Paper 4

Combining judgments with messy data to build Bayesian Network models for improved intelligence analysis and decision support

Norman Fenton[1]*, Anthony Constantinou[1], Martin Neil[1]

[1]Queen Mary University of London and Agena Ltd

*Corresponding author: n.fenton@qmul.ac.uk

Short abstract

Bayesian Network (BN) models are an increasingly important decision tool for critical real-world applications. While much research has focused on algorithms for learning the structure and parameters of BN models from data, such 'machine learnt' models require enormous volumes of data to discover the cause-and-effect, or other, relationships of interest. However, in critical applications like intelligence analysis, relevant data is often limited relative to the complexity of the problem. Even when the data are 'big', they tend to be 'messy': poorly structured and do not adhere to causal representations required to construct a BN model. In such situations we have to rely on a combination of knowledge and data to construct a useful BN that provides meaningful decision support for risk management rather than just 'prediction'. In particular, a useful BN in this context must be able to support 'interventional analysis', which requires knowledge about the 'actions' available to the decision maker. Such information typically needs to be hard coded into the model. Unfortunately, there are limited methods for building useful BNs in such scenarios.

We describe a rigorous and repeatable method for building effective BNs using a combination of knowledge and data. While much of the method is based on established work, it is novel in that it provides a rigorous consolidated and generalised framework that addresses the whole life-cycle of BN model development. The BN development process is applicable to any real-world application domain and challenges decision scientists to reason about models based on what information is really required for inference, rather than based on what data is available and hence, encourages decision scientists to use available data in a much smarter way.

Long abstract

Bayesian networks (BNs) are a well-established graphical formalism for encoding the conditional probabilistic relationships among uncertain variables of interest. A well-developed BN offers decision makers a range of powerful features, such as both predictive and diagnostic analysis and what-if analysis, including interventional and

even counterfactual reasoning. Hence, BNs have proven useful in a wide range of application domains [2] including recently in security and intelligence analysis [3] [4]. The major challenge is in how to build an accurate BN model that correctly captures the causal or other relationships between factors of interest. While much research has focused on algorithms that learn the structure and parameters of BN models from data, such 'machine learnt' models require enormous volumes of data to perform well. However, such models tend to optimise for accuracy, and prediction alone provides limited usefulness in the context of intelligence decision analysis, where we need models that provide decision support for intervention actions. Moreover, for such problems data are often limited relative to the complexity of the model; even when the data are 'big', they tend to be 'messy': poorly structured, not adhering to causal representations required to construct a meaningful BN model, and often involving repetitive, redundant and contradictory information.

Messy data make discovery of cause-and-effect relationships (where possible) much harder, and generate models that do not support causal intervention decision making. To overcome this, we have to rely on a combination of knowledge and data. Interventional analysis requires knowledge about the 'actions' available to the decision maker, and such information typically needs to be hard coded into the model. Yet, there are limited methods for building useful BNs in such scenarios. This is a primary reason why, despite their obvious demonstrable benefits, BNs still remain under-exploited in areas where they offer the greatest potential.

We describe a rigorous and repeatable method for building effective BNs using a combination of knowledge and data. The method has been validated on two applications in forensic psychiatry [1]. Most of the components of the method are based on established work on data and knowledge elicitation, parameter learning and interventional analysis for risk management. This includes the use of idioms to define the model structure and data management techniques to ensure data adhere to causal, or other, representation required by the BN model. The novelty of the method is that it provides a rigorous consolidated and generalised framework that addresses the whole life-cycle of BN model development.

The validation results from the forensic psychiatry BN applications [1] demonstrated superior predictive performance against the state-of-the-art rule-based and data-driven models. More importantly, the BN models go beyond improving predictive accuracy and enhance usefulness for risk management purposes through intervention, as well as decision support in terms of answering complex questions that are based on unobserved evidence.

This BN development process is applicable to application domains which involve large-scale decision analysis based on complex information, rather than based on data with hard facts, and in conjunction with the incorporation of knowledge. The novelty extends to challenging the decision scientists to reason about building models based on what information is really required for inference, rather than based on what data is available and hence, encourages decision scientists to use available data in a much smarter way.

References

[1] A. C. Constantinou, N. Fenton, W. Marsh, and L. Radlinski, “From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support.,” *Artif. Intell. Med.*, vol. 67, pp. 75–93, Jan. 2016.

[2] N. E. Fenton and M. Neil, *Risk Assessment and Decision Analysis with Bayesian Networks*. Boca Raton: CRC Press, 2012.

[3] L. D. Hudson, B. S. Ware, K. B. Laskey, and S. M. Mahoney, “An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners.” 2005.

[4] J. L. Regens, N. Mould, C. J. Jensen, M. A. Graves, and D. N. Edger, “Probabilistic Graphical Modeling of Terrorism Threat Recognition Using Bayesian Networks and Monte Carlo Simulation,” *J. Cogn. Eng. Decis. Mak.*, vol. June, pp. 1–17, 2015.