Safety Science 49 (2011) 1142–1155

Contents lists available at ScienceDirect

Safety Science

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

Experts, Bayesian Belief Networks, rare events and aviation risk estimates

Peter Brooker *

Aviation Consultant, 37 Woodland Way, West Wickham, Kent BR4 9LR, United Kingdom

ARTICLE INFO

Article history:
Received 5 September 2010
Received in revised form 28 January 2011
Accepted 25 March 2011
Available online 6 May 2011

Keywords:
Aviation safety
Bayesian belief nets
Causal models
Formal safety assessment
Risk analysis

ABSTRACT

Bayesian Belief Networks (BBN) are conceptually sensible models for aviation risk assessment. The aim here is to examine the ability of BBN-based techniques to make accurate aviation risk predictions. BBNs consist of a framework of causal factors linked by conditional probabilities. BBN conditional probabilities are elicited from aviation experts. The issue is that experts are not being asked about their expertise but about others' failure rates. A simple model of expertise, which incorporates the main features proposed by researchers, implies that a best-expert's estimates of failure rates are based on accessible quantitative data on accidents, incidents, etc. Best-expert estimates will use the best available and accessible data. Depending on the frequency of occurrence, this will be data on similar events, on similar types of event, or general mental rules about event frequencies. These considerations, plus the need to be cautious about statistical fluctuations, limit the accuracy of conditional probability estimates. The BBN framework assumes what is known as the Causal Markov Condition. In the present context, this assumes that there are no hidden common causes for sequences of failure events. Examples are given from safety regulation comparisons and serious accident investigations to indicate that common causes may be frequent occurrences in aviation. This is because some States/airlines have safety cultures that do not meet 'best practice'. BBN accuracy might be improved by using data from controlled experiments. Aviation risk assessment is now very difficult, so further work on resilience engineering could be a better way of achieving safety improvements.

1. Introduction

In recent decades, perhaps every possible means of preventing air transport accidents and reducing the risks to aircraft passengers has been investigated. The safety improvement process continues, and is a key part of the work of operational and research staff throughout the world. Risk assessment is a key part of this process. There are variety of ways of understanding risk factors in combination, including fault trees, event trees, and decision trees. A comparative newcomer in aviation risk analysis is the use of Bayesian Belief Networks (BBN). There are also slightly different names for BBNs, e.g. Bayesian Probability Nets.

The aim here is to examine the ability of BBN-based techniques to make accurate aviation risk predictions. This is usually categorised as *Validation*: "a demonstration that a predictive model within its domain of applicability possesses a satisfactory range of accuracy consistent with its intended application". This is distinct from *Verification*, which is a demonstration that the modelling formalisms – calculations, inputs, computer code, etc. – are correct: see Rykiel (1996) for a discussion on definitions. BBN risk assessments provide insight on the connections between risk factors, but how reliable are their quantitative estimates? Is there any danger that an estimate would 'prove' that part of the aviation operation is acceptably safe when in truth it is not? Would its predictions lead to an inefficient safety focus? Could BBN results wrongly indicate that a combination of factors of some type A generated more risks than a combination of a type B?

BBNs use experts to estimate the probabilities of events. The probabilities are conditional, i.e. the chance of something happening given that something else has happened. Probabilities are then combined to model the probabilistic behaviour of the system in question. The next section explains this further. The main task here is to try to answer the question: "Can an aviation expert estimate rare event conditional probabilities accurately?" Most of the analysis is about understanding what the words in the question might mean.

Several respected authors and research groups have studied the use of BBNs in aviation safety, and there is considerable commonality in their methodologies. Illustrative examples of three of the groups' work, not in a priority order, are:

> Neil et al. (2003): "…Bayesian Belief Network for an Air Traffic Control environment…a high level model of ATC operations spanning a number of defensive barriers from airspace design, through tactical control, the operation of aircraft safety net features to a potential accident."

* Tel.: +44 208 777 1718.
E-mail address: p_brooker@btopenworld.com

0925-7535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2011.03.006

> Luxhoj and Coit (2006): "In the aviation industry, accidents occur very infrequently, yet it is still critical to further reduce their rate of occurrence. Existing methods and models are already useful, but because of the importance of these failures, new modeling perspectives can add additional insights to further enhance safety…a model devoted to this class of 'low probability-high consequence' events…demonstrated with a model developed for a certain aircraft accident type known as Controlled Flight Into Terrain (CFIT)."
>
> Ale et al. (2009): "… a research effort has started to develop causal models for air traffic risks in the expectation that these will ultimately give the insight needed…In this paper, the backbone of the model and the way event sequence diagrams, fault-trees and Bayesian belief nets are linked to form a homogeneous mathematical model suitable as a tool to analyse causal chains and quantify risks…"

The examples used here are from Ale et al. because it is very comprehensive and is supported by extensive published work in peer-reviewed journals.

