A Tractable Inference Algorithm for Diagnosing Multiple Diseases[1]

David Heckerman

Medical Computer Science Group
Knowledge Systems Laboratory
Departments of Computer Science and Medicine
Stanford, California 94305

We examine a probabilistic model for the diagnosis of multiple diseases. In the model, diseases and findings are represented as binary variables. Also, diseases are marginally independent, features are conditionally independent given disease instances, and diseases interact to produce findings via a noisy OR-gate. An algorithm for computing the posterior probability of each disease, given a set of observed findings, called *quickscore*, is presented. The time complexity of the algorithm is $\mathcal{O}(nm^- 2^{m^+})$, where n is the number of diseases, m^+ is the number of positive findings and m^- is the number of negative findings. Although the time complexity of quickscore is exponential in the number of positive findings, the algorithm is useful in practice because the number of observed positive findings is usually far less than the number of diseases under consideration. Performance results for quickscore applied to a probabilistic version of Quick Medical Reference (QMR) are provided.

1   Introduction

One of the most common criticisms of the use of probability theory in expert systems is that the theory is impractical to apply in realistic situations (Buchanan and Shortliffe, 1984). In attempts to answer this criticism, several researchers at Stanford, myself included, have undertaken the task of converting to a probabilistic framework Quick Medical Reference (QMR) (Miller, 1987), one of the largest medical expert systems in existence.

We have made a straightforward and tractable transformation of the QMR knowledge base to a probabilistic model (Heckerman and Miller, 1986; Henrion, 1990; Shwe et al., 1990). Like the heuristic algorithms in QMR, the model allows for any combination of diseases to be present in a patient. Unfortunately, this feature leads to a time complexity
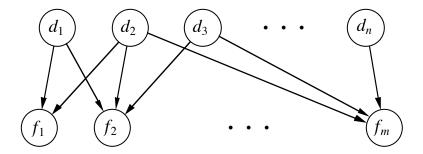
Figure 1: A belief network for diagnosing multiple diseases.
Diseases are marginally independent. Findings are conditionally independent given a disease instance.

for inferring the probability of each disease given a set of findings that is exponential in the number of *diseases* ($O(2^n)$) for all known algorithms. As there are over 600 diseases in QMR, these algorithms are intractable. Furthermore, this problem is known to be NP-hard (Cooper, 1987).

In this paper, I present *quickscore*, an algorithm with time complexity that is exponential in the number of *positive findings* (findings observed to be present rather than absent). More precisely, the algorithm has a time complexity of $O(nm^- 2^{m^+})$, where m^+ is the number of positive findings and m^- is the number of negative findings. Although quickscore has an exponential time complexity, it is useful in practice because the number of observed positive findings is often far less than the number of diseases under consideration. For many realistic patient cases, quickscore implemented on a Macintosh II produces an answer in less than 1 minute of real time.

2 The QMR model

The probabilistic version of QMR is called QMR-DT for Decision-Theoretic QMR. A belief network for QMR-DT, is shown in Figure 1. Each of the n nodes in the upper layer of the network represents a disease that may be present or absent in a patient. Each of the m nodes in the lower layer represents a finding that may be observed to be present or absent in the patient, or that may not be observed at all. The problem of interest is to compute the probability of each disease given a set of positive and negative findings.

As indicated by the network in Figure 1, we assume diseases to be marginally independent. Also, we assume that findings are conditionally independent given any instance of the set of diseases. An *instance of a set of diseases* is an assignment of present or absent to each disease in that set.

We also assume that diseases act independently to cause any given finding to be present. A belief network that represents this independency for two diseases is shown in Figure 2. The node labeled d_i–causes–f represents the event that d_i causes finding f to be present. The node with the double boundary, labeled f, is a deterministic node that says finding
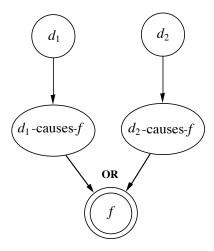
Figure 2: A belief network for the noisy OR-gate.

If disease d_1 is present, then with some probability it will act to cause finding f. This probability depends neither on the state of d_2 nor on whether d_2 acts to cause f. A reciprocal relationship holds for disease d_2. If either disease acts to cause f, then f will be observed to be present.

f will be present if and only if d_1 causes f to be present, d_2 causes f to be present, or both d_1 and d_2 cause f to be present. The arc between the node labeled d_1 and the node labeled d_1–causes–f reflects our assumption that the presence or absence of d_1 influences the probability that d_1 causes f to be present. In particular, we assume that, if d_1 is present, it may cause f to be present with some probability. If d_1 is absent, we assume that the disease cannot act to cause f. The same set of assumptions holds for the disease d_2. Finally, the lack of arcs between nodes in the upper two rows of the figure reflects the assertions that (1) the probability distribution for the variable d_1–causes–f depends neither on the absence or presence of d_2 nor on the absence or presence of the event d_2–causes–f, and (2) the probability distribution for the variable d_2–causes–f depends neither on the absence or presence of d_1 nor on the absence or presence of the event d_1–causes–f. We call this form of conditional independence *causal independence* to distinguish it from the type of conditional independence that is commonly represented in belief networks.

