Advances in Engineering Software 115 (2018) 126–148

Contents lists available at ScienceDirect

Advances in Engineering Software

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

Research paper

An open toolbox for the reduction, inference computation and sensitivity analysis of Credal Networks

Silvia Tolo[a], Edoardo Patelli*,[a], Michael Beer[a,b,c]

[a] Institute for Risk and Uncertainty, University of Liverpool, UK
[b] Institute for Risk and Reliability, Leibniz University Hannover, Germany
[c] Shanghai Institute of Disaster Prevention and Relief, Tongji University, China

ARTICLE INFO

Keywords:
Credal Networks
Bayesian Networks
Decision making
System reliability

ABSTRACT

Bayesian Networks are a flexible and intuitive tool associated with a robust mathematical background. They have attracted increasing interest in a large variety of applications in different fields. In spite of this, inference in traditional Bayesian Networks is generally limited to only discrete variables or to probabilistic distributions (adopting approximate inference algorithms) that cannot fully capture the epistemic imprecision of the data available. In order to overcome these limitations, Credal Networks have been proposed to integrate Bayesian Networks with imprecise probabilities which, adopting non-probabilistic or hybrid models, allow to fully represent the information available and its uncertainty.

Here, a novel computational tool, implemented in the general purpose software OpenCossan, is proposed. The tool provides the reduction of Credal Networks through the use of structural reliability methods, in order to limit the cost associated with the inference computation without impoverishing the quality of the information initially introduced. Novel algorithms for the inference computation of networks involving probability bounds are provided. In addition, a novel sensitivity approach is proposed and implemented into the Toolbox in order to identify the maximum tolerable uncertainty associated with the inputs.

1. Introduction

The fast technological growth that has characterized the last century has progressively provided more efficient and advanced instruments for everyday life as well as for industrial and scientific applications. This progress goes along with an ever increasing grade of complexity which concerns the engineering field on any level, leading to face new and more challenging tasks from both mathematical and computational points of view. As a consequence of this, novel and efficient tools are strongly needed to adequately predict the behaviour of complex systems, optimize their performance, estimate their reliability and evaluate the risks they are subject to, especially with regards to safety-critical installations (e.g., reservoirs, dams, nuclear and chemical installations etc.).

The accuracy in estimating the actual risks to which a system is subject is clearly bound by our ability of capturing and representing reality. This means that any engineering analysis has to face the challenging task of formulating suitable numerical models in a quantitative manner without ignoring significant information on the one hand, and without introducing unwarranted assumptions on the other [5]. If this balance is violated, computational results may deviate significantly from reality and the associated decisions may lead to serious consequences due to risk underestimation or, conversely, to unnecessary costs in the case of over-conservative approach.

This challenging task implies two main bottlenecks: the first one is associated with the technological complexity of the systems under study. This includes the representation of elaborate networks of dependencies among subsystems and components interacting directly or indirectly with each other and determining the correct functioning of the overall system. To provide an oversimplified representation of these mechanisms, hence to introduce large model errors, can easily result in misleading estimates of the system state or its reliability. The criticality of this task is also worsened by the difficulty of adequately quantifying the significance of the model error introduced and hence the weight of the simplifications adopted in the analysis. For this reason, it is of extreme importance to be able to adequately reproduce these interconnections in order to take into account the possible failure of each individual component and its consequences on a more or less wide range of others, potentially triggering a chain of events and leading to more serious accidents scenarios or failure modes.

* Corresponding author.
E-mail addresses: s.tolo@liverpool.ac.uk (S. Tolo), edoardo.patelli@liverpool.ac.uk (E. Patelli).

http://dx.doi.org/10.1016/j.advengsoft.2017.09.003
Received 19 August 2016; Received in revised form 27 July 2017; Accepted 14 September 2017
Available online 22 September 2017
0965-9978/

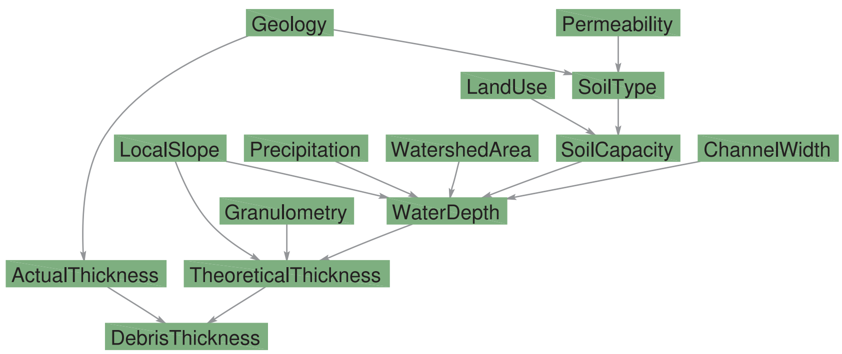

Fig. 1. CN model for hazard assessment of debris flows proposed by Antonucci et al.

