Risk Assessment of Ship Navigation Using Bayesian Learning

HU Shenping, CAI Cunqiang, FANG Quangen
Merchant Marine College, Shanghai Maritime University, Shanghai, China

Abstract - **Risk has a random uncertainty. Risks associated with a ship navigation at sea are analyzed to solve the problem of uncertainty and a developing method is applied to be feasible to work out. Based on Bayes' point estimate and Bayesian learning to estimate the traffic accidents related to ship navigation, an analysis model is established for the quantitative risk assessment (QRA) of the vessel traffic system at sea. After the analysis on occurrence likelihood of the accidents related to ship traffic, a structure on the basis of Bayesian networks is developed to obtain the QRA of their relative risks. QRA is also put forward after analyzing the features and situations of the vessel traffic system and identifying the corresponding feature including characteristics of those hazards. The risk distributions of ship navigation are described and results are presented on QRA in relation to various features by using this method. This method, verified in the cases of QRA, turns out to be feasible by the use of machine learning.**

Keywords - **Ship, risk assessment, Bayesian network, machine learning, safety factor**

I. INTRODUCTION

With the fast development of technology, people have obtained further understanding of accidents in ship navigation and safety management. Such changes are essentially the results of the conversion from qualitative analysis of accidents and safety to quantitative analysis of risks [1]. Quantitative risk analysis (QRA) in ship traffic is one of the most important issues under research.

Risk has a random uncertainty. In the case of model uncertainty, when there exist several possible models to describe a phenomenon, a Bayesian approach can be used to include all the candidate models by assigning model weight (the probability of each model being correct) and integrating the effects of all the models. When there is observation/data available, the model may be updated and shifted to the more appropriate model. This approach has been applied to the uncertainty of probability distribution type and linear regression model uncertainty problems in statistics, and was recently used to account for mechanical model uncertainty.

Here, the accidents are viewed as evident, this is, instantiations of some or all of the random variables describing the domain. So, with the development of artificial intelligence, QRA can be achieved using statistical learning methods. Approaches for learning probability models, primarily network reasoning can store and recall specific instance.

This paper, on the basis of analysis on the risk influence factors in a ship navigation system, introduces a new risk estimation method, in which the risk assessment model of the ship navigation system is established by using Bayesian statistics and network learning.

II. RISK ASSESSMENT MODEL OF SHIP NAVIGATION

When the uncertainty of incident occurrence is concerned, risk is defined as the potential of system failure that would result in losses. This kind of risk, therefore, is involved in two factors, that is, the likelihood of a incident occurrence, and the potential consequence caused by its occurrence. Ayyub[3] has put forward the following equation to evaluate risk using the likelihood and consequence factors.

$$Risk\left(Occurence, Consequence, Time\right) = Likelihood\left(\frac{Occurrence}{Time}\right) \times Impact\left(\frac{Consequence}{Occurrence}\right) \quad (1)$$

In Equation (1), the risk can be used to refer to ship traffic risks, *Likelihood* means the probability or frequency of the event concerned and *Impact* is the consequence of its occurrence. The analysis of *Likelihood* (or frequency)requires to establish a particular status (or to assume the frequency of events concerned) in a defined probability universe (Ω, Θ, C).

Suppose a hazard $R_j(x, y, t)$ results in the occurrence of a particular accident or incident. The possible range of impact C is a universe Ω, which covers a planar area Θ. At the occurrence of the accident or incident, the risks of potential danger to a certain target in the universe resulted from the occurrence should be in accordance with the following Eq.(2).

$$\begin{aligned} R_j(x,y,t) &= \sum_{i=1}^{n} P_i C_i(x,y,t),\ j = 1,2,\cdots n \\ Risk &= \sum_{j=1}^{m}\sum_{i=1}^{n} P_{ij} C_{ij}(x,y,t) \end{aligned} \quad (2)$$

where x, y shows the 2-D location particular of risks, while t the time particular of risks, 3-D factors.

For instance, a certain hazard $R_j(x, y, t)$ in a ship traffic system has resulted in a ship accident *i*, which, in

1-4244-1529-2/07/$25.00 ©2007 IEEE

general, can be classified into the traffic accident set *(collision, stranding, contact, etc)* . P_{ij} stands for the possibility of occurrence of i caused by hazard j , and $C_{ij}(x,y,t)$ refers to the causal relationship between hazard j and accident *i*.

The computation of P_i has always been a major subject in the study of risk management. The appropriate likelihood can be obtained using historical data, mathematical models and expert judgments.

