Probabilistic Similarity Networks

By David E. Heckerman
Cambridge, MA: The MIT Press, 1991.
Pp. xx + 234. $35.00.

Reviewed by: Stephen F. Roehrig
School of Urban and Public Affairs
Carnegie Mellon University
Pittsburgh, PA 15213-3890
roehrig@andrew.cmu.edu

This is a book about *normative* expert systems, that is, systems which treat uncertainty using the normative framework of probability theory. The gauntlet of probabilistic reasoning in expert systems was thrown down by the developers of MYCIN when they originated certainty factors. It has been picked up by quite a few, and serious battles have been joined. Now one of the owners of the glove has admitted a success by a probabilist.

In his foreword to this book, Ted Shortliffe lays down his major criterion for such a success—the development of "a nontrivial, efficient and effective system using formal probabilistic methods." On reading the book, it is clear why Shortliffe has conceded. *Probabilistic Similarity Networks*, which won Heckerman the 1990 ACM Doctoral Dissertation Award, lays out in detail the development of Pathfinder IV, an expert system for the diagnosis of lymph-node diseases. In it we find a new technique for easing the difficulty of constructing the system's network skeleton (similarity networks), an approach for simplifying the assessment of conditional probabilities in the resulting knowledge graph (partitions), a description of the application of these theoretical ideas to an important real world problem, and a careful evaluation of the results.

Work on probabilistic methods during the 1980s was crucial in paving the way for the current work. But only some of the apprehensions about using true probabilities in rule-based systems were calmed by the fruit of that decade (hence the reservations of Shortliffe and others). Bayesian networks (a.k.a. probabilistic networks, belief nets, influence diagrams *sans* decision nodes, and now, knowledge maps) are currently seen as an extremely powerful way to break down the exponential problem of specifying a joint distribution over a collection of discrete random variables. Various computational techniques for updating a distribution in light of fresh evidence have appeared, and these too have quieted some arguments against the use of probabilities. Many arguments still remain, of course, and while hardly anyone maintains that probabilistic networks are a universal means of knowledge representation in AI, this body of results is impressive.

Assuming that you believe that a network representation is suitable for your problem domain, and that you want your system backed by the normative force of probability theory, there remain the nagging problems of the actual construction and assessment of the network. Heckerman's first achievement is the theory of similarity networks, which helps to simplify the elicitation of the network which records conditional dependences among the variables at hand. By allowing the expert to focus on simpler conditioning events, the similarity network and associated "local knowledge maps" simultaneously streamline the development of the overall network structure and reduce the number of conditional probabilities which must be assessed.

A small example may help. Suppose node A (designated the "distinguished node") has three possible outcomes, a_1, a_2, and a_3. We may think of node A as a "disease" node, with the a_is representing three outcome diseases which are mutually exclusive and exhaustive in the domain at hand. Other nodes in the graph might represent symptoms. Suppose further that, in the judgement of the expert, outcomes a_1 and a_2 are quite similar, so that the set of symptoms that can distinguish them is a subset of the complete symptom set for the disease node taken as a whole. The local knowledge map for this disease pair, which records only the distinguishing features, will be correspondingly less elaborate than the global knowledge map, which must account for all diseases and symptoms. Heckerman shows that these local knowledge maps, constructed for pairs of outcomes, can be used as building blocks for the global map. In the carefully worked out theory, he shows that the global representation, formed as the graph union of the local maps, is necessarily correct given the opinions of the expert at the local level.

A similarity network, distinct from but related to, the overall knowledge map, has as its nodes the possible disease outcomes, and each arc represents the expert's view that the nodes (diseases) joined by it are similar from the diagnostic point of view (and thus difficult to distinguish). If two diseases are linked by an arc—are similar—then presumably only a few symptoms will be useful as clues in differentiating them. These are just the symptoms included in the local map for that pair. The local maps thus encode instances of *subset independence*. Heckerman's point is that quite often the probability of any one of a subset of possible diseases is independent of a particular symptom, given that the true disease is in the subset, while the probabilities diseases outside the subset are not.

By using the same general strategy of grouping over similar outcomes, the assessment of probabilities is also simplified. If a certain symptom doesn't help to distinguish between a pair of diseases, the conditional probabilities of that symptom given either of the diseases must be the same. More generally, a *partition* of the set of outcomes in the disease node, relative to a symptom, is a grouping of diseases such that the symptom is not relevant in terms of distinguishing which disease actually obtains. In Pathfinder IV, the similarity network idea and the use of partitions reduced the number of conditional probabilities which needed to be assessed more than fivefold.

