Reliability Engineering and System Safety 144 (2015) 265–284

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Review

Non-parametric Bayesian networks: Improving theory and reviewing applications

Anca Hanea [a,*], Oswaldo Morales Napoles [b], Dan Ababei [c]

[a] Centre of Excellence for Biosecurity Risk Analysis (CEBRA), The University of Melbourne, VIC 3010, Australia
[b] Hydraulic Engineering, Faculty of Civil Engineering & Geosciences, Delft University of Technology, Netherlands
[c] Light Twist Software, Australia

ARTICLE INFO

Article history:
Received 28 August 2014
Received in revised form
22 May 2015
Accepted 29 July 2015
Available online 14 August 2015

Keywords:
Bayesian networks
Model selection
Quantification
Inference
Structured expert judgement
Risk
Safety
Reliability
Parameter estimation

ABSTRACT

Applications in various domains often lead to high dimensional dependence modelling. A Bayesian network (BN) is a probabilistic graphical model that provides an elegant way of expressing the joint distribution of a large number of interrelated variables. BNs have been successfully used to represent uncertain knowledge in a variety of fields. The majority of applications use discrete BNs, i.e. BNs whose nodes represent discrete variables. Integrating continuous variables in BNs is an area fraught with difficulty. Several methods that handle discrete-continuous BNs have been proposed in the literature. This paper concentrates only on one method called non-parametric BNs (NPBNs). NPBNs were introduced in 2004 and they have been or are currently being used in at least twelve professional applications. This paper provides a short introduction to NPBNs, a couple of theoretical advances, and an overview of applications. The aim of the paper is twofold: one is to present the latest improvements of the theory underlying NPBNs, and the other is to complement the existing overviews of BNs applications with the NPNBs applications. The latter opens the opportunity to discuss some difficulties that applications pose to the theoretical framework and in this way offers some NPBN modelling guidance to practitioners.

Contents

1. Introduction . . . 266
2. Hybrid Bayesian networks . . . 267
3. Non-parametric Bayesian networks . . . 267
   3.1. Quantification/specification . . . 268
      3.1.1. Data . . . 269
      3.1.2. Structured expert judgement . . . 270
   3.2. Inference . . . 270
4. Applications of NPBNs . . . 271
   4.1. Risk analysis applications . . . 271
      4.1.1. Earth dams safety in the State of Mexico . . . 271
      4.1.2. Linking $PM_{2.5}$ concentrations to stationary source emissions . . . 274
      4.1.3. Causal models for air transport safety . . . 275
      4.1.4. The benefit-risk analysis of food consumption . . . 275
      4.1.5. The human damage in building fire . . . 275
   4.2. Reliability of structures . . . 276
      4.2.1. Bayesian network for the weigh in motion system of The Netherlands . . . 276
   4.3. Properties of materials . . . 276
      4.3.1. Technique for probabilistic multi-scale modelling of materials . . . 276
   4.4. Dynamic NPBNs . . . 276
      4.4.1. Permeability field estimation . . . 276

* Corresponding author. Tel.: +61 420 880 503.
E-mail addresses: anca.maria.hanea@gmail.com (A. Hanea), O.MoralesNapoles@tudelft.nl (O. Morales Napoles), dan@lighttwist.net (D. Ababei).

http://dx.doi.org/10.1016/j.ress.2015.07.027
0951-8320/

4.4.2. Traffic prediction in The Netherlands . . . 276
4.5. Summary of completed applications . . . 276
4.6. Ongoing applications . . . 276
4.6.1. Filtration techniques . . . 276
4.6.2. Flood defences . . . 276
4.6.3. Train disruptions . . . 277
5. Conclusions . . . 277
Appendix A. . . . 277
A.1. Proof of Theorem 3.2 . . . 277
A.2. Validation tests – essential properties and formulae . . . 278
Appendix B. Details of applications . . . 278
B.1. CATS . . . 278
B.2. BENERIS . . . 278
B.3. Building fires . . . 279
B.4. WIM . . . 279
B.5. Multi-scale materials modelling . . . 280
B.6. Permeability estimation . . . 282
B.7. Traffic prediction . . . 282
References . . . 283

1. Introduction

Understanding and representing multivariate distributions along with their dependence structure is a highly active area of research. A large body of scientific work treating multivariate models is available. This paper advocates graphical models in general and Bayesian networks (BNs) in particular to represent high dimensional distributions with complex dependence structures.