III. BAYESIAN METHODOLOGY

Bayesian theory has made great achievement covering most research areas in artificial intelligence (AI), which includes deduction of causality, description of uncertainty, identification of modes and clustering analysis. The achievements have been introduced to the study of risk analysis recently[4][5].

A. Bayes' point estimation of likelihood

Assume that sample set $A_1, A_2, \cdots A_n$ constitutes a complete and independent universe Θ . $A_i (i \in [1,n])$ is the occurred events in E and B refers to one existing event $p(B) \neq 0$ .It can be concluded that that the formula of the Bayes' rule can be expressed in accordance with the following Eq.(3):

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)}, i = 1,2,\cdots n \qquad (3)$$

At the occurrence of A_j , the discrete random parameter θ is defined as $\theta = \theta_j = A_j (j = 1,2,\cdots n)$, when the probability should conform to the following equation:

$$\pi(\theta_j) = P(\theta = \theta_j) = P(A_j), j = 1,2,\cdots n \qquad (4)$$

Therefore, $\{\pi(\theta_j), j = 1,2,\cdots n\}$ constitutes the prior probability distribution of random variables. This probability can be achieved on the basis of historical statistics or subjective judgments on the probabilities of incidents. Prior probability can normally be identified through three approaches mentioned above.

Suppose X , a discrete random variable related to B , accords with the following Eq.(5):

$$X = \begin{cases} x_1, \ when\ B = yes \\ x_2, when\ B = no \end{cases} \qquad (5)$$

From the equation (1), we can deduct the posterior probability (or conditional probability) under the condition of B occurred:

$$\pi(\theta_i \mid X = x_1) = \frac{P(x_1 \mid \theta_i)\pi(\theta_i)}{\sum_{j=1}^{n} P(x_1 \mid \theta_j)\pi(\theta_j)}, i = 1,2,\cdots n \qquad (6)$$

As a result, this kind of probability distribution $\{\pi(\theta_i \mid x_1), i = 1,2,\cdots n\}$ can be viewed as the distribution of $\{\pi(\theta_j), j = 1,2,\cdots n\}$ through the sample observation x .

B. Bayesian network structure learning

A Bayesian network (BN) is used to model a domain containing uncertainty in some manner. In the past, the term causal probabilistic networks have been used. A BN is a directed acyclic graph (DAG) where each node represents a random variable. Each node contains the states of the random variable it represents and a conditional probability table (CPT) or in more general terms a conditional probability function (CPF). The CPT of a node contains probabilities of the node being in a specific state given the states of its parents. The following example demonstrates what all this means.

Assuming variables set X and a discrete variable θ representing uncertainty of network structure, its hypnoses of possible network is S^h , and its prior probability conforms to $P(S^h)$. Under the condition of random samples D , its posterior probability is $P(S^h \mid D)$.

$$P(S^h \mid D) = \frac{P(S^h, D)}{P(D)} = \frac{P(S^h)P(D \mid S^h)}{P(D)} \qquad (7)$$

where $P(D)$ is a normalized constant.

One highly practical Bayesian learning method is the naive Bayes learner, often called the naive Bayes classifier. Probably the most common network model used in Bayesian learning is naive Bayes model. In this model the class variable X is the root and the attribute variables A are the leaves. With observed attribute values $a_1 \quad a_2 \quad \cdots \quad a_m$, the likelihood of each class is given by Eq.(8)

$$P(X \mid a_1 \quad a_2 \quad \cdots \quad a_m) = \alpha P(X) \prod_{j}^{m} P(a_j \mid X) \qquad (8)$$

The Bayesian approach to classifying the new instance is to assign the most probable target value, υ_{MAP} , given the values that describe the instance. Eq.(9)

$$\upsilon_{MAP} = \arg\max_{x_i \in B} P(x_i \mid a_1 \quad a_2 \quad \cdots \quad a_m) \qquad (9)$$

Substituting this into Eq.(10), we have the approach used by naive Bayes classifier.

$$\upsilon_{NB} = \arg\max_{x_i \in B} P(x_i) \prod_{j}^{m} P(a_j \mid x_i) \qquad (10)$$

where υ_{NB} denotes the target value output.

