MEDINFO 86, R. Salamon, B. Blum, M. Jørgensen (Editors)
Elsevier Science Publishers B.V. (North-Holland)

The Use of a Heuristic Problem-Solving Hierarchy to Facilitate the Explanation of Hypothesis-Directed Reasoning

Eric Horvitz[1], David Heckerman[1], Bharat Nathwani[2], Lawrence Fagan[1]

[1]Medical Computer Science Group, Knowledge Systems Laboratory, Stanford University School of Medicine, Stanford, CA 94305, [2]Section on Hematopathology, University of Southern California School of Medicine, Los Angeles, CA 90033 U.S.A.

We examine the problem of explaining expert system recommendations derived from complex reasoning strategies. We focus on the difficulty of justifying advice generated by hypothesis-directed reasoning. In this context, we discuss research on the PATHFINDER system for the diagnosis of lymph node pathology. We present our efforts to manage the complexity of an information optimizing reasoning strategy to facilitate explanation.

1. INTRODUCTION

Physicians have stressed the importance of understanding the reasoning of computer-based decision support systems [1]. The crucial role of explanation has made work on computer justification a focus of expert systems research. Recent explanation research has centered on the refinement of knowledge used by expert systems. Investigators are attempting to produce more meaningful explanations of computer advice by representing increasingly explicit medical knowledge [2, 3].

While, the current focus of explanation research is important, we turn our attention to a different problem with explanation. We have found that an important component of explanation difficulty is rooted in the intrinsic *naturalness* of expert system reasoning strategies. We refer to reasoning strategies that humans perceive as simple and intuitive as *natural* reasoning strategies. While unnatural strategies may at times be deemed optimal for efficient reasoning, they can be extraordinarily difficult to explain. The notion of natural versus unnatural reasoning strategies has been a concern of investigators in logic programming and automatic theorem proving research for over a decade [4]. We hope to stimulate similar concerns within the medical informatics community.

We begin by looking at the problem of unnatural reasoning strategies. We will then introduce the PATHFINDER system for lymph node pathology and describe a difficulty we had in explaining the rationale behind recommendations generated by one of the system's reasoning strategies. Finally, we discuss the development of a more natural strategy that facilitates explanation.

2. COMPLEXITY OF REASONING STRATEGIES

A variety of different reasoning strategies have been explored in artificial intelligence research. We have found it useful to characterize the naturalness of reasoning strategies along three dimensions: the nature of the fundamental inference, the complexity of the strategy, and familiarity with the high-level control knowledge [5]. In this paper, we focus on strategies that are unnatural primarily because of their complexity. The complexity of a reasoning strategy is a function of the number of objects and relations that must be considered in a single inference step. Complex strategies are difficult to understand and explain because they require the simultaneous consideration of manipulations on a great number of conceptual entities. Such strategies may demand consideration of quantities of elements and relations that burden the limitations of human working memory.

Problems associated with the comprehension of complex tasks like the operation of complex reasoning strategies have been a research focus within cognitive psychology since the field's inception [6]. Classic research in cognitive psychology has demonstrated severe limitations in the ability of humans to consider more than a handful of concepts in the short term [7]. In fact, studies have shown that humans cannot retain and reason about more than two concepts in an environment with distractions (as is often the case in medical practice) [8]. Such work has implications for the use and explanation of reasoning strategies in medical systems.

Recent research by Ben-Bassat [9] as well as by investigators in our group [10] has touched upon problems with the use of complex strategies in expert systems. Both groups are interested in the simplification of complex reasoning strategies in order to make them more understandable and explainable. Ben-Bassat has stressed the usefulness of making available several user selectable strategies. In this paper, we present the use of a heuristic problem-solving hierarchy to simplify a complex reasoning strategy. First we will present an unnatural information optimizing reasoning strategy. We will then describe how we generate a less-optimal but more explainable strategy.

*This work was supported in part by the Josiah Macy, Jr. Foundation, the Henry J. Kaiser Family Foundation, the Ford Aerospace Corporation, and the SUMEX-AIM Resource under NIH Grant RR-00785. Dr. Fagan is the recipient of a New Investigator Grant from the National Library of Medicine.

3. HYPOTHESIS-DIRECTED STRATEGIES

Hypothesis-directed strategies are used to refine the differential diagnoses in hypothetico-deductive systems. The hypothetico-deductive method (also referred to as the method of *sequential diagnosis* [11]) has been studied in several medical computer science research projects including the Acute Renal Failure [12] system, the INTERNIST-1 [13] system for diagnosis within the field of internal medicine, the MEDAS [14] system for emergency medicine, and the PATHFINDER [10] system for lymph node pathology.

