

From complex questionnaire and interviewing data to intelligent Bayesian Network models

Medical data is very often badly structured, incomplete and inconsistent. This limits our ability to generate useful models for prediction and decision support if we rely purely on machine learning techniques. That means we need to exploit expert knowledge at various model development stages. This paper, published in *Artificial Intelligence in Medicine*, tackles this problem which is common in many application domains. The paper describes a rigorous and repeatable method for building effective Bayesian Network (BN) models from complex data.

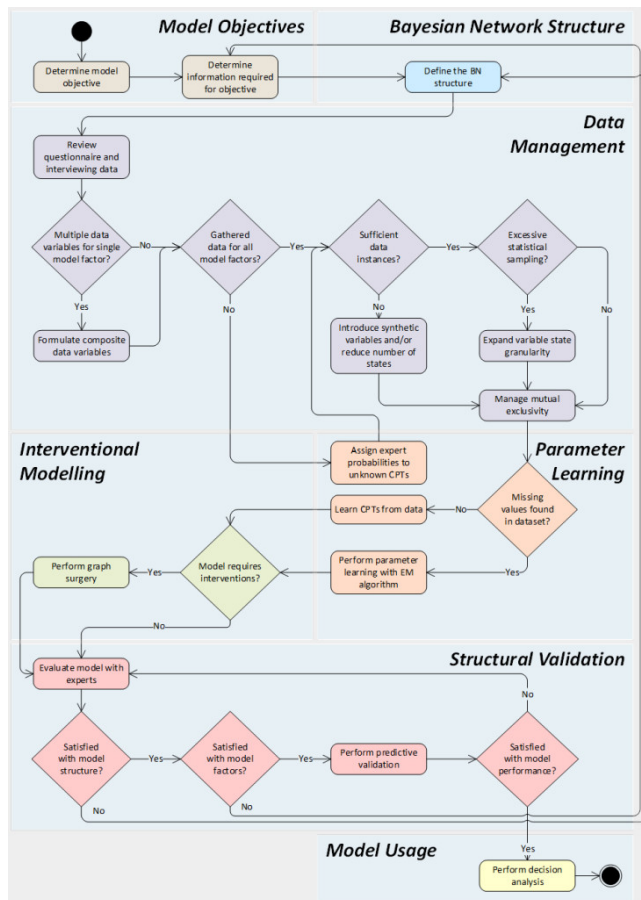


Fig. 1. The proposed expert Bayesian network development process based on questionnaire and interviewing data.

What is a BN? It is a type of statistical model and which represents the probabilistic relationships between factors of interest. BNs are also known as probabilistic graphical models, since they consists of nodes which represent the factors of interest, and arcs which usually represent the

direction of influence. Underpinning BNs is Bayesian probability inference that provides a way for rational real-world reasoning. Any belief about uncertainty of some event is assumed to be provisional upon experience or data gained to date, and which is then revised by new experience or data.

What do we mean by complex data? In the paper, complex data refers primarily to the data that comes from poorly structured questionnaires and interviews. This is data that was never designed to obtain information suitable for causal modelling. It involves answers to hundreds of questions, but including inevitably examples of repetitive, redundant and contradictory responses.

In the absence of expert knowledge, learning a BN model from such data alone is especially problematic where we are interested in simulating causal interventions for risk management. The novelty of this work is that it provides a rigorous consolidated and generalised framework that addresses the whole life-cycle of BN model development. The development process is demonstrated in Figure 1.

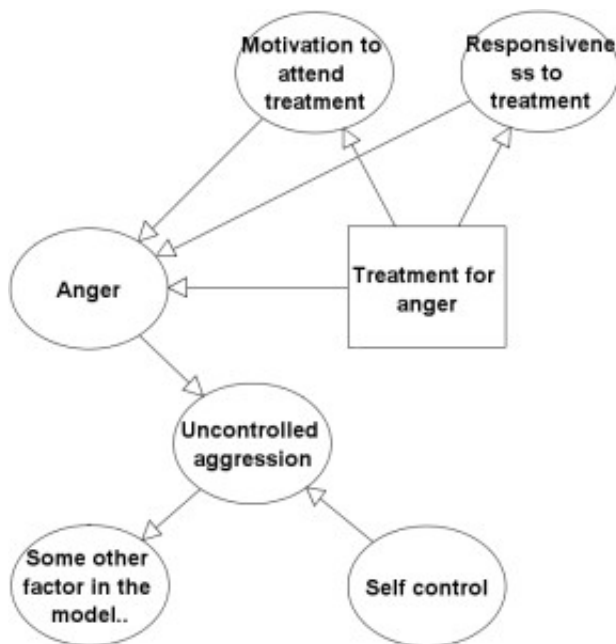


Fig. 2. An example of a risk management process available to the decision maker via simulation of intervention/s (squared node).

The method is validated using data from forensic psychiatry. The resulting BN models demonstrate competitive to superior predictive performance against the data-driven state-of-the-art models employed within this area of research. More importantly, the resulting BN models go beyond improving predictive accuracy and into usefulness for risk management through intervention, and

enhanced decision support in terms of answering complex clinical questions that are based on unobserved evidence.

Fig. 2 demonstrates a risk management example where, in the BN, the decision maker has the option to simulate the impact of an intervention (represented by the squared node). More specifically, by enabling the treatment the decision maker can observe the impact the intervention is expected to have on factor 'Anger', taking into account other relevant factors such as motivation to attend and responsiveness to treatment. The manipulation of anger will subsequently affect 'Uncontrolled aggression' and through that, some other factors of interest incorporated into the model.

The method is applicable to any application domain involving large-scale decision analysis based on such complex and unstructured information. It challenges decision scientists to reason about building models based on what information is really required for inference, rather than based on what data is available. Hence, it forces decision scientists to use available data in a much smarter way.

Anthony Constantinou, Norman Fenton, William Marsh and Lukasz Radlinski

Publication

[From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support.](#)

Constantinou AC, Fenton N, Marsh W, Radlinski L.
Artif Intell Med. 2016 Feb