Chapter 8 of Ale et al. (2009) is devoted to Validation in general. It focuses on Verification aspects, but is unambiguous about validation problems in their studies:

> "Validation against an independent data set was unfortunately impossible, because all available data were used to populate the model with numbers and calibrate the model on the historical accident rates.
>
> Case validity: Check whether the model gives answers for a specific case which correspond to reality … The application of the model on specific case studies has been postponed to a later project."

The most recent published journal articles – Groth et al. (2010) and Roelen et al. (2011) – into wider aspects of BBN modelling do not offer much progress on this aspect. Groth et al. has a section called Validation but it is actually about Verification. Roelen et al. does not examine Validation explicitly.

2. Bayesian Belief Networks: a sketch

There are very many sources on BBNs. Two classic books are Pearl (2009), who originated many of the ideas in the field, and Spirtes et al. (2000). A more recent book with relevance to some of the issues examined here is Williamson (2005). A recent review of applications on dependability, risk analysis and maintenance is Medina-Oliva et al. (2009). For present purposes, it is necessary to explain some of the basic vocabulary of BBNs. From Faber (2001), a BBN's elements are:

> Group of nodes – parents and descendants is one jargon – and a group of directed arcs,
>
> Table of prior probabilities $P(Z)$ associated to the top-level parent node Z,
>
> Table of conditional probabilities $P(Y|X)$ attached to the node Y whose parent is X, which are the probability distributions over the states of Y given the states X.

Fig. 1 is a sketch of a BBN for a power supply system of an engine, main fuel supply to the engine, and backup fuel supply with limited capacity. Electric cables deliver power to the consumers. There will be no supply to the consumer if the power supply is cut off. This happens if both the main and backup fuel supplies fail and/or the supply cables fail. The boxes show the unconditional probabilities for the top-parent events and the conditional probabilities for the descendent events. Working through the probabilities produces the probability structure for the different states of the system – Fig. 2. Thus, BBNs can simply generate fault trees and event trees: see Faber for further details.

The additional ingredient in BBNs is 'Belief'. In many cases, the probabilities in the boxes derive from 'expert' judgements. These are 'subjective probabilities', rather than measured probabilities (i.e. relative frequencies) of failure events. *Elicitation* is the process of deriving these probabilities (and probability *distributions*) from experts. It converts the thoughts in an individual's mind into quantitative statements about his/her beliefs. Good elicitation means that the probability statements accurately represent the expert's (imperfect) knowledge. A very good general reference is Garthwaite et al. (2005).

The psychological literature reports that in some circumstances people's judgements are error-prone, because they use specific heuristics, and are subject to biases in how they respond to situations involving uncertainty. Tversky and Kahneman (1974) were the first to examine these effects comprehensively. There is now a large literature on the topic. Kynn (2008) presents a recent review, which inter alia argues that substantive experts are much less subject to such errors and biases.

An obvious question is: "What do we mean by an expert?" Then there is a range of more focused questions relevant to the present context. "How do experts become expert?" "What are the characteristics of expert analysis?" "How do we know how good an expert is?" "What limits the accuracy of an expert's assessment?" These kinds of questions are explored in the following sections. The literature on expertise is huge, so the text here is brief and highly simplified.

3. What is meant by 'an expert'?

Historically, expertise arose in pre-scientific technology, e.g. in smelting before metallurgical chemistry and controlled experiments. Specialist skills passed from one generation to the next. Dictionary definitions of an expert are typically: 'someone with great skill in, or extensive knowledge of, a particular subject.' In the following, expert specifically means possessing some kind of technical expertise, not simply stakeholders in a particular field. A few examples indicate the kinds of characteristics in the development of – and performance by – such experts. A nice background paper is Ericsson et al. (2007), and a useful general review is Farrington-Darby and Wilson (2006).

Probably the most studied area of expertise is the game of chess. This is of interest in its own right but also as a 'finite world' means of understanding the nature of intelligence. It is often studied to aid the development of artificial intelligence methodologies (Gobet and Chassy, 2009). Grandmaster performances are hugely different from both novices and experienced leisure players. Typically, experts do not have a higher general intelligence or a special talent, i.e. their expertise is largely acquired through deliberate practice. It typically takes 6–10 years to become an expert. Specific skills are in pattern recognition and search techniques: experts have very specialized knowledge, partly coded as *perceptual chunks*. The rapid understanding shown by experts when they face a problem and carry out these highly focused searches is termed 'intuition'. When necessary, chess experts can search the consequences of potential moves to much greater depths. Feedback is crucial – experts become more expert by keeping track of their successes and their failures – and their analyses of the reasons for failure.

Of immediate practical relevance is the estimation of software costs (e.g. see Jørgensen and Boehm, 2009). Surveys of software projects suggest that these estimates are quite inaccurate and strongly biased towards over optimism. On average, there is an effort overrun of about 30%. There have been at least four decades of