Article

Third-Party Damage Model of a Natural Gas Pipeline Based on a Bayesian Network

Baikang Zhu [1], Xu Yang [2], Jun Wang [2], Chuanhui Shao [3], Fei Li [4], Bingyuan Hong [1,5,*], Debin Song [1,*] and Jian Guo [1,*]

[1] National & Local Joint Engineering Research Center of Harbor Oil & Gas Storage and Transportation Technology, Zhejiang Key Laboratory of Petrochemical Environmental Pollution Control, School of Petrochemical Engineering & Environment, Zhejiang Ocean University, Zhoushan 316022, China

[2] School of Shipping and Maritime, Zhejiang Ocean University, Zhoushan 316022, China

[3] Zhejiang Zheneng Natural Gas Operation Co., Ltd., Hangzhou 310052, China

[4] Sinopec Sales Co., Ltd. Zhejiang Quzhou Petroleum Branch, Quzhou 324000, China

[5] Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Fuxue Road No. 18, Changping District, Beijing 102249, China

* Correspondence: hongby@zjou.edu.cn (B.H.); sdb18857057903@163.com (D.S.); 025125@zjou.edu.cn (J.G.)

Abstract: Natural gas plays an important role in the transition from fossil fuels to new energy sources. With the expansion of pipeline networks, there are also problems with the safety of pipeline network operations in the process of transportation. Among them, third-party damage is a key factor affecting the safety of pipelines. In this paper, the risk factors of third-party damage are analyzed, and an evaluation model of natural gas pipeline damage is established using the GeNIe Modeler. Through Bayesian network reverse reasoning and a maximum cause chain analysis from the four aspects of personnel, environment, management, and equipment, it was found that the top five factors that have significant influence on third-party damage, are safety investment, the completeness of equipment, safety inspection frequency, the management of residents along the pipeline, and safety performance, with the posteriori probability in the model of 97.3%, 95.4%, 95.2%, 95.1%, 95.1%, respectively. Consequently, it is necessary for pipeline operation companies to secure investment on safety, to make sure that the safety equipment (system) works and is in a good condition, to maintain the safety inspection frequency in an organization, to build a management system for residents along the pipeline, and to conduct routine safety performance assessments accordingly.

Keywords: natural gas pipeline; Bayesian network; third-party damage; evaluation model

1. Introduction

In September 2020, China proposed a "dual carbon strategy", proclaiming it will strive to achieve carbon emissions peaking by 2030 and carbon neutrality by 2060. In this context, with the benefits of lower carbon emissions compared with coal and oil, natural gas will become an important intermediate energy source in the next decade and will play an important role in China achieving its dual-carbon strategic target. As one of the models of transportation of natural gas [1], third-party damage to pipelines caused by factors such as engineering construction and man-made damage has become an important risk factor affecting the operational safety of natural gas pipelines. To prevent the third-party damage accidents of pipelines, it is necessary to construct an evaluation model, in which the systematic analysis of the occurrence mechanism, the key influencing factors, and the indicators of third-party damage to natural gas pipelines are studied.

Domestic and abroad scholars have carried out a series of researches on the safety evaluation of natural gas pipelines. Safety evaluation methods such as the bow-tie model, analytic hierarchy process (AHP), fault tree analysis (FTA), and grey comprehensive evaluation (GCE) have been widely used in oil and gas pipeline risk evaluation. Chen et al.

combined the bow-tie model and AHP to obtain the ranking of various risk factors [2], conducted a risk assessment of pipelines according to their ranking results, and concluded that third-party damage is the main reason for pipeline failure in China. Zhang et al. [3] used the principal component-clustering analysis method to divide the pipeline risk index into three principal component factors, which reduced the number of risk factors and reduced the error caused by the uncertainty of evaluation effectively. Guo et al. [4] proposed a comprehensive evaluation method combining FTA, AHP, and grey theory for urban gas pipelines. FTA was used to determine the risk factors of urban gas pipelines, then a risk evaluation index system for urban gas pipelines was established through the analytic hierarchy process, then, the risk evaluation indicators were divided into two levels, and finally, the pipeline risk value was obtained through the grey comprehensive evaluation method. Yavorskyi et al. [5] built a structure of the risk management system for safe pipeline maintenance under geodynamic influences, and developed a complex application of methods for inspection of a pipeline and the adjacent rock mass to obtain the level of geodynamic hazards on a pipeline. Yatsyshyn et al. [6] launched a comprehensive study of both the internal technological processes, the state of equipment, and external influences, where the factors affecting oil and gas industry emergencies were studied and the main types of environmental hazards were identified.

The traditional risk analysis methods, such as fault tree, event tree, etc., can be used to analyze the causes and failure consequences of gas accidents qualitatively, but they cannot be adapted to the dynamic evolution process of accidents well, due to their inherent static structure [7–9]. For example, when a natural gas pipeline leakage happens, the scope of the leakage expands over time, the emission gas keeps accumulating in the underground space, and the probability of an explosion varies accordingly; thus, the fault tree and event tree analysis methods are not able to reflect the dynamic evolution process of gas leakage accurately.

Compared with traditional risk analysis methods, a Bayesian network has the advantage of analyzing the uncertain factors existing in a system better, to express the causal relationship between event nodes, and to analyze the cause and effect of the accident quantitatively. Moreover, newly obtained information could also be processed through inference to update the probability [10]. Mamdikar et al. [11] proposed a system reliability research method for a nuclear power plant, based on the fault tree and dynamic Bayesian network. In the framework created, the information on the component's failure probability could be updated based on observed data. Wang et al. [12] proposed a dynamic risk analysis method for gas pipeline networks based on a Bayesian network and verified the characteristics of gas pipeline network failures and accident consequences that change with time. Li et al. [13] employed the Markov chain Monte Carlo algorithm for hazard liquid pipeline leakage position and coefficient sampling, which can quickly assess the impact on the surrounding environment. Considering that pipelines are affected by comprehensive factors such as personnel, environment, management, and equipment, the disaster mode of pipelines is a dynamic evolution process. Moreover, studies have shown that the combination of the pipeline failure mode, gas leakage, gas cloud ignition time interval, and the degree of confinement of the surrounding space will determine the final disaster mode [14], but due to the uncertainty of these factors, it is difficult to obtain reliable results by using historical statistical data.

