Reliability Engineering and System Safety 102 (2012) 27–40

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

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

Influences of variables on ship collision probability in a Bayesian belief network model

Maria Hänninen [a,*], Pentti Kujala [b]

[a] Aalto University, Department of Applied Mechanics, Kotka Maritime Research Centre, Heikinkatu 7, FI-48100 Kotka, Finland
[b] Aalto University, Department of Applied Mechanics, P.O. Box 15300, FI-00076 Aalto, Finland

ARTICLE INFO

Article history:
Received 21 March 2011
Received in revised form
3 February 2012
Accepted 12 February 2012
Available online 3 March 2012

Keywords:
Maritime accidents
Bayesian networks
Sensitivity analysis
Mutual information
Causation probability

ABSTRACT

The influences of the variables in a Bayesian belief network model for estimating the role of human factors on ship collision probability in the Gulf of Finland are studied for discovering the variables with the largest influences and for examining the validity of the network. The change in the so-called causation probability is examined while observing each state of the network variables and by utilizing sensitivity and mutual information analyses. Changing course in an encounter situation is the most influential variable in the model, followed by variables such as the Officer of the Watch's action, situation assessment, danger detection, personal condition and incapacitation. The least influential variables are the other distractions on bridge, the bridge view, maintenance routines and the officer's fatigue. In general, the methods are found to agree on the order of the model variables although some disagreements arise due to slightly dissimilar approaches to the concept of variable influence. The relative values and the ranking of variables based on the values are discovered to be more valuable than the actual numerical values themselves. Although the most influential variables seem to be plausible, there are some discrepancies between the indicated influences in the model and literature. Thus, improvements are suggested to the network.

1. Introduction

Estimating the probability of a ship–ship collision is an important step in quantitative collision risk assessment. In the most commonly used approach (see e.g. [1–3]), the number of collisions N within a certain ship traffic area and time is calculated as a product of a number of geometrical collision candidates N_A and a causation probability P_C

$$N = N_A \cdot P_C \tag{1}$$

N_A describes the geometrical number of collisions within the time period, i.e., the number of collisions if no evasive maneuvers were made, and it depends on ship traffic properties in the area. There are various approaches to estimating N_A such as [2–7]. P_C denotes the fraction of geometrical collision candidates failing to avoid the collision. It is affected by technical, environmental, and human factors. In the ship collision probability models found in the literature, human and organizational factors variables have explicitly had an influence only on P_C (e.g. [3,8–11]).

A comprehensive ship collision risk model should describe human error-related accident causation mechanisms, as human and organizational errors have been reported causing a large proportion of maritime traffic accidents (e.g. [12–15]); in reality the role of human and organizational errors might be even larger, as studies such as [13,16] have indicated their underreporting. However, modeling these mechanisms might be challenging as accident occurrence is a result of multiple causes having complex interrelations. Maritime traffic accidents are typically a result of chains of events occurring on many organizational levels while each of these events in an accident chain has normally involved one or more human errors [17].

Given the large number of variables with complicated, partly unknown dependencies, a collision risk model will very likely contain uncertainty. One technique for modeling complicated systems with uncertainty is Bayesian belief networks. A Bayesian network approach has been adopted in several probabilistic risk assessments such as [18–21]. They have been applied to modeling human and organizational factors in other domains (e.g. [22–25]) and in the maritime transportation [26,27]—and, also to the causation probability estimation (e.g. [9–11]). However, these existing Bayesian network causation probability models have varied both in the number and the nature of their variables. The causation probability estimates resulting from these models have been compared with previously published values in the literature, but there were no detailed analyses of the model variables and their influences on the causation probability value. While some

* Corresponding author. Tel.: + 358 50 410 6518; fax: + 358 5 225 0570.
E-mail address: maria.hanninen@aalto.fi (M. Hänninen).

0951-8320/$ - see front matter
doi:10.1016/j.ress.2012.02.008

authors including [28] and [29] have found Bayesian network models in general to be rather insensitive to changes in the network probability parameters, more recent studies such as [30–32] have showed that occasionally this does not apply.

The aim of this study is to discover the variables which have the largest influences on the causation probability of ship collisions in the Gulf of Finland modeled with a Bayesian network. Further, the validity of the magnitudes of the influences is discussed, i.e, if the model is realistic. In order to examine the influences thoroughly, three different methods are utilized. The rest of the paper is organized as follows. The analyses are carried out for a model described in Section 2. Section 3 presents the methods applied for the analyses. The results are presented in Section 4, followed by the discussion in Section 5. Finally, conclusions are drawn and recommendations for further study are given in Section 6.

