Risk Assessment of Ship Navigation Collision in Inland Waterway Transportation System Based on Bayesian Method

Lupeng Yang\textsuperscript{1*}, Zhiwei Xing\textsuperscript{1} and Te Shi\textsuperscript{1}

\textsuperscript{1} School of Energy and Power Engineering, Wuhan University of Technology, Wuhan, Hubei, 430063, China

*Corresponding author's e-mail: 273495@whut.edu.cn

Abstract. Due to the high risk and uncertainty of ship navigation, the identification, early warning and control of ship navigation risks have always been studied by scholars. In recent years, water traffic accidents have occurred frequently, resulting in an incalculable number of casualties and property losses, which makes ship navigation risk management more and more a research hotspot. Ship navigation risk is a complex system consisting of a number of risk factors. The assessment of it is a systematic project aimed at clarifying the risk factors that need to be controlled. The purpose of this paper is to use Bayesian methods to assess the risk of navigational collisions in inland waterway transport systems. In this paper, the Bayesian network and the inland waterway transportation system ship navigation collision risk simulation are combined, and the concept of the ship navigation collision risk analysis model of the Bayesian network inland waterway transportation system is proposed. This method essentially utilizes three important characteristics of Bayesian network itself—complex relation representation ability, probability uncertainty representation ability and causal reasoning ability. Through the expert experience and ship navigation collision risk simulation of inland waterway transportation system the data is learned, the knowledge contained therein is mined, and the ship navigation collision risk analysis model of the inland waterway transportation system is established, and the risk assessment analysis is carried out based on the model. Through a large number of experiments, the proposed method can better model the navigation state of the ship, quantify the navigation risk of the ship, and identify the navigation anomaly data.

1. Introduction

With the rapid development of global trade, inland water transportation is becoming increasingly busy, and ships are developing in the direction of large-scale and high-speed. The increase in the tonnage of the ship not only reduces the kinetic performance of the stern itself, but also increases the requirements for the navigation channel and the port [1]. In recent years, with the increasing traffic flow of ships, the risks faced by navigation vessels as the main body of inland traffic operations are also increasing, and the complexity and pressure of inland river traffic management are increasingly prominent [2]. Every year, a large number of river collisions such as collision, stranding, anchoring and drifting occur in various sea areas around the world, which not only causes huge economic losses and a large number of casualties, but also may pollute the ecological environment of the accident inland river and cause serious ecology. Destruction [3]. Therefore, real-time assessment and degree assessment of single-system and systemic risks such as ship collision, shallowness, error and drifting of the inland river traffic system under the jurisdiction has become one of the important research topics in the field of intelligent transportation in the sea, which is not only important. The theoretical research significance, while having important social and economic value in the risk management and
control of the ship transportation system. The number of casualties may also contaminate the ecological environment of the accidental inland river and cause serious ecological damage. Inland river traffic risk identification and assessment research has important application value for improving ship traffic management level, reducing traffic management pressure, and continuously improving ship traffic management methods and methods, implementing ship traffic risk management and risk in key monitoring waters of relevant maritime bureaus. Effective identification is very necessary [4-5]. Therefore, based on the real-time AIS data of ship navigation, it is of great theoretical significance to carry out quantitative measurement calculations, identification and evaluation of risk, such as system collision, stranding, walking and drifting towards the inland river transportation network. The performance of traffic safety early warning and control system has important technical and practical significance.

The Vessel Traffic Management (VTS) system plays an important role in improving the safety of inland river traffic and preventing ships from polluting the inland environment [6]. In recent years, China has built a number of VTS systems, covering basically the main ports and important waters along the coast of China. With the increasing traffic flow of ships, the risk of inland traffic is also increasing, and the complexity and pressure of inland traffic management are becoming more and more obvious. How to use the risk identification tools to effectively and quickly identify inland traffic risks has become the face of inland management One of the problems [7-8]. In recent years, the international maritime community has stepped up research on the identification of inland river traffic risks. For example, the European Maritime Safety Administration uses real-time traffic data to analyze ship traffic characteristics, and combines ship safety status information to establish waters covering the waters of all EU member states. High Risk Ship (HRV) Identification System. The US and Canadian Coast Guard also developed a tragic system for dangerous ships in inland waters [9-10].

