Safety Science 87 (2016) 53–62

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Review

Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review

Guizhen Zhang [a], Vinh V. Thai [a,b,*]

[a] DHI-NTU, Nanyang Environment and Water Research Institute, Interdisciplinary Graduate School, Nanyang Technological University, 1 Clean tech Loop, Singapore 637141, Singapore
[b] School of Business IT & Logistics, RMIT University, 124 La Trobe St, Melbourne, VIC 3000, Australia

ARTICLE INFO

Article history:
Received 17 April 2015
Received in revised form 27 September 2015
Accepted 18 March 2016
Available online 24 March 2016

Keywords:
Bayesian Network
Shipping accidents
Risk modeling
Experts' elicitation
Uncertainty

ABSTRACT

The Bayesian Network (BBN) has been a popular method for risk assessment especially for the modeling of rare accidents. It could make use of experts' domain knowledge when historical data were not enough to support the use of other statistical methods. In the maritime domain, the Bayesian Network has been widely used for risk prediction by modeling the causal relationship of shipping accidents where a lot of human and organizational factors are involved. Most of the models depend on experts' elicitation for model construction and parameterization. The involvement of experts' judgment brings uncertainty and biases. In contrast, data-driven BBN is considered more objective since it is learnt from empirical data. However, even though researchers started to explore the application of data-driven BBN in recent years, its application is still constrained due to the rare occurrence of maritime accidents and the incompatibility of accident databases. As a result, experts' knowledge continues to be an important source for modeling. Reducing the elicitation workload and facilitating the elicitation of individual conditional probability are the two most important tasks for BBN modeling with experts' knowledge. Different techniques that facilitate experts' elicitation process were reviewed in this paper. Some of these methods have been applied in the maritime risk model while new techniques should be developed and applied as well to address the uncertainty and improve accuracy of modeling shipping accidents.

Contents

1. Introduction . . . 54
2. Introduction to Bayesian Belief Network . . . 54
3. Maritime accidents and Bayesian Network . . . 56
   3.1. Maritime accidents and human and organizational factors . . . 56
   3.2. Bayesian Network as a method for maritime modeling . . . 56
   3.3. Experts involvement in Bayesian Network application for maritime modeling . . . 56
4. Challenges with expert elicitation in BBN application . . . 57
   4.1. Reducing the number of conditional probability entries . . . 57
   4.2. Individual conditional probability elicitation . . . 57
5. BBN and experts' elicitation in maritime risk modeling . . . 58
6. Discussion about the literature . . . 60
7. Conclusions and future work . . . 60
   Acknowledgements . . . 61
   References . . . 61

* Corresponding author at: School of Business IT & Logistics, RMIT University, 124 La Trobe St, Melbourne, VIC 3000, Australia.
E-mail addresses: GZHANG004@e.ntu.edu.sg (G. Zhang), vinh.thai@rmit.edu.au (V.V. Thai).

http://dx.doi.org/10.1016/j.ssci.2016.03.019
0925-7535/

1. Introduction

The nature of most shipping accidents (collisions, groundings, fire and explosions, etc.) is low probability–high consequence accidents (Elliott et al., 2008). Catastrophic shipping accidents may cause great loss to the economy, human lives and the environment. One of the primary concerns in the maritime domain is to improve safety and reduce pollution caused by shipping accidents. Statistics show that navigational accidents such as collision, contact and grounding are the most predominant types of shipping accidents (Kuehmayer, 2008; Kujala et al., 2009). Since it is impossible to totally eliminate shipping accidents, a reasonable target is to mitigate accidents in terms of decreasing the probability of their occurrence and minimizing the severity of the associated consequences, i.e. minimizing the risk of accidents.

A risk is defined through event A, consequence C and probability P, namely, Risk $\sim$ (A, P, C) (Goerlandt and Kujala, 2014). Risk assessment is a structured science-based process to estimate the likelihood and severity of risks with attendant uncertainty (Coleman and Marks, 1999). It mainly deals with questions like 'what could go wrong', 'what could be the potential consequences' and 'what is the probability of the occurrence' (Ayyub et al., 2002). Risk assessment has been a really hot topic in the last twenty years in the maritime and offshore industry. The International Maritime Organisation (IMO) proposed the systematic Formal Safety Assessment (FSA) for risk assessment to help decision making in safety management (Kontovas and Psaraftis, 2009). Specifically, there are five steps to conduct a standard FSA study: 1. Hazard identification, 2. Risk assessment, 3. Risk control options, 4. Cost-benefit assessment, 5. Recommendations for decision making. Many methods have been applied for risk analysis the past few years, including Hazard and Operability Studies (HAZOP), Failure Mode and Effects Analysis (FMEA), Fault Tree Analysis (FTA), Event Tree Analysis (ETA) and Bayesian Network (BBN), etc. BBN is becoming more and more popular for maritime risk modeling during the last decade as the number of publications containing the keywords of "maritime safety" and "Bayesian network" has increased from 0 to 16 during the period of 2004–2013 (Hänninen, 2014). This paper therefore aims to review the cases in which the Bayesian Network is applied as a tool for risk assessment for the maritime transportation system, with a focus on the data source and detailed methodology on how to obtain the probability numbers. It was found that experts' knowledge plays an important role in both the establishment of Bayesian Network structures and defining the relative probabilities. Using experts' knowledge to assign probabilities is called expert elicitation (Knochenhauer et al., 2013). Through a review of experts' knowledge elicitation methods and those applied in maritime cases, some useful conclusions could be made.