Under these assumptions, we can compute the probability of f given any instance of the pair $\{d_1, d_2\}$ from the two assessments p_1 and p_2, where p_i is the probability that d_i causes f, $i = 1, 2$. If both diseases are absent, finding f must be absent. If only one disease is present—say, d_i—then

$$p(f^+|\text{only } d_i) = p_i$$

and

$$p(f^-|\text{only } d_i) = 1 - p_i$$

where f^+ and f^- denote the presence and absence of finding f, respectively. That is, the probability that d_i causes f is just the probability of f given that only disease d_i is present. If both d_1 and d_2 are present, finding f will be absent only if both diseases fail

to cause f to be present. Therefore,

$$p(f^-|d_1^+, d_2^+) = (1 - p_1)(1 - p_2) = p(f^-|\text{only } d_1)p(f^-|\text{only } d_2)$$

where d_i^+ denotes the presence of disease d_i.

More generally, suppose that n diseases can potentially cause finding f, and that these n diseases are the only causes of f. As in the simpler case,

$$p(f^+|\text{only } d_i) = p_i$$

where p_i is the probability that d_i causes f. Let D_i denote a particular instance of the n diseases and let D_i^+ denote the set of diseases that are present in the instance D_i. Given instance D_i, f will be absent only if none of the diseases in D_i^+ cause f to be present. It follows that

$$p(f^-|D_i) = \prod_{d \in D_i^+} p(f^-|\text{only } d) \tag{1}$$

Thus, using Equation 1, we can compute the probability of f given any of the 2^n disease instances from only n assessments.

The assumptions underlying Equation 1, including the assumption of causal independence, have been described previously (Good, 1961), and several researchers have suggested that these assumptions be used to model medical domains (Habbema, 1976; Peng and Reggia, 1986). Pearl has called this canonical model of cause and effect the *noisy* OR-*gate* (Kim and Pearl, 1983; Pearl, 1986). Other researchers have extended the model to accommodate situations where a finding may appear in the absence of all explicitly represented diseases, a leaky OR-gate (Suppes, 1970; Henrion, 1989).

The leaky OR-gate model was assumed implicitly by the developers of the QMR expert system (Miller, 1987). In conjunction with the independence assumptions shown in Figure 1, the model made tractable the transformation of QMR into a probabilistic framework. As we see in the following section, these assumptions also accommodate an inference algorithm that is tractable for many patient cases. In Section 5, we examine the appropriateness of these assumptions, and discuss how they might be relaxed.

3   The Quickscore Algorithm

The goal of quickscore is to compute the probability of each disease d_i, $i = 1, 2, \ldots n$, given a set of positive findings F^+ and a set of negative findings F^-, under the assumptions described in the previous section.[2] To understand how quickscore works, we first compute the probability that a single finding f will be absent. Using the expansion rule, we get

$$p(f^-) = \sum_{D_k \in D} p(f^-|D_k)p(D_k) \tag{2}$$

---

[2]Here, we develop the algorithm for nonleaky OR-gates. We can, however, extend easily the development to include leaky OR-gates.

where D is the set of all disease instances. Using the assumption that diseases are marginally independent and the assumptions underlying the noisy OR-gate, Equation 2 becomes

$$p(f^-) = \sum_{D_k \in D} \left[\prod_{d \in D_k^+} p(f^-|\text{only } d) \prod_{d \in D_k^+} p(d^+) \prod_{d \in D_k^-} p(d^-) \right] \tag{3}$$

where D_k^- denotes the set of diseases that are absent in the instance D_k.

Now consider the expression

$$\prod_{i=1}^{n} \left[p(f^-|\text{only } d_i)p(d_i^+) + p(d_i^-) \right] \tag{4}$$

If we multiply out this expression, we see that it is just the right-hand side of Equation 3. Thus, we obtain

$$p(f^-) = \prod_{i=1}^{n} \left[p(f^-|\text{only } d_i)p(d_i^+) + p(d_i^-) \right] \tag{5}$$

The difference in time complexity between the computations of Equations 3 and 5 is striking. The computation in Equation 3 is a sum over 2^n terms, whereas the computation in Equation 5 is a linear product over n sums. Pearl (Pearl, 1988, pp. 187-188) was the first to note the equivalence between these two computations. As we shall see, quickscore derives its speed from this equivalence.