In this paper, a third-party damage evaluation model for natural gas pipelines is established based on Bayesian networks. The evolution process of natural gas pipeline accidents from the failure cause to the consequence is analyzed thoroughly and the correlation between the factors is established by a correlation analysis of the influencing factors. The structure of the Bayesian network model is further optimized, and a third-party damage evaluation model of the pipeline based on the Bayesian network is obtained. Then, the parameters and a Bayesian network reverse inference are conducted with the method of a maximum cause analysis. With the rapid diagnosis of the accident cause, the prediction of a future risk, and the evaluation of an emergency response to a gas accident, a comprehensive assessment of a pipeline's third-party damage accident was obtained.

This paper is structured as follows. Section 2 presents the methodology of this paper, and introduces the framework of third-party damage risk assessment procedures. In Section 3, variables were set for a risk identification and a correlation analysis with different factors was conducted. Section 4 contains the construction of the Bayesian network for the third-party damage model, the network structure learning, network optimization, parameter learning, reverse reasoning, and a maximum causal chain analysis.

2. Methodology

This paper presents a comprehensive application of the Bayesian network method on the third-party damage of a natural gas pipeline combined with an understanding of the pipeline's operation. Firstly, relevant data on the third-party damage to the pipeline was collected and analyzed, and the third-party damage factors of the pipeline were classified. Before constructing the Bayesian network model, the correlation analysis was carried out on the influencing factors, and the relationship between the factors was clarified through the Spearman rank correlation coefficient test method which is widely used. As an artificial intelligence modeling and machine learning software based on Bayesian networks, the GeNIe Modeler provides algorithms such as Naive Bayes, Greedy Search, Bayesian Search, etc. The software was applied to learn the Bayesian network structure in this paper. Additionally, the method of the node sorting and local scoring-search algorithm was used to optimize and adjust the Bayesian network, and to calculate the prior probability when damage does not occur, and the posterior probability when damage occurs, respectively. Furthermore, since the maximum likelihood estimation method does not need to consider prior knowledge during analysis, it is suitable for datasets with full samples; therefore, it was used to analyze the main causes of the third-party damage to the pipeline. Lastly, in order to find the most important factors affecting the target variable by comparing the posterior probability of each node, a Bayesian network reverse inference was used to calculate the posterior probability of other node variables in this paper. The framework of the methodology is shown in Figure 1.
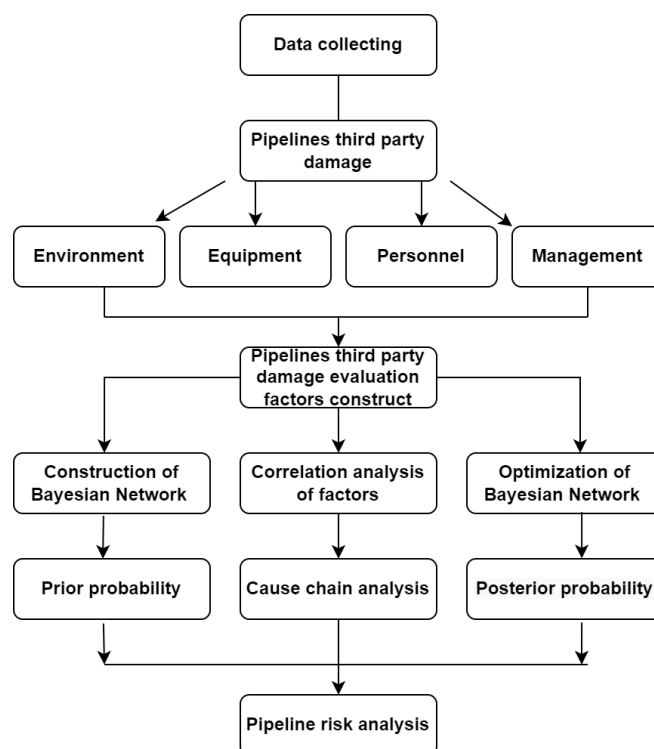

Figure 1. Pipeline third-party damage risk assessment procedure.

3. Pipeline Third-Party Damage Risk Identification and Correlation Analysis

3.1. Risk Variables Setting

The mechanism of pipeline damage and degradation may differ from each other in different climate zones. In this paper, the third-party damage accidents database from the Zhejiang Zheneng Natural Gas Operation Co., Ltd. (Hangzhou, China) was used in this study, as most of the cases occurred in the Zhejiang province of China, which has a subtropical monsoon climate. The risk assessment indicators were obtained according to expert opinions, with a combination of quantitative and qualitative analysis, and external and internal analysis. By analyzing the causes of third-party damage accidents of natural gas pipelines, it was divided into four aspects, namely, environment, personnel, management, and equipment, with a total number of 28 risk factor evaluation indicators.

Considering the characteristics of Bayesian network modeling, discrete variables are usually required to use the Bayesian network model, to ensure the accuracy [15]; therefore, before constructing the Bayesian network model, continuous variable discretization processing is needed. There are two types of variable states in this paper, which are represented by 0 and 1, respectively, (1 means the indicator event did happen, and 0 means it did not happen). The variable settings are shown in Table 1.

Table 1. Risk factor property settings.