In this paper, the Bayesian network and the inland waterway transportation system ship navigation collision risk simulation are combined, and the concept of the ship navigation collision risk analysis model of the Bayesian network inland waterway transportation system is proposed. This method essentially utilizes three important characteristics of Bayesian network itself—complex relation representation ability, probability uncertainty representation ability and causal reasoning ability. Through the expert experience and ship navigation collision risk simulation of inland waterway transportation system The data is learned, the knowledge contained therein is mined, and the ship navigation collision risk analysis model of the inland waterway transportation system is established, and the risk assessment analysis is carried out based on the model. Through a large number of experiments, the proposed method can better model the navigation state of the ship, quantify the navigation risk of the ship, and identify the navigation anomaly data.

2. Proposed Method

2.1. Bayesian Network Parameter Learning

The parameter learning of Bayesian network is to learn the probability distribution parameters of each node under the condition that the network nodes and their distribution relationships, ie network structure, are known. Initially, the probability distribution parameters of each node of the Bayesian network are determined according to the relevant domain knowledge. However, this method based on empirical knowledge often has a large deviation from the actual node observation data. At present, the main method used is to learn the probability distribution parameters of these parameters from the actual node observation data. This method driven by the example data has strong adaptability. Data refers to a set of actual observations of variables in the network:

$$D = \{x_1, x_2, \ldots, x_n\}, x^i = \{x^i_1, x^i_2, \ldots, x^i_m\}$$  \hspace{1cm} (1)

According to the actual observation results of the network node data, it can be divided into a complete observation data set and an incomplete observation data set. For each set of observations in the complete observation data set, each node has complete observation data. The incomplete data set
refers to the case where some network node parameters are missing or the node observations are abnormal. For the study of network parameters of incomplete data sets, the relevant approximate solution method should be used.

The purpose of node probability parameter learning for observation data with a complete data set is to find each node parameter $\theta_i$ that can summarize the network data sample set in a probabilistic form $p(x_i | \theta_i)$. For the learning of network parameters, it is generally necessary to first determine a certain probability distribution family, and then use some methods to estimate these distribution parameters.

2.2. Maximum Likelihood Estimation (MLE) Method

The maximum likelihood estimation MLE method is based on traditional statistical analysis ideas. The general form of the likelihood function is:

Assuming that the distribution function of the variable is known, the maximum likelihood value can be obtained by performing the Lagrangian multiplier method on the equation (2), and then the estimated value of the parameter is obtained. According to the principle of statistics, we can know that the MLE method has the following advantages:

(1) Consistency. As the number of observations on the sample increases, the parameters will gradually converge to the best possible value (actual probability value).

(2) Progressive effectiveness. Look for the parameter $\theta_i$ that can make the sample occur with the maximum probability. The parameter $\theta_i$ is as close as possible to the actual probability value. The more the number of sample instances, the better the degree of the actual probability value is close.

(3) Indicates flexibility. The estimated probability distribution effect is not affected by the different form of parameter distribution.

2.3. Bayesian Method

The biggest difference between the Bayesian method and the traditional statistical method is the difference in perception between the two. Traditional statistical methods simply regard the probability of an event as an infinite proximity to the frequency of occurrence of an event, while the Bayesian method considers uncertainty to be a cognitive experience of an event, which is preceded by the object. The combination of subjective experience perception and the observed phenomena now determines the level of awareness of the thing. Therefore, Bayesian method learning network node parameters should be composed of two aspects: the prior knowledge of the problem before observation and the instance data obtained after observation. In the Bayesian network parameter learning, the selection of the prior distribution of the node parameters and the determination rules of the distribution parameters are included. The purpose of learning is to find a specific algorithm to link the two. There are two main rules:

(1) Conjugate distribution family

The selection of the conjugate distribution, that is, the posterior distribution and the prior distribution are all of the same distribution type. Its general description is:

Let the sample $X_1, X_2, \ldots, X_n$, the conditional distribution of the parameter $\theta$ is $p(x_1, x_2, \ldots, x_n | \theta)$, and if the posterior density $P(\theta | x)$ determined by the prior distribution density function $\pi(\theta)$ and the $\pi(\theta)$ belong to the same type distribution, then $\pi(\theta)$ is called the conjugate distribution of $P(\theta | x)$. When the density distribution of the sample distribution and the prior distribution are exponential functions, the results of multiplication are added, and the results still belong to the same type of exponential function, only a constant proportional factor. The family of exponential functions includes binomial distribution, polynomial distribution, normal distribution, Poisson distribution and Poisson distribution. Since the computational difficulty of the non-conjugated distribution is very large, in contrast, the conjugate distribution is only required to multiply the prior
distribution, and the calculation is very simple. It can be said that the use of the conjugate distribution family paves the way for the Bayesian network in actual engineering use.