The remaining of this paper is organized as follows. Section 2 is a detailed description of the BBN theory and some of its modeling advantages and challenges. Section 3 addresses the detailed reasons why BBN is most suitable for risk modeling of shipping accidents. One of the main reasons is that BBN could be used to model human and organizational factors, which are the main causes of maritime accidents. The importance of experts' knowledge for BBN modeling was also discussed in this section. Section 4 carries on with the discussion on the challenges when using experts' elicitation in BBN modeling and reviews techniques that have been developed in the literature to deal with these challenges. Section 5 reviews the BBN modeling of maritime accidents with a focus on the elicitation techniques. It was found that some of the techniques in Section 4 were applied while more techniques could be developed and used. Section 6 discusses the findings of the review and implications for the current research. Section 7 summarizes the paper and draws some conclusions about future research directions.

2. Introduction to Bayesian Belief Network

Bayesian Belief Network (BBN), often known as Bayesian Network or Bayesian Nets for short, is a directed acyclic graph (DAG) and belongs to the family of graphical models (GMs). The detailed definition and features of BBN will be discussed as follows.

A classical BBN structure is composed of nodes and arcs. The value of the nodes may be discrete or continuous, and the most widely used are the discrete nodes. There are mainly three types of discrete node: Boolean nodes, ordered values and integral values, depending on the number of values they may take. The values of Boolean nodes are binary, being either 'True' or 'False'. The ordered value nodes may take several values. For instance, the node 'Pollution' may take the value of 'high', 'medium', or 'low'. Integral values, in contrast, may take more than a hundred values (Kjrćulff and Madsen, 2013). Arcs represent the influence of one node on another. The nodes connected by an arc are called the parent nodes and child nodes respectively. One child node may have several parent nodes, meaning this node is affected by several factors. Similarly, a parent node could have several child nodes, meaning that this factor may have influences on several other factors. Fig. 1 shows a simple Bayesian Network model.

In this example, there are four nodes, "Cloudy" "Sprinkler" "Rain" and "Wet Grass". All of these nodes are binary nodes or Boolean nodes. From the arcs we can see the Wet Grass may be due to two sources: use of Sprinkler or rain. We can also see that cloudy may influence the use of Sprinkler as well as determining the rain probability. We can also see a lot numbers in the example. They are probability numbers. In fact, there are three types of probabilities data in a BBN: prior probability, conditional probability and posterior probability. Prior probabilities are the probability distribution before taking into consideration of any evidence. Conditional probabilities are the probabilities that reflect the degree of influence of the parent nodes on the child node. For BBNs with discrete nodes, the probabilistic dependence is often represented via a table called a Conditional Probability Table (CPT).

To obtain the CPT, we should first find out the possible combination values of the parent nodes, called an instantiation. For each instantiation, the probability that the child node will take a possible value is the conditional probability. They could be calculated using statistical or computational methods or elicited from domain experts (Ben-Gal et al., 2007). For nodes that do not have parent nodes, CPT reduces to prior probabilities. Posterior probabilities are the probability distribution calculated given the evidence. In example one, the distribution: $P(C = T) = 0.5$; $P(C = F) = 0.5$ are the prior probabilities, meaning that the weather maybe either be cloudy or not at 50% probability. The probability that $P(R = T|C = T) = 0.8$ is a conditional probability. We can see that for the node "Sprinkler" and "Rain", there are four probability numbers in the CPT. For the node "Wet grass", there are eight probability numbers. In fact, the number of CPT entries increase exponentially with the number of parent nodes, and the number of states of the parent nodes (Achumba et al., 2013). For a node with i states and k parent nodes and if each parent node has n states, $i \times n^k$ conditional probability values are required while $(i - 1) \times n^k$ values need to be elicited (Knochenhauer et al., 2013). The demand of a large number of CPTs is one of the biggest problems often criticized of BBN. The sheer number of probabilities would not only lead to heavy elicitation loads but will also cause inconsistency of the judgement (Coutts). Ways to reduce the number of CPTs will be reviewed in Section 4.1.

When new evidence (observation) is obtained, inferences could be made, i.e. posterior probabilities could be calculated. Making inferences is also called probability propagation, conditioning or belief updating. The evidence may take several forms, like specific evidence ($X = x$), negative evidence ($Y \neq y_1$, and virtual evidence or