Hypothetico-deductive systems are presented with an initial set of disease manifestations. The initial evidence is used to assign a probabilistic or quasi-probabilistic score to each disease. A differential diagnosis of possible disease hypotheses is formulated from the scores. Then, questions are selected which can help narrow the number of diseases under consideration. After a user answers requests for new information, a new set of hypotheses is formulated and the entire process is repeated until a diagnosis is reached.

The question selection strategies are termed hypothesis-directed in that reasoning strategies operate on the current list of hypotheses under consideration to generate test recommendations. Investigators in the INTERNIST-1 and PATHFINDER research groups have explored the usefulness of tailoring different reasoning strategies to the current differential. For example, the strategy selected to narrow the differential may depend upon the number of the diseases on the differential, the probability distribution over the differential, or both.

The advice generated by hypothesis-directed strategies is often difficult to explain because of the complexity of their operation. This is especially true if recommendations are the result of inferences based on a differential diagnosis containing a large number of diseases. Hypothesis-directed strategies may consider the relevance of hundreds of diseases in a single inference step.

4. THE PATHFINDER SYSTEM

PATHFINDER is a hypothetico-deductive expert system for the diagnosis of lymph node pathology based upon the appearance of microscopic features in lymph node tissue. Disease manifestations in lymph node pathology are microscopic *features*. Features are each subdivided into a mutually exclusive and exhaustive list of *values*. Features are evaluated by the selection of a value that reflects the status of the feature in the case being reviewed. We say that the assignment of a value to a feature constitutes a *piece of evidence*. The PATHFINDER system reasons about 80 diseases, considering over 500 pieces of evidence.

The scoring scheme employed by PATHFINDER is based upon the theory of subjective probability [15]. The subjective probabilities of experts are used to infer the probability that each disease is responsible for the evidence that has been entered into the system. Depending on the number and on the distribution of probabilities among diseases on the differential diagnosis, PATHFINDER chooses one of several alternative diagnostic strategies for selecting questions. As in other hypothesis-directed systems, it is the goal of the question selection strategies to suggest the optimal test to be evaluated next in an effort to reduce the uncertainty in the differential diagnosis.

Several PATHFINDER strategies discriminate among large numbers of diseases and features in the generation of advice. We shall not describe all of the hypothesis-directed reasoning strategies used by PATHFINDER. Rather, we explore issues surrounding the explanation of a particular PATHFINDER hypothesis-directed reasoning strategy termed *entropy-discriminate* and its descendant, *group-discriminate*.

4.1 The Entropy-discriminate Strategy

The PATHFINDER *entropy-discriminate* reasoning strategy was originally used to refine differential diagnosis disease lists ranging in size from two to eighty diseases. The strategy makes recommendations about evidence-gathering by searching for tests that maximize a measure of information contained in the differential diagnosis. Similar information-maximizing strategies have been examined in the MEDAS and Acute Renal Failure systems.

Entropy-discriminate makes use of a measure of information known as *relative-entropy*. In this context, relative entropy is a measure of the additional information provided by a piece of evidence E_i about a differential diagnosis DD. Formally,

$$H(DD,E_i) = \Sigma_j \; p(D_j|E_i) \; \log[p(D_j)/p(D_j|E_i)]$$

where $p(D_j)$ is the probability that disease D_j is present before evidence E_i is known, the *prior* probability of the disease, and $p(D_j|E_i)$ is the probability that disease D_j is present after evidence E_i is known, the *posterior* probability of the disease. For a justification of relative entropy as a measure of information gain, see [16].

Entropy-discriminate selects features which give the highest expected relative entropy

$$\langle H(DD,E_i)\rangle = \Sigma_i \; p(E_i) \; H(DD,E_i),$$

where $p(E_i)$ is calculated using the expansion rule

$$p(E_i) = \Sigma_j \; p(E_i|D_j) \; p(D_j).$$

In an information-theoretic sense, the questions selected by the entropy-discriminate strategy are optimal assuming that the goal of the pathologist is to reduce uncertainty in the differential as much as possible.

4.2 Problems with Entropy Discriminate

Soon after the implementation of entropy-discriminate mode, we discovered that several expert pathologists, including the expert that provided the system's knowledge, often found that selected questions were difficult to understand when the differential contained more than approximately ten diseases. The entropy-discriminate strategy of selecting questions that best discriminate among *all* diseases on a differential diagnosis often seemed to be too complex for experts. This is not surprising in light of the limitations of human short term memory discussed above.

We also had problems explaining the recommendations of entropy-discriminate whenever there were more than two diseases on the differential. Attempts were made to provide textual and graphical explanations for the powerful strategy's recommendations. One such graphical explanation justified questions by listing, for each disease, the feature value that would most favor the disease. Physicians found such complex summarizations to be difficult to understand.