Several methods (e.g., Fault Tree, Event Trees, Reliability Block Diagrams etc.) are available in the scientific literature and widely used in several fields of engineering for the representation and analysis of complex systems. Among these, Bayesian Networks (BNs) have attracted an increasing interest in the last decades spreading to several industrial and scientific application fields. The large success of BNs is linked to their capability of providing efficient factorization of joint probability distributions exploiting information about the conditional dependencies existing among the variables involved on the basis of a robust mathematical background such as Bayesian Statistics. In addition, their intuitive graphical framework has consolidated their attractiveness in quite different fields of science and engineering, from artificial intelligence to medical and economic areas [42]. They can be considered as the general case of more common methodologies, such as Fault Tree analysis [18], with respect to which they offer several advantages. Indeed, Bayesian Networks can model complex dependencies among components, uncertainty can be included in modelling and both forward and backward analysis is allowed, making them particularly attractive for both diagnosis and inference purposes [7].

Often, it is necessity to deal with our ignorance, in terms of uncertainty, scarcity of data or even a more general lack of knowledge regarding the actual mechanisms of interaction involved in the analysis. This must be explicitly included in the analysis in order to be able to quantify its impact on the accuracy and robustness of the results obtained. Nevertheless, information available in real-world application involves sparse data, poor measurements and subjective information, hence results difficult to quantify and model. The adoption of traditional mathematical models built on poor or scarce data, can lead the information modelled to be far from that actually available, introducing biases that lower the credibility and can even invalidate the results of the analysis [5]. For instance, this can occur when the samples related to certain event are not enough, or not robust enough, to be clearly attributed to a particular distribution family or to be associated confidently with precise parameter values for the distribution selected. A wide range of solutions are available to adequately represent different degrees of information without introducing biases or strong assumptions but reflecting in the predictions the uncertainty in input. The definition of imprecise probabilities embraces a wide set of approaches (e.g., interval probability, fuzzy probabilities, p-boxes) which provide a mathematical basis to deal with the representation of information when it is not sufficient for probabilistic modelling or rather suggests a set-theoretical approach.

The aim of this work is to provide a novel computational tool for the efficient computation of Credal Networks (CNs). The approach proposed is based on the use of system reliability methods to integrate traditional Bayesian Networks with probabilistic, non-probabilistic and hybrid frameworks without renouncing to the robustness of traditional inference algorithms. The proposed algorithms have been implemented in the general purpose software OpenCossan [17,29,30]. They allow to reduce the initial user defined models into a simpler but equivalent network. This, conversely from the initial model which can include a wide variety of variables type, embraces only crisp and interval probabilities and hence results into a BN containing probability bounds. The integration of these two approaches (namely BNs and probability intervals), strongly enhances the robustness of the analysis, but also introduces significant challenges. First, the inference computation on these models can easily become highly demanding. Second, the capability to track the propagation of uncertainty within the model is essential in order to ensure the significance of the output and to obtained the desired level of accuracy of the analysis at the lowest cost. This study aims to analyse and propose novel solutions to deal with these issues. Few available software packages for manipulation of graphical models with imprecise probabilities are available, so that this field is highly open to new contributions [11] and this study is characterized by strong novelty.

1.1. Aim and motivation

The aim of the proposed methodology and related computational tool is to offer a novel approach for the implementation of credal networks able to include different mathematical frameworks for the representation of the available data. In spite of the restricted number of CNs applications available in the literature, the limitations of such approach in terms of data representation are blatant: most of the inference algorithms and software available for the computation of Credal Networks are indeed restricted to random variables which assume finitely many values (also called discrete or categorical variables). In many cases, this implies the necessity to adapt the representation of data to a discrete number of possible outcome states regardless the nature of the information available and hence potentially impoverishing its quality and the accuracy of the overall analysis.

As an example, Antonucci et al. [4] propose a CN model for the hazard assessment of debris flows, as shown in Fig. 1. The model, developed for the Ticino canton in Switzerland, allows to fully exploit the advantages offered by CNs, such the potential of credibly dealing with qualitative uncertainty through the use of imprecise probabilities, enhancing the robustness of the inference computation. On the other hand, the authors highlight computational issues related to the updating process which could require different and more expensive solutions in the case of application of the model to other geographical areas. For computational reasons the model is restricted to discrete nodes, in spite of the large dataset available and the random nature of the phenomenoa modelled (e.g. precipitation intensity, granulometry). Conversely, the treatment of continuous variables in the network would be more appropriate, avoiding the loss of information associated with discretization. The computational approach implemented in the present study offers a solution strategy for both these issues: on the one hand it allows to depict the data available through the use of continuous variable, representing the randomness of the phenomenon involved as well as the imprecision affecting the available knowledge. On the other hand, it allows to simplify the initial network, reducing significantly the computational costs of updating, even for identical inference