A BN consists of a directed acyclic graph (DAG) and a set of (conditional) distributions. Each node in the graph corresponds to a random variable and the arcs represent direct qualitative dependence relationships. The absence of arcs guarantees a set of (conditional) independence facts. The direct predecessors (successors) of a node are called parents (children). A marginal distribution is specified for each node with no parents, and a conditional distribution is associated with each child node. The (conditional) distributions serve as the quantitative information about the strength of the dependencies between the variables involved. The graph with the conditional independence statements encoded by it, together with the (conditional) distributions represent the joint distribution over the random variables denoted by the nodes of the graph. The quantitative information needed can be either retrieved from data, when available, or from experts.

The relatively simple visualization of the complicated relationships among the random variables is one of the most appealing features of a BN model. The main use of BNs is to update distributions given observations. This is referred to as inference in BNs.

BNs have been successfully used to represent uncertain knowledge, in a consistent probabilistic manner, in a variety of fields [62]. Nevertheless, most of the applications use discrete BNs, i.e BNs whose nodes represent discrete random variables. These models specify marginal distributions for the nodes with no parents, and conditional probability tables for child nodes. Inference in discrete BNs can be performed using either exact algorithms (e.g. [68,51]), or approximate algorithms (e.g. [31,47]). Despite their popularity, discrete BNs suffer from severe limitations. Applications involving high complexity in data-sparse environments are severely limited by the excessive assessment burden which leads to rapid, informal and indefensible quantification. Furthermore, this type of representation, using only discrete variables, is inadequate for many problems of practical importance. Discrete BNs have been extensively studied, hence we omit their presentation here, but refer to Weber et al. [62] and references therein. Many domains require reasoning about the joint behaviour of a mixture of discrete and continuous variables. These domains are often called hybrid domains. Hence, BNs involving both discrete and continuous variables will be called hybrid BNs (HBNs). Working with HBNs proves considerably more challenging than working with their discrete counterparts. Several methods for HBNs are available and discussed in the literature (e.g. [38,34]). We will only briefly mention some of them, insist on the most recent ones, and present in more detail the method called non-parametric BNs (NPBNs) that is overlooked in many HBNs review papers. The emphasis here is on the application of the NPBN methods in different fields of science and engineering in the course of approximately 10 years since NPBNs were first presented.

A study which is perhaps the most comprehensive overview study of BNs applications [62] incorporated information about most methods available for working with HBNs up to the date of its publication. Given the amount of research available on the theoretical and practical aspects of BNs, it is not surprising that even such an impressively large study overlooked a couple of methods and their applications. For instance only one application of NPBNs was mentioned, without any specification of the methodology applied. We would like to complement their presentation of NPBN applications.

We aim to provide a short introduction to the NPBNs, present a couple of theoretical refinements and an overview of their use in practice. Moreover a number of difficulties and challenges occur when working with HBNs in general, and with NPBNs in particular. We address these challenges from the practitioner's perspective, and in this way we provide guidance in using the NPBN framework.

The remainder of the paper is organized as follows. Section 2 provides some background on methods available for HBNs. Section 3 describes the NPBN model and gives details about quantification and inference using NPBNs. The main theorem of NPBNs is reformulated and proved for more general settings. Two NPBN structure learning algorithms are detailed. The first one represents the initially proposed algorithm used in most of the applications presented later in the paper, whereas the second one is a new proposal that shows promise. NPBNs have been or are currently being used in at least twelve professional applications. Section 4 reviews these applications. Two of the applications are described in more detail with the intention of presenting some more theoretical details on real examples. The other applications are only very briefly presented,