Factor Name	Code	Value	
Monitoring equipment completeness	A01	1	0
Pipeline safety protection	A02	1	0
Equipment completeness	A03	1	0
Equipment replacement	A04	1	0
Equipment failure rate	A05	1	0
Equipment maintenance	A06	1	0
Geological structure	B07	1	0
Operating temperature	B08	1	0
Precipitations	B09	1	0
Management system completeness	C10	1	0
Safety inspection frequency	C11	1	0
Safety investment	C12	1	0
Government regulation	C13	1	0
Completeness of emergency plan	C14	1	0
Timeliness of emergency rescue	C15	1	0
Emergency supplies	C16	1	0
Rescuer allocation	C17	1	0
Management of residents along the pipeline	C18	1	0
The average age of employees	D19	1	0
The average level of education of employees	D20	1	0
Average years of employee training	D21	1	0
Job match	D22	1	0
Safety performance	D23	1	0
Employee attendance	D24	1	0
The physical examination pass rate	D25	1	0
Employee mental health	D26	1	0
Employee violations	D27	1	0
Status of residents along the pipeline	D28	1	0

A, B, C and D represent the four aspects of risk factors which are personnel, equipment, environment and management, respectively.

3.2. Correlation Analysis of Risk Factors

Before constructing the Bayesian network model, the correlation analysis was carried out on the influencing factors of the natural gas pipelines' third-party damage, and the relationship between the factors was clarified. The commonly used methods for variable correlation analysis include the Spearman rank correlation coefficient test and the Pearson correlation coefficient test [16,17]. Since the risk variables constructed in this paper are

ordinal, we used the Spearman rank correlation coefficient test method to analyze the correlation of the variables. The formula is as follows:

$$r = 1 - \frac{6 \sum_{i=1}^{n} D_i^2}{n(n^2 - 1)} \quad (-1 \leq r \leq 1) \tag{1}$$

$$D_i^2 = \sum_{i=1}^{n} (Ui - Vi)(Ui - Vi) \tag{2}$$

U and *V* are the factors of the correlation analysis, and *n* is the sample size. *r* is the correlation coefficient and the larger value stands for a stronger correlation between variables. $r > 0$, means the two variables are positively correlated; $r < 0$, means the two variables are negatively correlated; $r = 1$, means the two variables are completely positively correlated; $r = -1$, means the two variables are completely negatively correlated; $r = 0$, means the two variables are not correlated. Since the third-party damage to natural gas pipelines mainly involves three aspects: personnel, management, and equipment, we carried out a correlation analysis mainly for these aspects accordingly.

Twenty-two industry practitioners and experts were invited to score for the 28 indicators through a questionnaire survey to evaluate the influence of each indicator on the pipelines' third-party damage. Then, the Spearman's rank correlation coefficient method was applied to calculate the difference between the U rank and the V rank for each pair of data (the process of Formula (2)), and finally, the strength of the rank correlation between the variables was obtained (the process of Formula (1)), as shown below in Tables 2–4.

Table 2. Spearman correlation analysis between management and equipment.

	A01	A02	A03	A04	A05	A06
C10	0.291	0.139	0.046	0.313	0.139	0.265
C11	0.602 **	0.662 **	0.584 **	0.624 **	0.662 **	0.683 **
C12	0.374	0.373	0.360	0.511 *	0.373	0.422
C13	0.533 *	0.393	0.411	0.465 *	0.377	0.646 **
C14	0.298	0.325	0.433 *	0.662 **	0.419	0.421
C15	0.539 **	0.638 **	0.339	0.563 **	0.556 **	0.628 **
C16	0.559 **	0.568 **	0.288	0.361	0.485 *	0.631 **
C17	0.318	0.286	0.430 *	0.555 **	0.375	0.360

* $p < 0.05$, indicating a significant correlation at the 0.05 level (two-sided); ** $p < 0.01$, indicating a significant correlation at the 0.01 level (two-sided).

Table 3. Spearman correlation analysis between personnel and equipment.

	A01	A02	A03	A04	A05	A06
D19	0.519 *	0.359	0.450 *	0.599 **	0.247	0.434 *
D20	0.656 **	0.415	0.330	0.458 *	0.233	0.450*
D21	0.328	0.441 *	0.251	0.469 *	0.536 *	0.393
D22	0.461 *	0.302	0.101	0.288	0.250	0.321
D23	0.509 *	0.372	0.251	0.545 **	0.392	0.390
D24	0.384	0.125	−0.097	0.092	0.276	0.084
D25	0.372	−0.055	0.101	0.229	−0.172	−0.011
D26	0.526 *	0.116	0.126	0.319	0.177	0.175
D27	0.325	0.078	−0.027	−0.030	0.025	0.182
D28	0.391	0.203	0.156	0.422	0.311	0.253

* $p < 0.05$, indicating a significant correlation at the 0.05 level (two-sided); ** $p < 0.01$, indicating a significant correlation at the 0.01 level (two-sided).

Table 4. Spearman correlation analysis between personnel and management.

	C10	C11	C12	C14	C15	C16	C17
D19	0.462 *	0.260	0.512 *	0.706 **	0.544 **	0.523 *	0.603 **
D20	0.338	0.297	0.275	0.507 *	0.426 *	0.525 *	0.473 *
D21	0.346	0.504 *	0.463 *	0.600 **	0.621 **	0.404	0.553 **
D22	0.493 *	0.318	0.449 *	0.366	0.641 **	0.431 *	0.230
D23	0.584 **	0.634 **	0.664 **	0.621 **	0.691 **	0.480 *	0.493 *
D24	0.411	0.464 *	0.319	0.288	0.348	0.386	0.336
D25	0.323	0.144	0.111	0.299	0.169	0.121	0.433 *
D26	0.424 *	0.187	0.151	0.437 *	0.278	0.299	0.440 *
D27	0.366	0.096	0.195	0.135	0.417	0.337	−0.056
D28	0.523 *	0.227	0.327	0.551 **	0.417	0.358	0.382

* $p < 0.05$, indicating a significant correlation at the 0.05 level (two-sided); ** $p < 0.01$, indicating a significant correlation at the 0.01 level (two-sided).