This equation can be equivalently expressed in terms of maximizing the *ln* or alternatively, minimizing the negative of this quantity.

$$\ln \upsilon_{NB} = \arg\min_{x_i \in A} \{-\ln P(x_i) - \sum_{j=1}^{m} \ln P(a_j \mid x_i)\} \qquad (11)$$

C. Bayesian parameter learning

The structure of the Bayes net is given and the parameter are trying to be learnt. Up to this point we have estimated probabilities by fraction of times the event is observed to occur over the total number of opportunities. To avoid difficulty this paper adopts a Bayesian approach to estimating the probability using the *m*-estimation defined as follows Eq.(12).

$$P(a_j \mid x_i) = \frac{n_c + m \bullet p}{n + m} \tag{12}$$

Here, n is defined as total number of training examples for which x_i occurred and n_c is the number of these for which a_j occurred. p is our prior estimate of the probability we wish to determine, and m is a sample size constant.

IV. BAYESIAN NETWORK ANALYSIS ON THE RISKS OF SHIP NAVIGATION

Vessel traffic is a ship operation system[1] that consists of crew, ship, information and environment, which can also be analyzed by dividing it into subjective factors and objective factors. The physical features of the system can be expressed as a Bayesian Belief Network in which such features are treated as samples.

A. Bayesian estimation of likelihood in ship navigation

This paper introduces historical data into calculation of the objective prior probability.

The statistical accidents in a summary activity can be used to describe the frequency of accident, which can act as the objective standard for risk assessment, and which is also a key parameter in simulation research on system safety. Ship traffic accident is random event, which conforms to binomial distribution of ship flow. The statistic laws of accident can be described with the digital features(mean, deviation, variation) of probability distribution. According to accident statistics, when there are sufficient statistic intervals or accident samples, the distribution function of accident samples in a ship navigation system meets the following formula:

$$P(X = k) = C_n^k \theta^k (1-\theta)^{n-k}, k = 0,1,2...,n \tag{13}$$

the parameter θ has its prior distribution in accordance with the following Eq.(14):

$$\pi(\theta) = \frac{1}{\beta(a,b)} \theta^{a-1} (1-\theta)^{b-1} \tag{14}$$

Furthermore, the posterior distribution of the accidents in ship's navigation system meets the following equation:

$$\pi(\theta \mid k) = \frac{1}{\beta(a+k, n+b-k)} \theta^{a+k-1} (1-\theta)^{n+b-k-1} \tag{15}$$

The formula describes the probability of k accidents in n ship summary activity. The posterior distribution meets the equation that follows Eq.(16):

$$\bar{\theta} = \frac{1}{m}\sum_{i=1}^{m} \theta_i, S_\theta^2 = \frac{1}{m-1}\sum_{i=1}^{m} (\theta_i - \bar{\theta})^2 \tag{16}$$

Then it is in accordance with the following Eq.(17):

$$\begin{cases} \hat{\theta}_E = \dfrac{a+k}{a+b+n}, \hat{\theta}_{MD} = \dfrac{a+k-1}{a+b+n-2} \\ \hat{a} = \bar{\theta}[\dfrac{(1-\bar{\theta})\bar{\theta}}{S_\theta^2} - 1] \\ \hat{b} = (1-\bar{\theta})[\dfrac{(1-\bar{\theta})\bar{\theta}}{S_\theta^2} - 1] \end{cases} \tag{17}$$

B. Statistic sample in ship navigation

The hazard identification(HazID) is a difficult job in QRA. But the probability influence diagram can be useful for doing it. Normally the navigation system of ship consists of 4 factors, namely liveware, hardware, software and environment which include risk of vessel traffic, ship's motion, waterways, fairway, and pilotage patterns. All features in the system can be viewed as a discrete and independent evidence. For instance, this item "nation" is used to express the features of navigation water areas including domestic, oversea and high sea.

Bayesian network model is constructed by means of topology. Each node can be illustrated in the following expression:

$$\begin{aligned} &A_0 = \langle very, serious, less, slight, safe \rangle \\ &A_1 = \langle domestic, oversea, highsea \rangle \\ &A_2 = \left\langle \begin{matrix} navigating, berthing, unberthing, anchoring, turning, \\ mooring, unmooring, alongside\_ship, departure\_ship \end{matrix} \right\rangle \\ &A_3 = \langle ocean, fairway, harbor, coastal, anchorage, fishingzone \rangle \\ &A_4 = \langle pilotage, non\_pilotage \rangle \\ &A_5 = \langle fog, heavyseas, normal \rangle \\ &A_6 = \langle collision, contact, agrouding, fire, wave, others \rangle \end{aligned} \tag{18}$$

V. RISK ASSESSMENT IN SHIP NAVIGATION

It is essential to compute the probability of the physical features in a ship' navigation system, in order to find out the overall and internal situation and tendency in ship safety operation within a specific time span, as well as to grasp the specific features of risks existing in the ship operation process. It is significant in implementing the detailed measures for the improvement of ship operation safety to analysis of the causes and effects in the systems, evaluate of the current safety situation, and to search for the defections and matters of events, etc. The introduction of the Bayesian probability method, therefore, is effective in the computation of risks in ship's operation.