Under the assumption that findings are conditionally independent given any disease instance, we can employ the transformation described in the previous paragraph to compute the probability that the set of negative findings, F^-, are observed. We have

$$p(F^-) = \prod_{i=1}^{n} \left(\left[\prod_{f \in F^-} p(f^-|\text{only } d_i) \right] p(d_i^+) + p(d_i^-) \right) \tag{6}$$

The situation is more complex for positive findings. Let us first examine the simple case where $F^+ = \{f_1, f_2\}$. Applying the expansion rule to $p(f_1^+, f_2^+)$ and using the assumption of conditional independence of findings, we get

$$p(f_1^+, f_2^+) = \sum_{D_k \in D} p(f_1^+|D_k)p(f_2^+|D_k)p(D_k) \tag{7}$$

Since $p(f_j^+|D_k) = 1 - p(f_j^-|D_k)$, Equation 7 becomes

$$\begin{aligned} p(f_1^+, f_2^+) &= \sum_{D_k \in D} p(D_k) - \\ &\quad \sum_{D_k \in D} p(f_1^-|D_k)p(D_k) - \\ &\quad \sum_{D_k \in D} p(f_2^-|D_k)p(D_k) + \\ &\quad \sum_{D_k \in D} p(f_1^-|D_k)p(f_2^-|D_k)p(D_k) \end{aligned} \tag{8}$$

The first sum in Equation 8 is equal to 1. The remaining terms are in the same form as the right-hand side of Equation 2. Thus, using the algebraic transformations derived previously, we obtain

$$
\begin{aligned}
p(f_1^+, f_2^+) \ = \ & 1 - \\
& \prod_{i=1}^{n} \left[p(f_1^- | \text{only } d_i) p(d_i^+) + p(d_i^-) \right] - \\
& \prod_{i=1}^{n} \left[p(f_2^- | \text{only } d_i) p(d_i^+) + p(d_i^-) \right] + \\
& \prod_{i=1}^{n} \left[p(f_1^- | \text{only } d_i) p(f_2^- | \text{only } d_i) p(d_i^+) + p(d_i^-) \right]
\end{aligned}
\tag{9}
$$

More generally,

$$
p(F^+) = \sum_{F' \in 2^{F^+}} (-1)^{|F'|} \prod_{i=1}^{n} \left(\left[\prod_{f \in F'} p(f^- | \text{only } d_i) \right] p(d_i^+) + p(d_i^-) \right)
\tag{10}
$$

where 2^{F^+} denotes the power set of F^+, and $|F'|$ denotes the number of elements in set F'.

In the most general case where some findings are present and some are absent, we can combine Equation 6 and Equation 9 to obtain

$$
p(F^+, F^-) = \sum_{F' \in 2^{F^+}} (-1)^{|F'|} \prod_{i=1}^{n} \left(\left[\prod_{f \in F' \cup F^-} p(f^- | \text{only } d_i) \right] p(d_i^+) + p(d_i^-) \right)
\tag{11}
$$

It is now a simple matter to compute $p(d_i | F^+, F^-)$. First, we compute $p(F^+, F^-)$. Then, we compute $p(d_i, F^+, F^-)$ by setting $p(d_i^+) = 1$ in Equation 11. The sought-after probability is then

$$
p(d_i | F^+, F^-) = \frac{p(d_i, F^+, F^-)}{p(F^+, F^-)}
\tag{12}
$$

by the product rule of probability.

The quickscore algorithm can provide intermediate results. In particular, suppose that we order the findings in F^+—say, f_1, f_2, ... f_{m^+}. In Equation 11, we can first compute the term in the power set of F^+ corresponding to f_1 alone. We can then compute the terms in the power set that correspond to combinations of only f_1 and f_2. Continuing in this way, the probability of each d_i given the first j findings can be recovered from the algorithm at any time, where time is an exponential function of j. The fact that quickscore can provide intermediate results may prove useful. The QMR knowledge base contains a partial ordering of findings by their general clinical importance in determining a diagnosis. This ordering might be used to degrade gracefully the performance of quickscore with increasing time constraints.
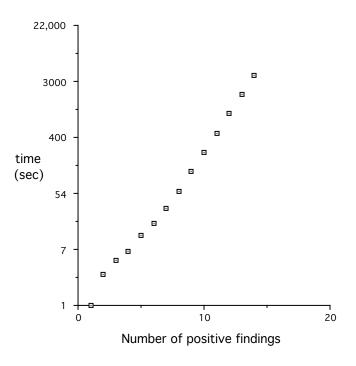
Figure 3: Run times for quickscore.
Run times for quickscore implemented in LightSpeed Pascal on a Macintosh II are plotted against the number of positive findings in a case. Nine positive findings typically can be scored in less than 1 minute.