3.2.1. Correlation Analysis of Management and Equipment

The Spearman correlation analysis between the management and equipment is shown in Table 2. Once the values of the Spearman coefficients were obtained, a statistical test was conducted to generate the p value, with the p value to be compared with α (which is usually set at 0.05 or 0.01). When the p value was $< \alpha$, the values were regarded as statistically significant. There was a correlation between the equipment completeness and the frequency of safety inspections, government regulation, timeliness of emergency rescue, and the emergency supplies. The correlation coefficient values were 0.602, 0.533, 0.539 and 0.559, respectively, which were all greater than 0, indicating there was a positive correlation between the above factors. Additionally, there was a correlation between the equipment completeness and the frequency of safety inspections, the completeness of emergency plans, and the rescuer allocation. The correlation coefficient values were 0.584, 0.433 and 0.430, respectively, which were all greater than 0, indicating that there was a positive correlation between the factors.

3.2.2. Correlation Analysis of Personnel and Equipment

The Spearman correlation analysis between the personnel and equipment is shown in Table 3. There was a correlation between the completeness of monitoring equipment and the average age of the staff, the average level of education of the employees, job matching, safety performance, and employee mental health. The correlation coefficient values were 0.519, 0.656, 0.461, 0.509 and 0.526, respectively, all greater than 0, indicating that there was a significant correlation between the factors mentioned above.

3.2.3. Correlation Analysis of Personnel and Management

The Spearman correlation analysis between the personnel and management is shown in Table 4. There was a significant correlation between the completeness of the management system and the average age of employees, job matching, safety performance, employee mental health, and the status of residents along the pipeline. The correlation coefficient values were 0.462, 0.493, 0.584, 0.424 and 0.523, respectively, all greater than 0, which indicates that there was a positive correlation between the factors mentioned above.

4. Construction of a Bayesian Network for Third-Party Damage of a Natural Gas Pipeline

Based on the network nodes of the risk variable system established above, a Bayesian network model for a third-party damage analysis of natural gas pipelines was constructed.

4.1. Bayesian Network Structure Learning

To quantitatively analyze the third-party damage factors of natural gas pipelines and to determine the main causes of third-party damage, it is necessary to learn the structure of the Bayesian network. In this paper, we used the GeNIe Modeler to learn the Bayesian

network structure, which provides algorithms including Naive Bayes, Greedy Search, Bayesian Search, etc. The core idea of the greedy search is to select the local optimal solution each time. Combined with the data characteristics and algorithm characteristics collected and processed in this paper, we used the greedy search to learn the structure of the Bayesian network [18]. After importing the processed data table into the GeNIe Modeler, the results are shown below in Figure 2.
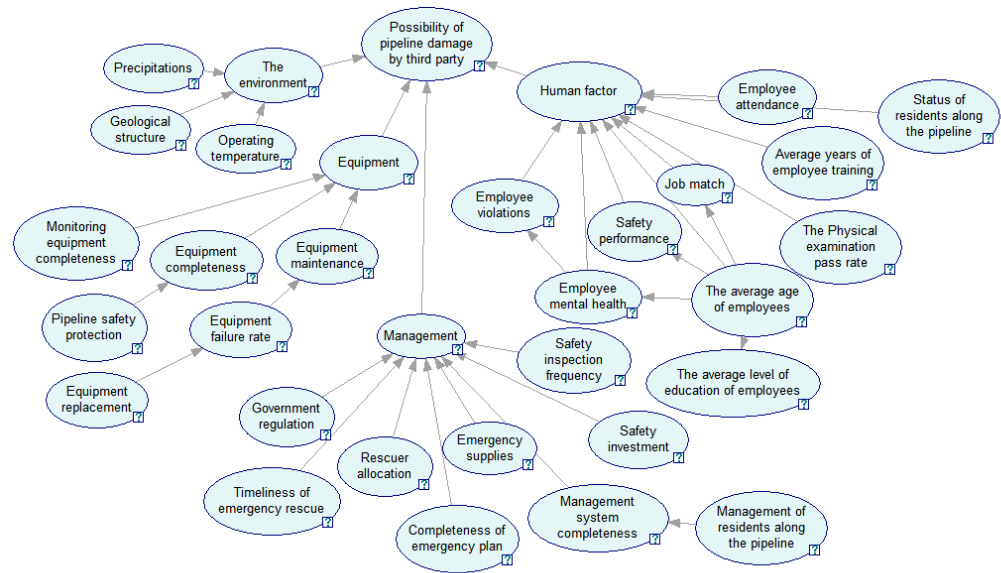

Figure 2. Initial Bayesian network diagram.

Once the Bayesian network of a pipeline's third-party damage is created, the causal relationships between the equipment, management, and personnel variables also require experts in relevant fields to optimize the Bayesian network based on certain objective facts combined with their experience.

4.2. Optimization of Bayesian Network Structure

In this section, the method of the node sorting and the local scoring-search algorithm was used to optimize and adjust the Bayesian network [19]. T represents the third-party damage event of natural gas pipelines. According to the setting of the risk factors in Table 1, monitoring equipment completeness, pipeline safety protection, equipment completeness, equipment replacement, equipment failure rate, and equipment maintenance are set as $X_1^t, X_2^t, X_3^t \cdots X_6^t$. The geological structure, operating temperature, and precipitations are set as X_7^t, X_8^t, X_9^t. The management system completeness, safety inspection frequency, safety investment, government regulation, the completeness of emergency plans, timeliness of emergency rescue, emergency supplies, rescuer allocation, and the management of residents along the pipeline are set as $X_{10}^t, X_{11}^t, X_{12}^t \cdots X_{18}^t$. The average age of employees, the average level of education of the employees, the average years of employee training, job matching degree, safety performance, employee attendance, physical examination pass rate, employee mental health, employee violations, and the status of the residents along the pipeline are set as $X_{19}^t, X_{20}^t, X_{21}^t \cdots X_{28}^t$.

