference in knowledge-based probabilistic model for medical diagnosis," *Knowledge-Based Systems*, vol. 138, pp. 58–68, 2017.
- [15] Q. Liu, M. H. Han, L. Y. Jiang et al., "Two-tier random walk based relational inference algorithm," *Chinese Journal of Computers*, vol. 40, no. 6, pp. 1275–1290, 2017.
- [16] 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.
- [17] S. L. Lauritzen, *The EM Algorithm for Graphical Association Models with Missing Data*, Elsevier Science Publishers, Amsterdam, Netherlands, 1995.
- [18] C. Liu, S. Lv, X. Peng et al., "Analysis and comparison of different impacts of aging and loading frequency on fatigue characterization of asphalt concrete," *Journal of Materials in Civil Engineering*, vol. 32, no. 9, Article ID 04020240, 2020.
- [19] C. Liu, S. Lv, D. Jin et al., "Laboratory investigation for the road performance of asphalt mixtures modified by rock asphalt-styrene butadiene rubber," *Journal of Materials in Civil Engineering*, vol. 33, no. 3, Article ID 04020504, 2021.