How Bayesian Networks are pioneering the 'smart data' revolution

Norman Fenton and Martin Neil ask what next after 'big data', focussing on how Bayesian Networks are pioneering the 'smart data' revolution

The era of 'big data' offers enormous opportunities for societal improvements. There is an expectation – and even excitement – that, by simply applying sophisticated machine learning algorithms to 'big data' sets, we may automatically find solutions to problems that were previously either unsolvable or would incur prohibitive economic costs.

Yet, the clever algorithms needed to process big data cannot (and will never) solve most of the critical risk analysis problems that we face. Big data, even when carefully collected is typically unstructured and noisy; even the 'biggest data' typically lack crucial, often hidden, information about key causal or explanatory variables that generate or influence the data we observe. For example, the world's leading economists failed to predict the 2008–2010 international financial crisis because they relied on models based on historical statistical data that could not adapt to new circumstances, even when those circumstances were foreseeable by contrarian experts. In short, analysts often depend on models that are inadequate representations of reality – good for predicting the past but poor at predicting the future.

These fundamental problems are especially acute where we must assess and manage risk in areas where there is little or no direct historical data to draw upon; where relevant data are

difficult to identify or are novel; or causal mechanisms or human intentions remain hidden. Such risks include terrorist attacks, ecological disasters and failures of novel systems and marketplaces. Here, the tendency has been to rely on the intuition of 'experts' for decision-making. However, there is an effective and proven alternative: the smart data approach that combines expert judgment (including an understanding of underlying causal mechanisms) with relevant data. In particular, Bayesian Networks (BNs) provide workable models for combining human and artificial sources of intelligence even when big data approaches to risk assessment are not possible

BNs describe networks of causes and effects, using a graphical framework that provides rigorous quantification of risks and clear communication of

results. Quantitative probability assignments accompany the graphical specification of a BN and can be derived from historical data or expert judgment. A BN then serves as a basis for answering probabilistic queries given knowledge about the world. Computations are based on a theorem by the Reverend Thomas Bayes dating back to 1763 and, to date, provides the only rational and consistent way to update a belief in some uncertain event (such as a decline in share price) when we observe new evidence related to that event (such as better than expected earnings).

The problem of correctly updating beliefs in the light of new evidence is central to all disciplines that involve any form of reasoning (law, medicine and engineering as well as finance and indeed AI). Thus, a BN provides a

general approach to reasoning, with explainable models of reality, in contrast to big data approaches, where the emphasis is on prediction, rather than explanation and on association rather than causal connection.

BNs are now widely recognised as a powerful technology for handling risk, uncertainty and decision making. Since 1995, researchers have incorporated BN techniques into software products, which in turn have helped develop decision support systems in many scientific and industrial applications, including: medical diagnostics, operational and financial risk, cybersecurity, safety and quality assessment, sports prediction, the law, forensics and equipment fault diagnosis.

A major challenge of reasoning causally is that people lacked the methods and tools to do so productively and effectively. Fortunately, there has been a quiet revolution in both areas. Work by Pearl (Turing award winner for AI), has provided the necessary philosophical and practical instruction on how to elicit, articulate and manipulate causal models. Likewise, our work on causal idioms and influence diagrams has been applied in many application areas to make model building and validation faster, more accurate and ultimately more productive.

Also, there are now software products, containing sophisticated algorithms, that help us to easily design the BN models needed to represent complex problems and present insightful results to decision makers. Compared to previous generations of software these are more powerful and easier to use – so much so that they are becoming as familiar and accessible as spreadsheets became in the 1980s. Indeed, this big leap forward is helping decision makers think both graphically,

Our projects
Much of the recent and ongoing BN research described here is from projects:
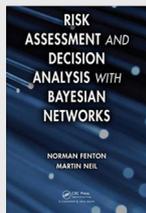www.bayes-knowledge.com
www.pambayesian.org
www.causaldynamics.org

Software
Much of the new BN functionality described here has been incorporated into version 10 of the AgenaRisk software (www.agenarisk.com)

Book
"Risk Assessment and Decision Analysis with Bayesian Networks" 2012, CRC Press by Fenton and Neil provides a thorough overview of BNs that is accessible to non-mathematical readers. Second edition available August 2018 (see https://www.crcpress.com/9781138035119)

about relationships and numerically, about the strength of these relationships, when modelling complex problems, in a way impossible to do previously.

Recent research has now made it easy to accurately incorporate numeric variables in the analysis, an obvious practical requirement, but one that the past generation of BN algorithms could not satisfy. There are now BN products that implement the latest and most accurate inference algorithms, as well as:

• Provide 'smart' learning of relationships from data – with or without missing values – incorporating as much or as little expert judgement as required.

• Automatically identify and select a decision strategy to maximise overall utility or minimise overall risk, using hybrid influence diagrams.

• Compute the 'value of Information' of uncertain variables in terms of how much should be paid to find more information out about them.

Our recent and ongoing research projects are providing ever more efficient algorithms both for building and deploying BNs (such as in patient-held medical devices and energy smart meters), including efficient cloud-based services for applications like cybersecurity risk analysis.

Many are asking what comes after 'big data'? Surprisingly, the ideas of Thomas Bayes, despite being pioneered over 250 years ago, may provide the answer in the form of smarter decisions from data and causal, uncertain knowledge.

Queen Mary
University of London

Norman Fenton
Professor of Risk and Information Management
Queen Mary University of London
Tel: +44 (0)20 7882 7860
n.fenton@qmul.ac.uk
www.eecs.qmul.ac.uk/~norman/
www.twitter.com/profnfenton