A. Ship navigation accident probability estimation

The probability of ship traffic can be conducted on the basis of the ship transportation data in recent years. For instance, Table 1 is a demonstration of the recent ship accidents of a certain shipping company.

TABLE 1
STATISTICS OF TRAFFIC ACCIDENTS IN A SHIPPING COMPANY

NO.	n_i	k_i	θ
1	25956	39	0.150%
2	26139	42	0.161%
3	25408	36	0.142%
4	24311	18	0.074%
5	24677	18	0.073%
6	21752	9	0.041%
7	18462	10	0.054%
8	19924	10	0.050%
9	20290	11	0.054%
10	20540	8	0.039%
11	19936	14	0.070%
12	19074	11	0.058%
13	20440	10	0.047%

Based on the Bayes' statistics in section IV, the probability in the traffic fits in with the following Eq.(19):

$$\begin{cases} \hat{a} = 3.2502, \hat{b} = 4163.6 \\ \pi(\theta) \sim \beta(3.2502, 4163.6) \\ \hat{\theta}_{MD} = 0.0819\% \end{cases} \tag{19}$$

B. Ship navigation accident sample structure

The appropriate probability can be obtained by using historical data, mathematical models and expert judgments. This paper establishes an accident sample structure through the statistic accidents from 1993 to 2005. All samples would be described as physical features such as weather, motion, waterways, fairway, and pilotage patterns.

The Bayesian Belief Network can be shown as follows after Bayesian learning.

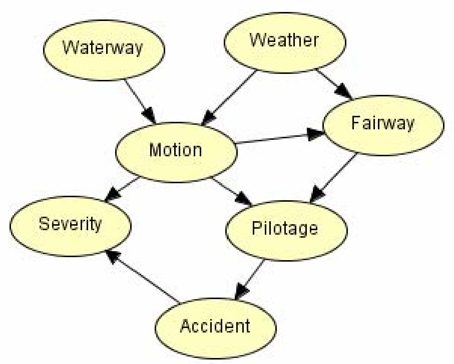

Fig.1 Bayesian Network topology diagram in ship navigation

C. Ship navigation accident parameter learning

Based on the above-mentioned calculations in section III, the conditional probability of various features $P_j(\theta \mid A_1, A_2, A_3, A_4)$ can be deducted under the condition of *Severity* occurred.

If the ranking of severity can be quantified [7] [8], the risk under the conditional probability of various features could be computed according to the following Eq.(20):

$$R_j(x,y,t) = \hat{\theta}_{MD} \cdot P_j \cdot \sum_{i=1}^{5} \omega_i C_i(x,y,t) \tag{20}$$

where ω_i shows the distribution of severity under P_j.

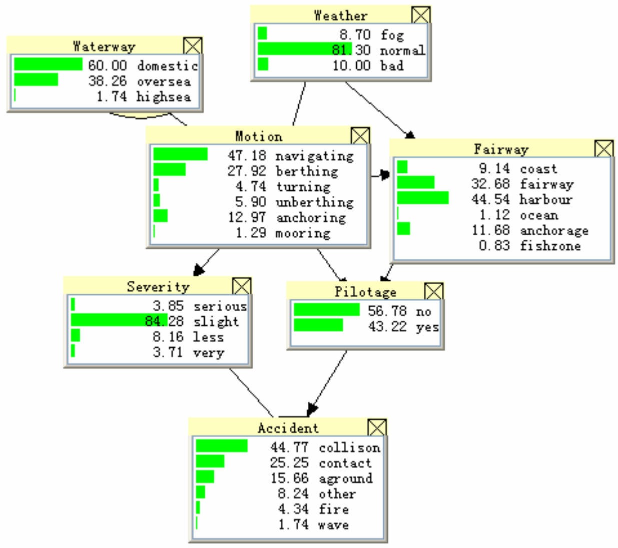

Fig.2 Bayesian learning in ship navigation

The results[2] [7] are analyzed in the following:

a) Concerning the ship motion in operating waterways A_1, A_2, higher risks exist mainly in the navigation of domestic waterway, secondly in navigating of overseas waterway, and thirdly in berthing of overseas waterway.