Using a local scoring method: let $T_i = \left(X_i^1, X_i^2, X_i^3, \cdots X_i^s \right)$, for any $X_i^k, X_i^j \in T_i$, when $P\left(X_i^k \to X_i^j \middle| H_i, D \right) > P\left(X_i^j \to X_i^k \middle| H_i, D \right)$ (where H_i is the hidden variable obtained from the Bayesian network learning and D is the sample set), the directed edge $X_i^k \to X_i^j$ is established, otherwise, the directed edge is $X_i^j \to X_i^k$, if the node X_{ai} satisfies:

$$MDL(P(X_{ai}|X_{ai}^l)|D) = \min\{MDL(P(X_{ai}|X_{ak}))|D\}X_{ak} \in (X_{a1}X_{a2} \cdots X_{ai-1}) < MDL(X_{ai}|D), \qquad (3)$$

Then, X_{ai}^l can be used as a parent node of X_{ai} [20,21], through continuous scoring by experts, and the optimized Bayesian network model of the third-party damage to natural gas pipelines is finally obtained, as shown in Figure 3.
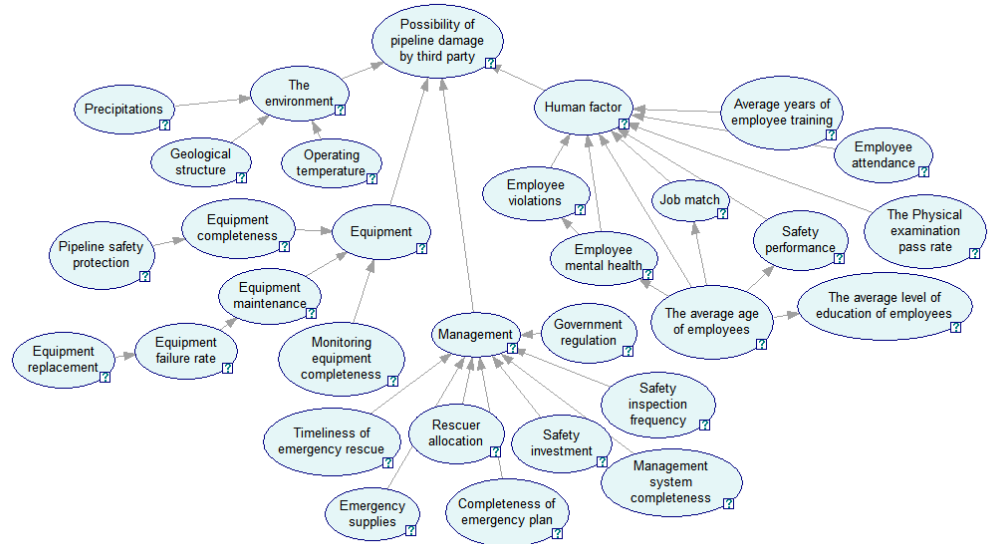

Figure 3. Optimized Bayesian network diagram.

4.3. Bayesian Network Parameter Learning

Through the optimized Bayesian network, the main causative factors in the network including the personnel, environment, management, and equipment are learned, and the values of the prior probability and posterior probability of each influencing factor are determined. Since the maximum likelihood estimation method does not need to consider prior knowledge during the analysis, it is suitable for datasets with full samples; therefore, we used the maximum likelihood estimation method for the parameter learning. Figure 4 shows the Bayesian network parameter learning, where YES represents the probability of a node event occurring, and NO represents the probability that a node event does not occur.
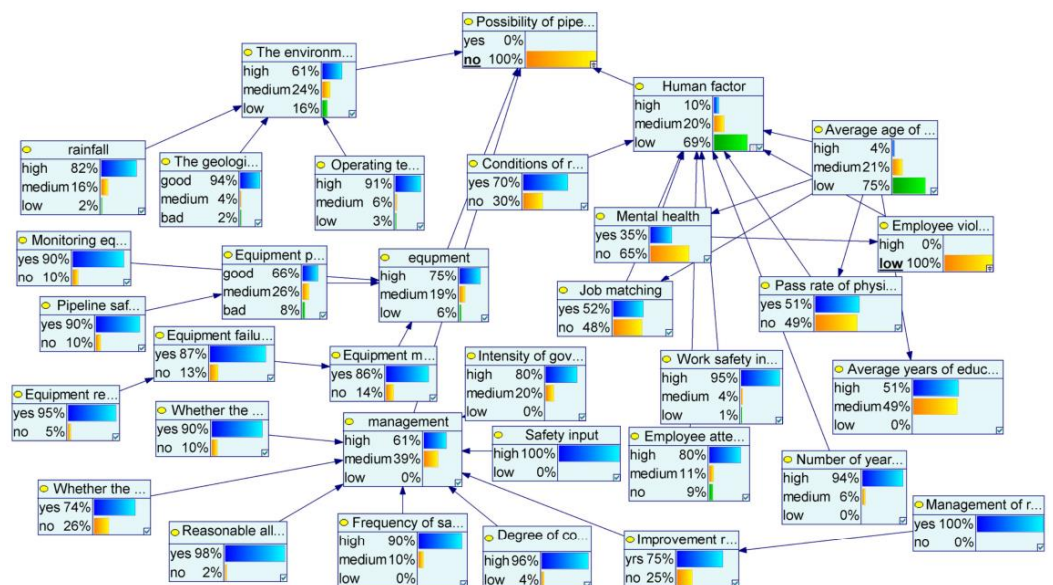

Figure 4. Bayesian network parameter learning.