(2) Maximum entropy principle
The principle of maximum entropy: the prior distribution without explicit information should take the distribution with the largest entropy within the range of the parameter $\theta$. It can be proved that a uniform distribution of random variables is a necessary and sufficient condition for its entropy to be maximized. Therefore, it is assumed that the prior distribution of the relevant variables in the non-information network is uniformly distributed in accordance with the principle of maximum entropy in information theory, which can maximize the entropy of random variables. Under the premise that there is no information to determine the prior distribution of network variables, it is reasonable to use uniform distribution as its prior distribution.

3. Experiments

3.1. AIS Data Environment
The AIS data obtained in this paper is the ship navigation AIS data provided by the Maritime Safety Administration within its jurisdictional waters in January 2019; the frequency of AIS data sampling is related to the speed of the ship’s navigation speed. The faster the speed, the higher the frequency of sampling. Raw data is not equally spaced. Then, the data is pre-processed first: because the data with long time interval means that the speed is relatively slow, then the frequency of the speed change can be considered to be limited, and then the possibility of a sudden change in speed between the two speeds can be assumed. The nature is small, and this assumption is realistic. In this paper, the cubic polynomial difference is used to process the data so that the ship data in the same time interval is equal. At the same time, the ship longitude data is converted to the virtual coordinate axis of the earth for subsequent analysis. At the same time, for the lack of data, the average of the values before and after is used. This part of the experiment used AIS data for ship navigation on January 6, 2019.

3.2. Collision Risk Severity Calculation
According to the analysis of the risk factors of navigational collisions mentioned above, in addition to the traffic flow and human error rate, the forecasting model must also consider the influence of geographical factors, hydrometeorological factors and ship factors. Drawing on the framework of the FSA risk matrix, this paper uses the fuzzy comprehensive evaluation method to synthesize the degree of comprehensive impact of other factors that may cause the event, which is defined as “the severity of the collision risk”; Equivalent to the probability of occurrence of the event, the combination of the two, the generation risk matrix, can estimate the collision risk level of the navigation ship.

4. Discussion

4.1. Collision Risk Utility Function
From $F(t)$, we can determine the trend of system collisions at several moments. Combined with the collision risk of subsystems, we can know which subsystems have higher collision risk. When the system collision risk experienced time 1 to 2, increasing from 0.3439 to 0.9271 (increase 0.5732), only one subsystem had a large change in collision risk. The change from 0.1 to 0.9; when the system collision risk experienced 2 to 5, from 0.9271 to 0.9999 (increase 0.0718), the collision risk of the three subsystems changed from 0.1 to 0.9. Therefore, when the risk of system collision does not change much, the collision risk of its subsystem may change sharply, indicating that the degree of change of the system collision risk cannot directly reflect the change of the magnetic collision risk of its subsystem. The degree of system risk change is shown in Table 1.
Table 1. Degree of system risk change.

| Time | \( f_1(t) \) | \( f_2(t) \) | \( f_3(t) \) | \( f_4(t) \) |
|------|----------|----------|----------|----------|
| 1    | 0.1      | 0.1      | 0.1      | 0.1      |
| 2    | 0.1      | 0.1      | 0.9      | 0.9      |
| 3    | 0.1      | 0.1      | 0.9      | 0.9      |
| 4    | 0.1      | 0.9      | 0.9      | 0.9      |
| 5    | 0.9      | 0.9      | 0.9      | 0.9      |

After introducing the utility function, when a subsystem risk increases from 0.1 to 0.9 in a large system, the value of the system's collision risk utility function increases by 2.1972, which is a clearer reflection of the degree of risk change than the increase in the system collision risk. Changes in utility function values allow maritime managers to better understand the relationship between subsystem collision risk and system collision risk, and take appropriate steps to mitigate risk.

4.2. Bayesian Network Sensitivity Analysis

Taking 10% as the step size, the sensitivity analysis of each risk factor layer node is carried out, and the probability that the node state is "1" increases by 10%, 20%, 30%, 40%, 50% respectively. The posterior probability of one of the accidents. It has been verified that the sensitivity of accidents at all levels is basically the same, so the sensitivity of each level of accidents is merged into the state of “accidents”, and the probability of accidents is used as the standard for measuring accidents. The sum of the probabilities of accidents at all levels. Taking the quality of the crew as an example, the data shown in Table 2 is obtained.