4 Run-Time Performance of Quickscore

Quickscore has been implemented in Lightspeed Pascal on the Macintosh II computer. Figure 3 shows the run time of the algorithm for cases of various size. The cases are taken from a library of classic cases used by the QMR research team to test periodically the diagnostic accuracy of the heuristic knowledge base. These cases contain only positive findings. As the graph indicates, nine findings can typically be scored in less than 1 minute. Overall, in 25 percent of the 400 cases in the library, quickscore requires 15 minutes or less to score each case.

5 Weaknesses of the Algorithm

The diagnostic model described in this paper contains several assumptions that may not be appropriate for many medical and nonmedical domains. For example, some diseases and findings can occur with different degrees of severity, and hence are not two-valued. In addition, certain diseases cause others to be present. Consequently, all diseases are not marginally independent. Furthermore, some findings are conditionally dependent, and certain diseases are caused by findings rather than vice-versa (e.g., a history of alcoholism tends to cause cirrhosis of the liver). None of these observations pose a significant barrier to the translation of the QMR's knowledge base to a probabilistic framework.

For example, several researchers have generalized the noisy OR-gate model, constructing other prototypic models that embody the assumption of causal independence (Kim and Pearl, 1983; Heckerman, 1987; Henrion, 1989). We can use these prototypic models to accommodate multiple-valued diseases and findings, and to accommodate causal interactions among diseases. Also, by introducing hidden or unobservable pathophysiologic states, we can capture many of the dependencies among findings. Again, we can use the noisy OR-gate and its extensions to model interactions between diseases and pathophysiologic states. Unfortunately, we cannot extend quickscore in a straightforward manner to treat these extensions to the current QMR-DT model.

In addition, quickscore is unstable numerically when $p(F^+, F^-)$—the probability of observations for a given case—is small, because $p(F^+, F^-)$ is a sum of positive and negative terms on the order of 10^0 (see Equation 11). IEEE-standard extended-precision arithmetic on the Macintosh II provides about 19 decimal places of precision. Thus, inferences using the current implementation of quickscore (or any implementation on hardware that uses the IEEE standard) are unreliable when $p(F^+, F^-)$ is less than approximately 10^{-19}. Typically, for the current QMR-DT model, such situations arise when the number of positive findings exceeds 15.

6 Conclusion

Despite its shortcomings, quickscore has been useful in the development of QMR-DT. For example, the QMR-DT group recently has experimented with several approximation algorithms for inference that are based on Monte-Carlo techniques (Shwe et al., 1990). These algorithms are suited to inference in the extensions of the QMR-DT model discussed in the previous section, but their convergence properties are not well characterized. Quickscore has provided gold standards for evaluating such convergence properties in the context of the current model. In general, quickscore is likely to be a useful tool for knowledge engineers who develop decision-theoretic expert systems.

7 Acknowledgments

References

Buchanan, B. and Shortliffe, E., editors (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, MA.

Cooper, G. (1987). Probabilistic inference using belief networks is NP-hard. Technical Report KSL-87-27, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA.

Good, I. (1961). A causal calculus (I). *British Journal of Philoshophy of Science*, 11:305–318.

Habbema, J. (1976). Models for diagnosis and detection of combinations of diseases. In de Dombal, F. and Gremy, F., editors, *Decision Making and Medical Care*, pages 399–411. North Holland, New York.

Heckerman, D. (1987). Formalizing heuristic methods for reasoning with uncertainty. Technical Report KSL-88-07, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA.

Heckerman, D. and Miller, R. (1986). Towards a better understanding of the INTERNIST-1 knowledge base. In *Proceedings of Medinfo, Washington, DC*, pages 27–31. North Holland, New York.

Henrion, M. (1989). Some practical issues in constructing belief networks. In Kanal, L., Lemmer, J., and Levitt, T., editors, *Uncertainty in Artificial Intelligence 3*, pages 161–174. North Holland, New York.

Henrion, M. (1990). Towards efficient probabilistic inference in multiply connected belief networks. In Oliver, R. and Smith, J., editors, *Influence Diagrams, Belief Nets, and Decision Analysis*, chapter 17. Wiley and Sons, New York.

Kim, J. and Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference engines. In *Proceedings 8th International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany*, pages 190–193. International Joint Conference on Artificial Intelligence.

Miller, R. (1987). Personal communication.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

Peng, Y. and Reggia, J. (1986). Plausibility of diagnostic hypotheses. In *Proceedings AAAI-86 Fifth National Conference on Artificial Intelligence, Philadelphia, PA*, pages 140–145. Morgan Kaufmann, San Mateo, CA.

Shwe, M., Middleton, B., Heckerman, D., Henrion, M., Horvitz, E., Lehmann, H., and Cooper, G. (1990). A probabilistic reformulation of the Quick Medical Reference System. Technical Report KSL-90-11, Stanford University, Stanford, CA. Submitted to the *Proceedings of the Symposium on Computer Applications in Medical Care*.

Suppes, P., editor (1970). *A Probabilistic Theory of Causality*. North Holland, New York.