The following main causative factors were identified and calculated through the risk variable correlation analysis in Section 3.2. Among them, the equipment maintenance

rate in the management factor and the employee attendance, physical examination pass rate, employee mental health, employee discipline violation rate, and the condition of residents along the pipeline in the personnel factor were relatively weakly correlated with the pipelines' third-party damage; thus, these six factors were excluded from the test probability. The prior probability was obtained from the databases and investigation reports, as shown in Table 5.

Table 5. The prior probability of pipeline third-party damage relevant nodes.

Item	Node Name	Prior Probability	Item	Node Name	Prior Probability
1	Monitoring equipment completeness	0.095	12	Government regulation	0.019
2	Pipeline safety protection	0.096	13	Completeness of emergency plan	0.078
3	Equipment completeness	0.046	14	Timeliness of emergency rescue	0.076
4	Equipment replacement	0.021	15	Emergency supplies	0.054
5	Equipment failure rate	0.012	16	Rescuer allocation	0.063
6	Geological structure	0.057	17	Management of residents along the pipeline	0.049
7	Operating temperature	0.017	18	The average age of employees	0.012
8	Precipitations	0.075	19	The average level of education of employees	0.023
9	Management system completeness	0.016	20	Average years of employee training	0.035
10	Safety inspection frequency	0.048	21	Job match	0.015
11	Safety investment	0.027	22	Safety performance	0.046

Once the weights of the above node factors were determined, they were substituted into the model. The state probability of each node was less than 0.1 after calculation, indicating that the state of each node was stable and there was no abnormal situation.

4.4. Reasoning with the Bayesian Network

Bayesian network reverse inference is the process to calculate the posterior probability of other node variables when the state of the model node variables is known, to find the most important factor affecting the target variable by comparing the posterior probability of each node [22,23].

With the combination of the characteristics of third-party damage accidents of natural gas pipelines and the Bayesian network model constructed in this paper, the parent node "natural gas pipelines third-party damage", and the child nodes "personnel factors", "environmental factors", and "management factors" were selected. The posterior probability of other related target nodes was calculated through the Bayesian network reverse reasoning, and the main factors affecting the target node variables were found.

Below is the reverse reasoning process for third-party damage of natural gas pipelines in the GeNIe Modeler. First, the parent node "natural gas pipelines third-party damage "is set as the evidence node, then, by clicking the option box, the "Set Evidence" function is selected, and the state probability of the parent node is set to 100%. Then, by clicking update to calculate the posterior probability value of each sub-node, the main influencing factors that lead to the damage of a natural gas pipeline are inferred, through analyzing the posterior probability value of each sub-node, as shown in Figure 5.

The posterior probability of each related accident node was obtained by reverse reasoning on the third-party damage Bayesian network constructed above, as shown in Table 6.

From Table 6 we can see that when a third-party damage accident occurs, the management of residents along the pipeline, the completeness of the equipment, the safety investment, the frequency of safety inspections, the safety performance, the timeliness of emergency rescue, the completeness of emergency plans, the completeness of monitoring equipment, pipeline safety protection, equipment failure rate, etc., are the main causes for an accident, with the probabilities of 95.1%, 95.4%, 97.3%, 95.2%, 95.1%, 92.4%, 92.2%, 90.5%, 90.4% and 87.2%, respectively. This shows that these are the key factors that affect the occurrence of accidents.
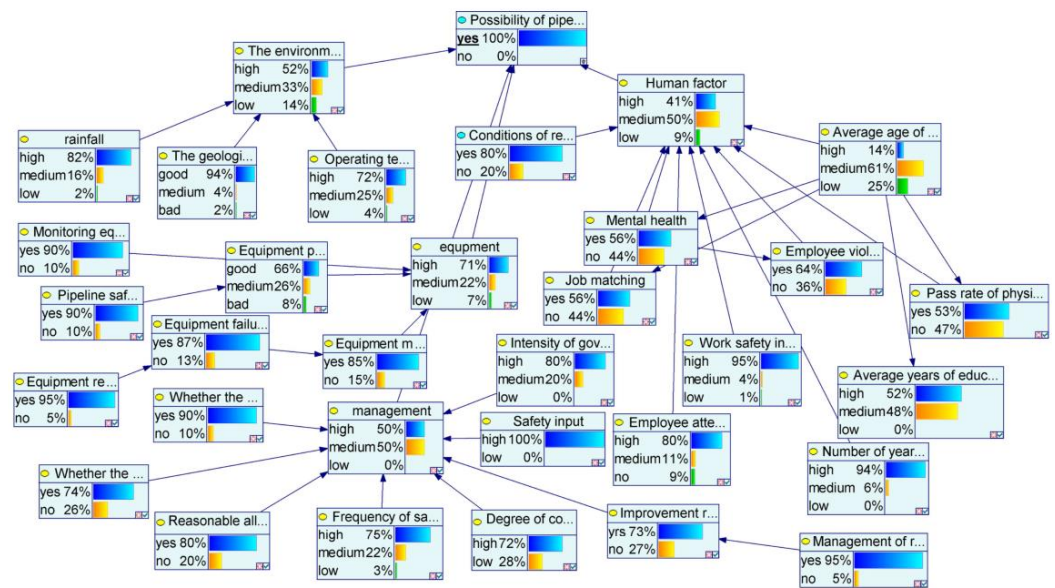
Figure 5. Reverse inference of pipeline third-party damage.

Table 6. Posteriori probability of pipeline third-party damage related nodes.

