The Knowledge Engineering Review, Vol. 13:4, 1998, 409–412
Printed in the United Kingdom

Book reviews

Rationalizing Medical Work: Decision-Support Techniques and Medical Practices by Marc Berg, MIT Press, Cambridge, MA, 1997, ISBN 0-262-02417-9, pp 256, $30.00 (cloth)

A distinctive feature of artificial intelligence is that it is often the object of analyses proposed by academics in other fields. Marc Berg's thesis is written in the style and vocabulary of sociology, and readers used to digesting work in the somewhat leaner language of computer science may find its phraseology rather rich: for example, Berg offers at one point a "contrast between the domination of the sole reductionist voice and the rich multivocality of the reduced". The book, however, is short, the argument lucid and Berg's criticism of AI intended to be constructive: it is not that AI is impossible or unacceptable, but that realising its benefits requires a fuller understanding of the consequences of its application.

The topic of the book is the use of formal tools in medicine, and the discussion opens with an account of how ideas of what would constitute 'rational' medicine have shaped, and been shaped by, the development of rationalising tools such as expert systems, decision analysis and clinical protocols.

Berg charts the post-war debate about the scientific character of medicine and proposals for rationalising medical practice. Doctors were first thought to be involved in the application of scientific knowledge. By the 1960s, however, it was being argued that not just medical knowledge but also medical practice should be scientifically based. Berg suggests that statistics began to be seen not just as a tool to be used in establishing medical knowledge but as a model for the kind of thinking that should inform medical decisions. In 1959 Ledley and Lusted characterised diagnosis as a process of reasoning on the basis of uncertain, or probabilistic, information. Information about the frequency of a disease in a population and the frequency with which given symptoms were associated with the disease could be combined to give a precise estimate of the probability that a patient showing these symptoms had that disease. Determining the most likely diagnosis would be the matter of a calculation.

A different model of medical reasoning was being developed by research into artificial intelligence. Many early AI systems attempted to represent the knowledge and reasoning of expert diagnosticians. Clinicians were not statisticians but logicians, and computerised "expert systems" would be perfect logicians with access to rule bases representing the wisdom of a plenitude of experts. The third kind of formal system considered by Berg is that of the protocol. Protocols are attempts to rationalise the practice of medicine conceived as a "scientific process of distinctive clear-cut steps". Clinical trials can be used to compare investigations, treatments and combinations of treatments and the protocols can be used to promulgate their findings; ensuring that all patients have the benefit of those procedures which have been proved to be the most effective.

Berg considers an example of each of these three kinds of decision aids: De Dombal's statistical system for advising on the treatment of patients with acute abdominal pain; Emerson and Wyatt's work on ACORN, a rule-based system for advising on the admission of patients to a coronary care unit and Eddy's protocol for the management of breast cancer. In each case the developers were obliged to make compromises which limited the value of their work. The tools became "localised" as they became tailored to the requirements of a particular setting ("localisation in space"), the number of tasks to be handled has to be restricted ("localisation in scope") and authors have to back away from the original justification for their work ("localisation in rationale").

The central problem is the difficulty of reconciling medical practice with formal systems. Berg illustrates this with a set of anecdotes. One concerns the management of a patient who developed a fever following a bone marrow transplant. The value of this story is that the items of data—on which decision making is apparently based—are not obtained by observation or measurement but constructed as the patient's case is made to fit an emerging understanding of what is going on. Measurements of the patient's fluctuating temperature are successively defined as: "a single peak", "stabilising", "still rising" and "not responding to treatment". A blood sample reveals evidence of infection in one of four tubes analysed: the oncologist regards this as an argument for altering treatment but the bacteriologist later says "one out of four tubes is nothing" and that treatment must be determined by clinical evidence. The clinical evidence in this case concerns an anal fissure, which was not discovered until after the first course of treatment had started. The case against AI, as stated by Berg, is that doctors are not the kind of rational processors of information suggested by proponents of formal systems but find themselves thrust into the middle of events and compelled to act. Rules, statistics and procedures are not the foundation of action: they are "post-hoc, situation-dependent derivatives of action".

It is because aspects of medical practice are chaotic and unruly that "localisation" occurs, as system designers are forced to rethink what they do in the face of forces and circumstances they can't control. Perhaps the most interesting aspect of Berg's thesis is the suggestion that as well as tools being altered to fit existing practice, medical practice is disciplined to meet the requirements of formal tools. Making tools work is a process of negotiation. Inserting a formal tool into a clinical setting inevitably alters the setting, through the installation or reinforcement of bureaucratic hierarchies, an increased structuring of the medical personnel's work (for example through the use of standard data collection forms) and a concentration on data items which can be gathered reliably and unequivocally (indeterminate and unstructured information—such as details of patient's home life which a clinician might use in deciding whether to admit him or her—are necessarily ignored).