4.3 Simplifying the Problem-Solving Task

The observed problems with the entropy-discriminate strategy stimulated our interest in strategies for simplifying and explaining hypothesis-directed reasoning. We discovered that pathologists often manage the complexity of the diagnostic problem-solving task by reasoning about a very small number of disease categories or groups at any one time. Questions that discriminate among natural groups tend to be proposed.

Specifically, the chief expert pathologist on the PATHFINDER team often imposes a simple two-group discrimination structure on the problem-solving task. As opposed to a strategy of discriminating among all the diseases on the differential, the pathologist's discrimination task at any point in reasoning about a case is constrained to only two groups of diseases. As categories of disease are ruled out, the particular pairs of groups considered become increasingly specific. For example, if there are benign and malignant diseases on a differential diagnosis, the pathology expert often deems most appropriate those questions that best discriminate between the benign and malignant groups rather than questions that might best discriminate among all of the diseases. If all benign diseases have been rule-out, leaving only primary malignancies and metastatic diseases on the differential diagnosis, the pathologist will attempt to discriminate between the primary malignancy and the metastatic categories.

We found that the expert's diagnostic strategy can be described by the traversal of a hierarchy of disease categories. The problem-solving hierarchy (see Fig. 1), is a binary tree of disease groups. The hierarchy can be used to group the differential diagnosis at various levels of refinement.

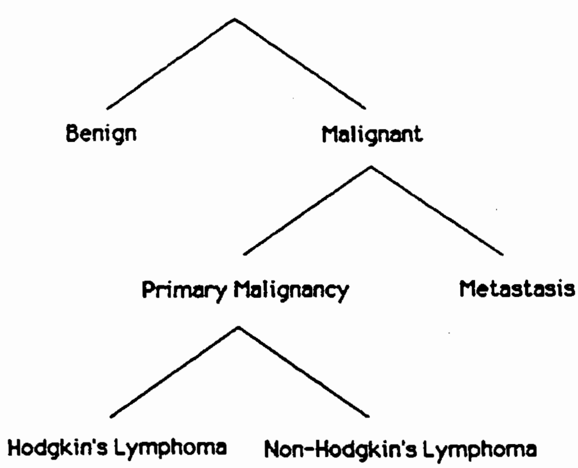

Figure 1: Heuristic problem-solving hierarchy

It is interesting to note that several previous studies of medical reasoning have identified similar problem-solving hierarchies [17]. Indeed, physicians in many areas of medicine may use and pass down historically-evolved heuristic strategies for managing the complexity of medical decision making.

The discovery of this expert reasoning strategy in lymph node pathology suggested the development of a new question-selection strategy that could discriminate among binary groups of diseases instead of individual diseases. It was hoped that design and application of such a strategy would make explanation clear, as the user would only have to consider the relevance of a recommendation to two groups.

4.4 The Group-discriminate Strategy

Our attempt to naturally constrain the discriminatory focus of the entropy-discriminate strategy led to a new reasoning strategy we named ***group-discriminate***. The group-discriminate strategy selects questions based on their ability to discriminate between the most specific pair of disease categories that account for all diseases on the differential.

For a given differential diagnosis, group-discriminate identifies the most specific grouping possible and then selects questions that best discriminate among groups of diseases. More formally, suppose the differential is split into two groups, G_1 and G_2, of n_1 and n_2 diseases respectively:

$$G_1 = \{D_{11}, D_{12}, \ldots D_{1n_1}\}$$

$$G_2 = \{D_{21}, D_{22}, \ldots D_{2n_2}\}$$

As we assume that only one lymph node disease is present in PATHFINDER, we can consider the diseases to be mutually exclusive events. We are interested in the probability that the true diagnosis will be in each group. To calculate this probability we add the probabilities of all the diseases within each group. That is, the probability that a group contains the true diagnosis is

$$p(G_j) = \Sigma_k \, p(D_{jk}), \quad j = 1, 2.$$

We can also calculate $p(G_j|E_i)$, the probability of the final diagnosis being contained in a group, considering a new piece of evidence E_i. This is

$$p(G_j|E_i) = \Sigma_k \, p(D_{jk}|E_i), \quad j = 1 \text{ or } 2.$$

Therefore, a relative entropy of the grouped differential can be defined. In particular,

$$H_G(DD, E_i) = \Sigma_j \, p(G_j|E_i) \log[p(G_j|E_i)/p(G_j)]$$

This quantity represents the additional information contained in E_i about the grouped differential diagnosis. Group-discriminate selects those features which give the highest expected relative entropy.