Item	Node Name	Prior Probability	Item	Node Name	Prior Probability
1	Monitoring equipment completeness	0.905	12	Government regulation	0.807
2	Pipeline safety protection	0.904	13	Completeness of emergency plan	0.922
3	Equipment completeness	0.954	14	Timeliness of emergency rescue	0.924
4	Equipment replacement	0.665	15	Emergency supplies	0.749
5	Equipment failure rate	0.872	16	Rescuer allocation	0.803
6	Geological structure	0.943	17	Management of residents along the pipeline	0.951
7	Operating temperature	0.725	18	The average age of employees	0.614
8	Precipitations	0.824	19	The average level of education of employees	0.523
9	Management system completeness	0.731	20	Average years of employee training	0.537
10	Safety inspection frequency	0.952	21	Job match	0.562
11	Safety investment	0.973	22	Safety performance	0.951

4.5. The Largest Cause Chain Analysis of Accidents

In the Bayesian network, the critical path of an accident and the corresponding risk source are found by analyzing the largest causal chain of the accident. Using the function "Strength of influence" of the GeNIe Modeler, we can find the main factors affecting the third-party damage to the pipeline, as shown in Figure 6.

Through the analysis of the largest causal chain of an accident, it can be found that the completeness of monitoring equipment, equipment completeness, pipeline safety protection, equipment failure rate, equipment maintenance, personnel operations, emergency plan, emergency rescue capabilities, and the management of residents along the pipeline, etc., are the main factors that cause pipelines' third-party damage accidents.
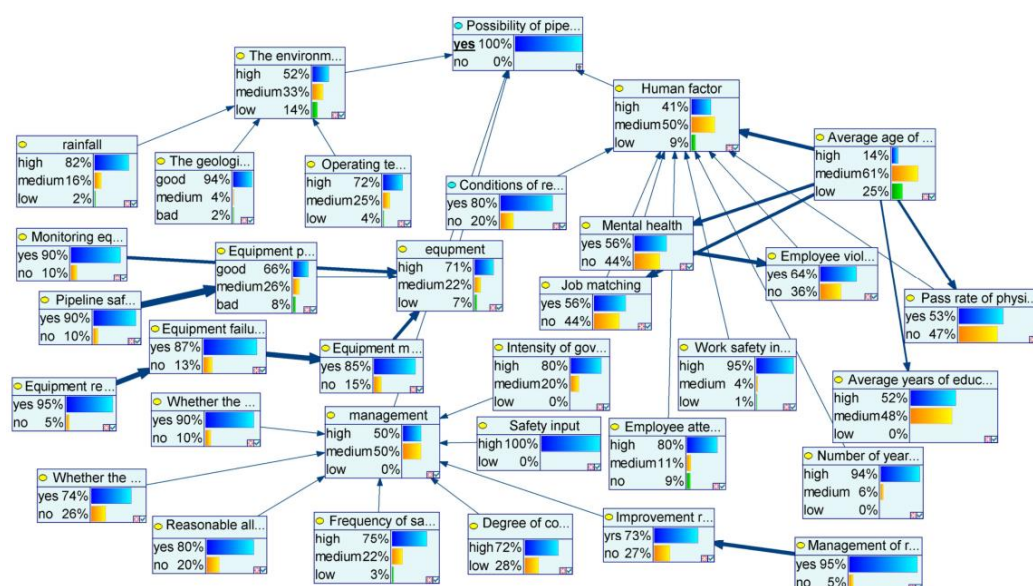
Figure 6. Analysis of the most probable cause of an accident.

5. Conclusions

This paper built a model for the third-party damage of natural gas pipelines based on a Bayesian network, and it was found the top five factors that have significant influence on the third-party damage are the safety investment, the completeness of equipment, safety inspection frequency, the management of residents along a pipeline, and safety performance, with the posteriori probability in the model of 97.3%, 95.4%, 95.2%, 95.1% and 95.1%, respectively. Based on this, some conclusions were obtained to improve the safety level of pipeline operations:

(1) It is necessary to secure investment for safety, both in hardware and software.
(2) Key equipment such as distributed optical fiber and UAV can prevent pipelines from being damaged by third-party construction effectively by measuring pipeline vibrations and detecting ground anomalies. Thus, it is important for pipeline operation companies to make sure they work well and are in a good condition.
(3) Conduct safety inspections with proper frequencies and routine safety performance assessments within an organization.
(4) Build a management system for the residents along a pipeline.
(5) Conduct routine safety performance reviews, including employer code of conduct reviews.

As the model was applied in the Zhejiang Zheneng Natural Gas Operation Co., Ltd., new cases will be collected and analyzed to assess the validity of the model periodically. The risk factors of third-party damage to the pipelines were analyzed based on the Bayesian network model from the four aspects of personnel, environment, management, and equipment. The pipeline accidents caused by factors such as the delayed manifestations of technological defects (technology impact), sudden emergency impacts (e.g., landslides) and stress-corrosion manifestations (degradations of the pipe wall) are not included in this paper, but these are also important factors that lead to pipeline accidents. We will take them into consideration in the research of comprehensive assessments of pipeline risks and make a division of the accidents in future work.

