The Knowledge Engineering Review, Vol. 15:3, 2000, 303–304
Printed in the United Kingdom

Book review

Expert systems and probabilistic network models by Enrique Castillo, José Manuel Gutiérrez and Ali S. Hadi, Springer-Verlag, 1997, ISBN 0-387-94858-9, pp 605, DM 98.00.

The use of probability theory within the field of artificial intelligence has an interesting history. Once the field got beyond building systems which interacted with toy domains like the blocks world, it rapidly became apparent that any artificial intelligence system which had to deal with the real world needed some mechanism for representing and reasoning with uncertain information. A good example of this is the expert system MYCIN (Shortliffe, 1976). Because the medical information encoded in MYCIN is not certain, and so the rules that encode this knowledge represent plausible connections rather than logical implications, it is necessary to annotate the rules to indicate the degree to which the consequent follows from the antecedent. That way it is possible to establish the strength of the conclusions that can be drawn from particular observations.

At the time that the need to handle uncertainty was becoming apparent to the pioneers of artificial intelligence, probability theory was the obvious choice as the means to do so. (It was not quite the only means to do so, though the choice of alternative methods was a lot more limited than it is today, since both Zadeh's seminal paper on fuzzy sets (Zadeh, 1965), and Dempster's work on what later came to be known as the Dempster–Shafer Theory (Dempster, 1968) had already appeared.) However, using probability theory threw up a major fundamental problem. This was the amount of information required about a domain before the theory could be used to reason about that domain. Basically, to fully specify a probability model involving N binary variables requires $2^N - 1$ probability values, and this is clearly a huge number for a reasonably large domain model. Furthermore, even if these values can be provided (for instance, using techniques such as maximum entropy to supply values which are unknown) for some intelligent system, this large number of probability values must be updated by the system when it obtains new information about the world. Thus, from the perspective of many people in artificial intelligence, attempting to use probability theory was ruled out for two reasons (though both stemmed from the same problem); it was difficult to build probability models and any models that might be built had very high computational overheads.[1] Instead such people turned to the development and use of models such as certainty factors and fuzzy logic, and the adherents of probability theory became somewhat detatched from the mainstream of research in artificial intelligence.

Undaunted by this, the detached group continued to investigate the ways in which probability theory could be used, in parallel with a study of alternative approaches, and many of the results that were obtained may be found in the *Proceedings of the Workshop* (later, Conference) on *Uncertainty in Artificial Intelligence*. Electronic versions of many of these proceedings may be found at `http://www2.sis.pitt.edu/~dsl/UAI/uai.html`. Of all the many results, the most significant is the development of probabilistic causal networks (also known as Bayesian networks).

These network models are significant because they address exactly the problem outlined above, that of the large number of probability values resulting in model building difficulties and high computational overheads. The way that the models handle this problem is by making it possible to explicitly represent the independencies between the propositions that make up the domain model. Once these independencies are identified, it is possible to dramatically decrease the number of probability values required to specify a model and to identify updating algorithms which run efficiently in most cases.

[1] This is a slight over-simplification. The state of the art at the time in question allowed for reasonably efficient updating, but only under a particularly restrictive assumption about the relationship between the propositions whose probabilities were in question.

While the classic reference for causal probabilistic networks is Pearl's book (Pearl, 1988), this being based upon the papers which originally identified the advantages of such models, there has been much development since then which has been unavailable in book form. Such topics included approximate algorithms for updating probability values and techniques for learning probabilistic causal networks from data, the latter being a subject which has been the centre of much attention in recent years. With Neapolitan's excellent book (Neapolitan, 1990) being out of print, there was clearly a space in the market for a book which summarised the state of the art in enough detail that even post-doctoral researchers in the field could use it, rather than the primary literature, as their source.[2] This gap is now filled. *Expert Systems and Probabilistic Network Models* is as comprehensive and detailed a treatment as I have read, and it is comprehensive and detailed enough to meet the needs of most people who find themselves interested in the field. Indeed, the only criticism I can find to make about it is that it is rather dismissive of non-probabilistic approaches to handling uncertainty (unlike the authors I happen to believe that such approaches have their place), but this is certainly not a reason to dismiss the book. On the contrary, I would enthusiastically recommend it to anyone who wants a detailed understanding of how to use probability theory in artificial intelligence, with the rider that it might be worth reading Jensen's book first.

References

Dempster, AP, 1968. A generalisation of Bayesian inference. *Journal of the Royal Statistical Society*, *B* 205–232.

Jensen, FV, 1996. *An Introduction to Bayesian Networks*. UCL Press, London.

Neapolitan, RE, 1990. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley, New York.

Pearl, J, 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Palo Alto.

Shortliffe, EH, 1976. *Computer-based Medical Consultations: MYCIN*. Elsevier, New York.

Zadeh, LA, 1965. Fuzzy sets. *Information and Control* **8** 338–353.

Reviewed by Simon Parsons, Queen Mary and Westfield College, University of London, UK

[2] Those who required less detail on the underlying mathematics were already well served by Jensen's recent volume (Jensen, 1996).

ERRATUM

The Knowledge Engineering Review, Vol. 15:2, 2000, 181–185

Game theoretic and decision theoretic agents

Simon Parsons and Michael Wooldridge

An incorrect URL was published on page 181. The correct version appears below.

`http://www.csc.liv.ac.uk/~sp/events/gtdt/gtdt99/`

The publishers would like to apologise for this error to the authors and their readers.