Table 2. Sensitivity analysis of crew quality.

| Risk factor probability change | Initial value | 15%  | 25%  | 35%  | 45%  | 55%  |
|-------------------------------|---------------|------|------|------|------|------|
| Probability that the node status is "1" | 40%           | 45%  | 55%  | 65%  | 75%  | 85%  |
| The sum of accident probability at all levels | 11.6%         | 13.4%| 15.1%| 16.2%| 17.9%| 21.0%|

The sensitivity of each node is reflected in Figure 1 as the slope of each curve. Therefore, as can be seen from Figure 1, the sensitivity ranking of each key risk factor is: wind > crew quality > visibility > maritime supervision > health status > ship maintenance > flow. Among them, the sensitivity of wind and crew quality is much higher than other factors, and the visibility and visibility of maritime supervision are relatively obvious. The reason why the sensitivity of the flow is the smallest is that in this case, the ship A sails in the first-level waters of the Yangtze River. As long as there is no windy weather, in most cases, the hydrological conditions are relatively normal, so the water flow is relatively affected by the wind. Larger, the effect of increasing its prior probability alone is not obvious. When formulating the navigation risk management measures of Ship A, it is considered that avoiding the windy weather will largely avoid the adverse hydrological conditions.
5. Conclusions
This paper presents a new system collision risk definition and calculation model. In order to better meet the work requirements of the maritime management department, the concept of system collision risk for ship navigation in a certain sea area is proposed, and the information fusion method is applied based on the traditional collision risk based on TCPA and DCPA and the risk of collision on the ship. The calculation model of basic system collision risk is established. The experiments show that the proposed method can better model the navigation state of the ship, quantify the navigation risk of the ship, and identify the navigation anomaly data.

References
[1] Perera, L. P., & Soares, C. G. (2015) “Collision Risk Detection and Quantification in Ship Navigation with Integrated Bridge Systems” Ocean Engineering, 10(9), pp.344-354.
[2] Zhang, W., Kopca, C., Tang, J., Ma, D., & Wang, Y. (2017) “A Systematic Approach for Collision Risk Analysis Based on Ais Data”, Journal of Navigation, 70(5), pp.1117-1132.
[3] Jingxian Liu, Feng Zhou, Zongzhi Li. (2016) “Dynamic Ship Domain Models for Capacity Analysis of Restricted Water Channels”, Journal of Navigation, 69(3),pp.481-503.
[4] Barthélémy, C., & Armani, G. (2015) “Risk of Navigation for Marine Traffic in the Malacca Strait Using Ais”, Procedia Earth & Planetary Science, 14(6), pp.33-40.
[5] Ling, S., Zhang, H., Wei, L., & Fei, C. (2018) “Research on Risk Assesment and Control of Inland Navigation Safety”, International Journal of System Assurance Engineering & Management, 9(5), pp.1-10.
[6] Brewerdotherty, T., Colamosca, B., Gerhardtfalk, C., Livingston, D., Martin, L., & Elsayed, E. A., et al. (2016) “Collision Risk Modeling in the Northern Pacific Airspace under Separation Reduction and Improvements in Navigational Performance”, Air Traffic Control Quarterly, 14(4), pp.257-282.
[7] John, A., & Osue, U. J. (2017) “Collision Risk Modelling of Supply Vessels and Offshore Platforms under Uncertainty”, Journal of Navigation, 70(4), pp.1-17.
[8] Kang, Z., Chen, J., & Xiang, L. (2017) “Optimal Collision-Avoidance Manoeuvres to Minimise Bunker Consumption under the Two-Ship Crossing Situation”, Journal of Navigation, 71(1), 1-18.
[9] Perera, L. P., & Mo, B. (2016) “Data Analysis on Marine Engine Operating Regions in Relation to Ship Navigation”, Ocean Engineering, 12(8), 163-172.
[10] Inazu, D., Ikeya, T., Waseda, T., Hibiya, T., & Shigihara, Y. (2018) “Measuring Offshore Tsunami Currents Using Ship Navigation Records”, Progress in Earth & Planetary Science, 5(1), pp.38.