b) With regard to the operating waterways A_2, A_3, the following areas pose greater risks: airways in domestic waterway, harbor in overseas waterway, as well as costal areas in overseas waterway. The fairway in overseas waterway comes next in posing risks.

c) As for the ship motion in piloting water areas A_1, A_3, A_4, the main risk is in fairway – pilotage – navigation, secondly in costal area– non-pilotage – navigation as well as harbor – non-pilotage – navigation, and thirdly in fairway – non-pilotage – navigation, harbor – pilotage – berthing, and harbor – non-pilotage – berthing.

d) With reference to the comprehensive analysis of the multi-features A_1, A_2, A_3, A_4 that pose relative risks to vessel traffic, the scale of higher risk area is listed in descending order as follows: a) fairway – pilotage – domestic waterway – navigation, and b) costal area– non-pilotage– domestic waterway – navigation.

VI. CONCLUSION

The discussion of ship navigation risk has always been an important issue in the shipping industry. A lot of work has been done aiming to ensure ship safety navigation. This paper, with the introduction of Bayesian probability statistics and network reasoning, attempts to show how to conduct the quantitative computing of the system risk distribution. This method, verified in the cases of QRA, turns out to be feasible to work out the identified posterior probability by Bayesian learning, especially perception of accidents in the near future.

ACKNOWLEDGMENT

The research work was a part of work under the guidance of Shanghai Municipal Education Committee Research Project, with the contract number 05FZ26. The research was also partially supported by Shanghai Leading Academic Discipline Project, China under the contract number T0603.

The authors would like to express their sincere gratitude to Container Liners initiative for their sample support and Ocean Shipping Co. who offered their valuable feedback for evaluation of QRA research on vessel traffic.

REFERENCES

[1]C. Guedes Soars, A.P. Teixeira. Risk Assessment in Maritime Transportation. *Reliability Engineering & System Safety*, Volume 74, 2001,pp 299-309.

[2]SP Hu, QG Fang, GM Qiao, et al. Risk Assessment and Risk Controlling Option on ship navigation. *Journal of the China Navigation*, *Institute* 2005,(1), pp1-5。

[3] BM Ayyub, JE Beach, S Sarkani, et al. *Risk analysis and management for marine system*. Naval Engineers Journal , Issue 2, 2002, pp181-206

[4]CL Mei, JC Fan. Data Analysis Method, Xi'an Communication University Press, 2006.2, pp176-205.

[5]RX Zhang, S Mahadevan. Bayesian methodology for reliability model acceptance, *Reliability Engineering & System Safety*, Volume 80, 2003, pp95-103

[6] HC Wu, Bayesian system reliability assessment under fuzzy environments, *Reliability Engineering & System Safety*, Volume 83, 2004, pp277-286

[7] SP Hu, QG Fang, HB Xia. Study on Formal Safety Assessment in ship navigation and Modeling based on Relative Risk, *Proceeding of the 4th International Conference on Quality and Reliability*,August 2005, Beijing:Beijing Institute of Technology Press, pp575-582.

[8] SP Hu, QG Fang, HB Xia. et al. Formal Safety Assessment based on Relative Risks Model in ship navigation. *Reliability Engineering & System Safety*, Volume 92, Issue 3, March 2007,pp 369-377.

[9] QG Fang, SP Hu. Application of FSA to the risk assessment of the ship-pilotage, *Journal of Harbin Engineering University*, Volume 27, Issue 3, Sept 2006, pp329-33

[10] X. Jiang, S. Mahadevan. Bayesian risk-based decision method for model validation under uncertainty(J), *Reliability Engineering & System Safety*, *Volume 92, Issue 6*, *June 2007,pp 707-718*

[11] PR. Kannan. Bayesian networks: Application in safety instrumentation and risk reduction, *ISA Transactions*, *Volume 46, Issue 2*, *April 2007*, *Pages 255-259*

[12] P. Trucco, E. Cagno, F. Ruggeri and O. Grande. A Bayesian belief network modelling of organisational factors in risk analysis: A case study in maritime. *Reliability Engineering & System Safety*, In Press, Accepted Manuscript, *Available online 1 April 2007*

[13] I. Maglogiannis, E. Zafiropoulos, A. Platis and C. Lambrinoudakis Risk analysis of a patient monitoring system using Bayesian Network modeling, *Journal of Biomedical Informatics*, *Volume 39, Issue 6*,*December 2006*, pp6*37-64*