The crux of Berg's argument is that formal tools and medical practice are intertwined. The debate about such systems has tended to set up a contrast between tool and practice. For some the lack of diffusion of tools is due to resistances in medical practice, for others the tools fail to capture the essence of human practice. But tools and practice develop together and transform each other. This is particularly clear in the case of research protocols which have become part and parcel of modern oncology. We should not be asking if a formal tool adequately captures the essence of medical practice, any more than we would ask if a CT scanner does its job properly or whether humans should be left to do the work. The appropriate question to ask is: what would be the consequences of adopting a tool in a particular setting? Answering this involves stepping outside the discipline of the system designer and understanding the social, material and organisational aspects of medical work. The process of designing a tool is not one of shaping it to fit a pre-existing niche, it also involves constructing a niche. Poor user interfaces or an absence of inter-operability have perhaps hindered the take up of decision support systems but such systems could never be designed to fit seamlessly into existing ways of working: the use of formal tools inevitably changes medical practice. And it does so in ways which have drawbacks as well as benefits. The challenge, set by Berg, is to search for ways in which formal tools can become "familiar yet never totally transparent, powerful yet fragile agents of change".

Reviewed by Paul Taylor, CHIME, UCL Medical School, London

Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.

In 1990, a teacher preparing the notes for a course in *neural networks* found it hard to recommend relevant reading, besides the excellent compendium by Rummelhart and McClelland. Several authors had tried their hand and written textbooks, but none seemed satisfactory. The time that had elapsed since the outburst of interest in connectionism in mid-eighties had been too short, and many questions that a teacher wanted to ask were still open. What is fundamental and what is marginal? What deserves detailed analysis, and what should only be mentioned in passing? Familiarity with practical implementations and teaching experience were simply not enough in a discipline where 90% of major breakthroughs happened just recently. Textbooks soon became outdated because they overlooked vital innovations, while indulging on topics that were soon to become obsolete.

Fortunately, in the last few years the situation has improved. The basic paradigms have been established, the major issues defined. The interested reader can now choose from a whole spectrum of books ranging from popular introductions to extensive monographs. Rather than with time, authors now have to fight with competitors on a maturing market that has recovered from the peak of the neural-network fad.

Where does Simon Haykin's *Neural Networks: A Comprehensive Foundation* fit into this spectrum? The brief summary on the cover promises that "...this book is ideal for professional engineers and graduate students this exciting field," but the book was hardly conceived as an easy-to-read primer. For one thing, its sheer volume (about 700 pages) is discouraging for a newcomer. The text teems with formulae from matrix algebra, statistics and calculus, and an undergrad without strong mathematical background will soon get lost.

Rather, the work represents an *advanced reading*, invaluable for those who have already acquainted themselves with the field, and now seek deeper understanding of its fundamentals. It summarizes the status quo of the emergent technology and helps us find orientation in the never-ending stream of competing theories, so often oblivious of each other. And the presentation is serious. Whereas some computer scientists view their discipline as an armory of algorithms and techniques, Simon Haykin belongs to those that strive for rigorous science that rests on provable truths and detailed theoretical analysis. *Ad hoc* rules of thumb and heuristics do not have sufficient appeal to him.

To start with, he introduces some *basic concepts*, such as artificial neurons, networks of neurons, and the essentials of learning. A well-written overview of the historical evolution of these concepts, as well as their relation to neuroscience and the conventional computing completes the picture and sets stage for the sequel.

Then, several chapters address the most popular methods of *supervised learning* in neural nets. After a detailed treatment of the correlation matrix memory and of several techniques pertaining to single-layer networks, the author details the principles of the backpropagation of error through multilayer networks, explains network-pruning techniques, and analyzes the deep statistical theories behind the now popular radial-basis function networks. This broad background then enables him to proceed from purely feedforward networks to a much greater challenge: to networks with recurrent links. In particular, he studies the Hopfield-network based autoassociative memory, heteroassociation using the Boltzman machine, and the implications of the mean-field theory developed by modern physics.

A few chapters are devoted to the three major strands in *unsupervised learning* and in self-organizing systems whose task is to discover regularities in data: hebbian learning (with emphasis on the principal components analysis), competitive learning (covering the Kohonen-like approaches), and the information-theoretic models rooted in Shannon's information theory and in the principle of maximum information preservation. A natural extension of these techniques leads to the idea of modular networks involving the use of adaptive experts: the individual networks learn in a supervised way, whereas the expert overseeing these networks functions in a competitive mode.

A separate chapter is dedicated to the problem of *time* as a phenomenon inseparable from learning, and yet so difficult to model. Another extensive chapter discusses the principles of *neurodynamics*, with rigorous treatment of such issues as stability, attractors, chaos and neurodynamical models: neural nets are presented as dynamic systems that can be subjected to deep physical considerations and precise analysis. The last chapter deals with questions related to hardware implementations. The author provides the reader with the basic outline and describes examples of those products that can be viewed as prototypical.

The book thus covers all major facets of artificial neural networks, sometimes at great lengths, and sometimes perhaps too rigorously, even at the cost of clarity. Many issues might have been approached using simpler terms, and, at times, the maths seems to be somewhat superfluous. But I would hardly dare to criticize this because many researchers with strong background in statistics will appreciate this approach. Personally, I see the book rather as an excellent source of reference than as a textbook.

But the single most important feature is the solid physical and statistical roots the book builds on. Roots that have been around for decades, have stood the test of time, and yet have so far been largely ignored by many connectionists. Ten years from now, the preferences of the new generation of scientists will surely depart from today's views and preferences. But the core of Haykin's *Neural Networks* will not lose its value for those who prefer to see this world described by formulae, rather than by narratives.

Reviewed by Miroslav Kubat, University of Southwestern Louisiana.