Author Contributions: Conceptualization, B.Z., J.G. and C.S.; Methodology, B.Z., X.Y. and J.W.; Software, X.Y. and J.W.; Formal Analysis, J.G., F.L. and C.S.; Investigation, J.W. and B.H.; Resources, B.Z.; Data Curation, X.Y., J.G., F.L. and D.S; Writing—Original Draft Preparation, X.Y. and J.W.; Writing—Review and Editing, J.W. and B.H.; Visualization, J.W., B.H., F.L. and D.S.; Supervision, B.Z.; Funding Acquisition, B.Z. and B.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yuan, M.; Zhang, H.; Wang, B.; Shen, R.; Long, Y.; Liang, Y. Future Scenario of China's Downstream Oil Supply Chain: An Energy, Economy and Environment Analysis for Impacts of Pipeline Network Reform. *J. Clean. Prod.* **2019**, *232*, 1513–1528. [CrossRef]
2. Chen, Y.; Jiang, H.; Yaohan, W.U.; Yao, A.; Han, X. Study on Risk Assessment Method of Urban Oil Pipeline Based on Bow-Tie Model. *J. Saf. Sci. Technol.* **2016**, *12*, 148–152.
3. Zhang, J. Pipeline Risk Assessment Method Based on Principle Component-Clustering Analysis. *Oil Gas Storage Transp.* **2014**, *33*, 139–143.
4. Guo, Y.; Fu, X.; Meng, T.; Liu, K.; Wang, D. Risk Assessment of City-Gas PE Pipeline Based on Variable Weights and Grey Theory. *Press. Vessel Technol.* **2015**, *3*, 59–65.
5. Yavorskyi, A.V.; Karpash, M.O.; Zhovtulia, L.Y.; Poberezhny, L.Y.; Maruschak, P.O. Safe Operation of Engineering Structures in the Oil and Gas Industry. *J. Nat. Gas Sci. Eng.* **2017**, *46*, 289–295. [CrossRef]
6. Yatsyshyn, T.; Polutrenko, M.; Poberezhny, L.; Maruschak, P.; Martynyuk, R. Assessment of Factors of Technogenic and Environmental Risks at Paticular Stages of the Life Cycle of Oil and Gaswells. *Procedia Struct. Integr.* **2022**, *36*, 362–369. [CrossRef]
7. Jo, Y.D.; Ahn, B.J. A Method of Quantitative Risk Assessment for Transmission Pipeline Carrying Natural Gas. *J. Hazard. Mater.* **2005**, *123*, 1–12. [CrossRef]
8. Khakzad, N.; Khan, F.; Amyotte, P. Dynamic Safety Analysis of Process Systems by Mapping Bow-Tie into Bayesian Network. *Process Saf. Environ. Prot.* **2013**, *91*, 46–53. [CrossRef]
9. Khakzad, N.; Khan, F.; Amyotte, P. Safety Analysis in Process Facilities: Comparison of Fault Tree and Bayesian Network Approaches. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 925–932. [CrossRef]
10. Morales, J.; Yu, W. Improving Neural Network's Performance Using Bayesian Inference. *Neurocomputing* **2021**, *461*, 319–326. [CrossRef]
11. Mamdikar, M.R.; Kumar, V.; Singh, P. Dynamic Reliability Analysis Framework Using Fault Tree and Dynamic Bayesian Network: A Case Study of NPP. *Nucl. Eng. Technol.* **2022**, *54*, 1213–1220. [CrossRef]
12. Wang, W.; Dong, C.; Liu, L.; Feng, L.I. Dynamic Risk Analysis of Urban Buried Gas Pipeline Network Based on Bayesian Network. *J. Saf. Sci. Technol.* **2019**, *15*, 55–62.
13. Li, Z.; Feng, H.; Liang, Y.; Xu, N.; Nie, S.; Zhang, H. A Leakage Risk Assessment Method for Hazardous Liquid Pipeline Based on Markov Chain Monte Carlo. *Int. J. Crit. Infrastruct. Prot.* **2019**, *27*, 100325. [CrossRef]
14. Han, Z.Y.; Weng, W.G. An Integrated Quantitative Risk Analysis Method for Natural Gas Pipeline Network. *J. Loss Prev. Process Ind.* **2010**, *23*, 428–436. [CrossRef]
15. Beuzen, T.; Marshall, L.; Splinter, K.D. A Comparison of Methods for Discretizing Continuous Variables in Bayesian Networks. *Environ. Model. Softw.* **2018**, *108*, 61–66. [CrossRef]
16. Feng, L.A.; Ww, B.; Sd, C.; Fk, D.; Jiang, X.A.; Jy, A. Analysis on Accident-Causing Factors of Urban Buried Gas Pipeline Network by Combining DEMATEL, ISM and BN Methods. *J. Loss Prev. Process Ind.* **2019**, *61*, 49–57.
17. Wu, Z.; Wang, C. Limiting Spectral Distribution of Large Dimensional Spearman's Rank Correlation Matrices. *J. Multivar. Anal.* **2022**, *191*, 105011. [CrossRef]
18. Zhang, W.; Feng, W.; Zhao, H.; Zhao, Q. Rapidly Learning Bayesian Networks for Complex System Diagnosis: A Reinforcement Learning Directed Greedy Search Approach. *IEEE Access* **2020**, *8*, 2813–2823. [CrossRef]
19. Lee, C.; van Beek, P. Metaheuristics for Score-and-Search Bayesian Network Structure Learning. In *Advances in Artificial Intelligence*; Mouhoub, M., Langlais, P., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 129–141.
20. Han, M.; Wang, Z.; Liu, Y. Selecting One Dependency Estimators in Bayesian Network Using Different MDL Scores and Overfitting Criterion. *J. Inf. Sci. Eng.* **2014**, *30*, 371–385.
21. Krueger, R.; Bansal, P.; Buddhavarapu, P. A New Spatial Count Data Model with Bayesian Additive Regression Trees for Accident Hot Spot Identification. *Accid. Anal. Prev.* **2020**, *144*, 105623. [CrossRef]
22. Cui, Y.; Quddus, N.; Mashuga, C.V. Bayesian Network and Game Theory Risk Assessment Model for Third-Party Damage to Oil and Gas Pipelines. *Process Saf. Environ. Prot.* **2019**, *134*, 178–188. [CrossRef]
23. Cai, Q.; Abdel-Aty, M.; Lee, J.; Huang, H. Integrating Macro- and Micro-Level Safety Analyses: A Bayesian Approach Incorporating Spatial Interaction. *Transp. A Transp. Sci.* **2019**, *15*, 285–306. [CrossRef]