2. Analyzed model

The analyzed model is a Bayesian network for estimating the ship collision causation probability in the Gulf of Finland. Formally, the structure of a Bayesian network model is a directed acyclic graph (see e.g. [33]) consisting of nodes and arcs, representing discrete random variables and the dependencies between the variables, respectively [34]. Each variable consists of a finite set of mutually exclusive states. For each variable A with parent nodes $B_1,\ldots,B_n$, i.e., there is an arc from $B_1,\ldots,B_n$ to A, there exists a conditional probability table $P(A|B_1,\ldots,B_n)$ that describes the probabilities of all combinations of the states of A and B; if variable A has no parents, it is linked to unconditional probability table $P(A)$. The entries of the probability tables are the parameters of the Bayesian network model.

The structure of the analyzed model is to a large extent based on models presented by Det Norske Veritas (DNV) [10,35]. However, the model considers multiple ship types and models both of the encountering vessels. A simplification of the model structure can be seen in Fig. 1. A complete list of the model variables and their parents is presented in Appendix A, Table A1.

The data sources for the network probability parameters are listed in Table A1. The majority of the parameters is derived from

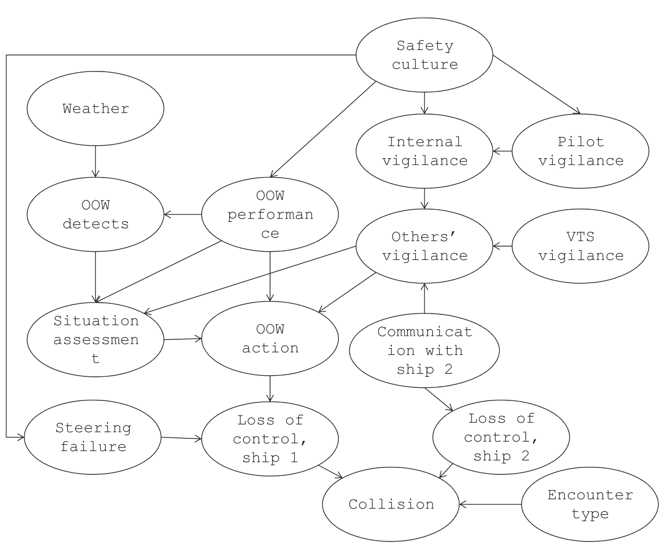

Fig. 1. A simplification of the model structure. The actual model has 100 variables and includes variables affecting the loss of control for both ships although no variables of ship 2 are visible in thisfigure. A complete list of the model variables and their parent nodes (excluding the ship 2 variables that are identical to the ones of ship 1) is presented in Table A1.

the DNV models (all zeros are substituted with 10^{-6} so that sensitivity analysis is technically possible for those parameters as well). Alternatively, the parameters of certain network variables are based on or adjusted with data from the Gulf of Finland or from Finnish officers, or from experts' judgment. These parameters, the details of their sources and the expert judgments are presented in the contents and captions of Fig. A1 and Tables A2–A13 of Appendix A. For the parameter values derived from the DNV models, see [10,35]. The default causation probability value the model estimates, that is, the value with no observation of the network variables, is 1.06E−04.

3. Applied methods

3.1. Difference in collision probability given observations on another variable's true state

The value of the causation probability is examined while observing each state of each model variable in turn. In other words, the probability of the observed state is set to one and the probabilities of the other states of that variable to zero. The difference in the causation probability for a state producing the largest causation probability and a state corresponding to the smallest probability is recorded for each variable Y

$$\Delta P_{C_Y} = \max_j p(\text{“Collision”} = \text{“yes”}|Y = y_j) - \min_k p(\text{“Collision”} = \text{“yes”}|Y = y_k) \quad (2)$$

where y_j and y_k the states producing the largest and smallest causation probabilities of variable Y, respectively. This difference describes the maximum change in causation probability a variable could produce. Bayesian network software Hugin Expert [36] is used in performing the analysis. It should be noted that in case of a pair of variables describing the same factor for both ships (see Table A1, the variables marked with an asterisk), only one of these variables is observed.

3.2. Sensitivity value

Causation probability's sensitivity to changes in all other network probability parameters is analyzed using one-way sensitivity analysis. In a one-way sensitivity analysis, every conditional and prior probability in the network is systematically varied in its turn while keeping the others unchanged and the effects on the output probabilities computed from the network are then examined [37].

One-way sensitivity analysis is conducted using a sensitivity value approach presented in [37,38]. A sensitivity function is established which describes the causation probability P_C as a function of the parameter $z = p(Y = y_i|\pi)$, where y_i is one state of a network variable Y, and π is a combination of states for Y's parent nodes. For a network with no observations on any of the network variables and when applying proportional covariation [38] for keeping the sum of Y's parameters equal to one, sensitivity function is linear

$$P_C(z) = c_1 z + c_2 \quad (3)$$

where the constants c_1 and c_2 are identified based on the model. The first derivative of the sensitivity function at the original value, called sensitivity value, describes the effects of minor changes in original parameter value on the output

$$P'_C(z) = c_1 \quad (4)$$

In case of no observations, the sensitivity value also describes the effects of major changes in z: it equals the change on the output probability if the parameter was changed from zero to one.