The book is well written and organized in a straightforward way. The first chapter is introductory, and Chapter 2 proceeds directly to a concrete example, again of the medical variety. Much of the work presented in the book has appeared in journal articles and working papers, but the approach in this chapter is more relaxed and intuitive. It might be possible to decide if similarity networks are suitable for one's own expert system project just from the description here. However, extensions of the theory are given in later chapters, so one should read through if in doubt.

Chapter 3 gives the full theoretical story. The basic problem is to show that the simplified local knowledge maps can be combined to give a global knowledge map which preserves the probabilistic meaning intended by the expert. By analogy with logic, Heckerman defines *soundness* and *completeness* for transformations of probability constraints. In these terms, he is able to show that the global map formed by transformations on the simplified maps (through some intermediate steps that needn't concern us here) is equivalent to the global map which would have been obtained by conventional means.

There are some technical caveats to this development. First, the overall distribution must be strictly positive, that is, no entry in the joint distribution may be zero. Second, the structure of the knowledge map is restricted; in particular, there must be a single top-level node on which the similarity network is constructed. For medical diagnosis problems, this is conveniently the disease node, consisting of mutually exclusive and exhaustive outcomes. All symptoms are taken as following causally from one or more diseases (but there may also be conditional dependencies among them). This certainly seems quite severe; if all the "action" in a knowledge map is not clustered around a central node, much of the simplification in construction and assessment disappears.

1.1.1992

Later in the book, Heckerman suggests a work-around for this problem when the distinguished node has only a few predecessors, but does not comment on the more general situation where the network is large in diameter. Finally, it is immediately apparent that similarity networks have little to offer probabilistic networks whose nodes are primarily propositional or binary. The method depends on analyzing subsets of outcomes, and there are only trivial subsets for nodes which are either true or false.

Chapters 4 and 5 are devoted to a discussion of Pathfinder. In Chapter 4, its similarity network is presented, and examples of local knowledge maps are given. Statistics on the construction cycle show, as mentioned before, that a large reduction in the number of assessed probabilities can be had, while still retaining the validity of the finished product. The inference (probability propagation) algorithm, admittedly not optimised for the application, is discussed briefly.

From the medical point of view, the two most important question remaining are, "How does Pathfinder IV perform?" and "Was it worth it?" Chapter 5 addresses these questions. The first is approached by having the domain expert rate Pathfinder's output (an ordering of diseases by likelihood) given input from case histories. This is both an overall rating and a comparison with the output generated by an version of Pathfinder built before the similarity network concept was developed. The second question attempts to use a decision-theoretic framework to judge the cost-effectiveness of the Pathfinder project. A utility model for diagnosis was constructed, and was used to compute the expected utility of Pathfinder's diagnoses. I leave it to other readers to decide how convincing this portion of the evaluation is.

For me the greatest contribution of Heckerman's work is that is shows us how to exploit a new form of independence (subset independence) in knowledge maps. Even armed with the knowledge that, in a network, node probabilities conditioned only on direct predecessors are sufficient to completely specify the entire joint distribution, huge complexities can still arise. If the graph is complete (each node connected to all others), we have gained essentially nothing. Even in sparse graphs, large numbers of conditional probabilities may need to be collected. Any relief, by means of subset independence or other insights which will undoubtedly follow, is welcome. Probabilistic networks have been the subject of intense study for many years now, but as Heckerman shows, there is still room for clever analysis.

This is not to shortchange the remainder of the work. In addition to the Pathfinder system itself, Heckerman and his collegues have constructed a program called SimNet, which organizes and partially automates the process of assembling the local knowledge maps and assessing them using partitions. From the expert's point of view, the program provides a kind of "graphical user interface" to the process of knowledge engineering.

Probabilistic Similarity Networks shows that there is often more to a probabilistic network than meets the eye. A certain level of qualitative structure is apparent from the unassessed network, since conditional independences between *nodes* can be read off once the rules are learned. But now we see that there is another level of structure beneath. Decomposing a node into subsets of outcomes provides new independencies to exploit. There is no reason to suppose that further insight along these and similar lines will not yield further increases in the expressiveness and compactness of network representations.