Notice that the group-discriminate strategy ignores information concerning the probabilities of diseases within each group. Only the probabilities that the true diagnosis lies within a group is considered in the calculations.

5. DISCUSSION

We integrated the group-discriminate strategy into the PATHFINDER system so that it continues to refine differential diagnosis lists until all diseases remaining on the differential diagnosis are in a category at one of the leaves of the binary problem-solving tree. At this point, other hypothesis-directed strategies are applied to continue pursuing a diagnosis. As the group-discriminate reasoning strategy has a simpler discriminatory focus and more closely follows the decision making protocol of the expert lymph node pathologist than entropy-discriminate, it is quite easy to explain.

Instead of having to present complex summaries explaining how each piece of evidence might impact on a number of diseases, an explanation of questions generated by group-discriminate must simply demonstrate how possible responses affect the two groups under consideration.

The PATHFINDER system justifies the usefulness of questions selected by group-discriminate with a graphical display. Fig. 2 presents a small portion of a PATHFINDER consultation. At the top of the figure is the differential diagnosis, grouped into benign and malignant categories at the current level of refinement. Below, several lymph node features recommended by group-discriminate are listed. The group-discriminate strategy has determined that these features can best discriminate between the benign and malignant diseases. In this case, the user requested explanation for the *follicles density* recommendation.

The positions of a set of asterisks in the justification graph at the bottom of the figure is used to indicate the degree to which each group of diseases is favored by each possible feature value. Specifically, the position of the asterisk is a function of the likelihood ratio $p(E_i|G_1)/p(E_i|G_2)$. In the example below, the values *separated* and *far apart* strongly support diseases on the differential diagnosis that are in the benign group, while the values *back-to-back* and *closely packed* strongly support the malignant disease hypotheses.

A user can easily ascertain how a question discriminates among two groups of diseases; evidence is either supportive for one group or the other. Even in an environment filled with distractions, the behavior of the strategy is adequately explained by such simple graphs.

Unfortunately, the more explainable group reasoning strategy has some disadvantages. A problem with the use of group-discriminate is that the differential diagnosis refinement process does not always proceed as quickly as it does with the application of the optimal entropy-discriminate. That is, group-discriminate is not as efficient as the more powerful entropy-discriminate; on average, a larger number of evidence-gathering requests will be made by group-discriminate to achieve a similarly refined differential diagnosis. This must be the case as detailed information about the plausibility of individual diseases within each group is discarded in the grouping process.

In general, simplifying a formerly optimal strategy will lead to a less-efficient strategy. In the context of cognitive psychology research on human limitations, it is not unexpected that a reasoning strategy derived through the constraint or decomposition of a complex problem-solving task may be easier to understand and explain. It seems that for a wide variety of reasoning strategies, there will frequently be an inverse relationship between the understandability and efficiency of complex reasoning strategies. In making decisions about alternative reasoning strategies for expert systems, computer scientists may have to face an explainability/efficiency trade-off.

6. SUMMARY

We have identified the difficulty of explaining advice generated by complex reasoning strategies in medical expert systems. We have found that strategies that are deemed optimal for information-gathering may be sub-optimal for assisting physicians. We then presented an example of a natural simplification of a complex hypothesis-directed strategy. It is clear that constraining the discriminatory focus of the strategy with a natural problem-solving hierarchy facilitated the explanation of system recommendations.

The problem of explaining complex reasoning strategies to users that may have little background in machine inference strategies is an important area of investigation warranting further attention. We hope that research on the use and modification of unnatural reasoning strategies in medicine will complement the concurrent explanation research on the refinement of knowledge.

```
> ask

Discriminating:

 Malignant

    Small cleaved, follicular lymphoma
    Mixed, small cleaved and large cell,
         follicular lymphoma
    Large cell, follicular lymphoma
    Kaposis sarcoma
    Small noncleaved, follicular lymphoma

 Benign

    Florid reactive follicular hyperplasia
    Reactive hyperplasia
    AIDS

I recommend that the following
features be evaluated:

  Status of follicles
  Follicles density
  Subcapsular sinuses
  Medullary sinuses
  Comparison of cytology inside and
        outside the follicles

> justify

 Which feature do you want justified?

> follicles density

The following table elucidates the
discriminating power of this feature.
The position of the asterisk indicates
which of the two groups of diseases is
favored by each value.

 Malignant
         Benign
 |       |
 v       v
 *........ back-to-back
 *........ closely packed
 .......*. separated
 ........* far apart
```

Figure 2: Portion of PATHFINDER consultation

Typing *ask* invokes the differential refinement strategy appropriate to the current differential. Group-discriminate is invoked because the current differential contains both benign and malignant diseases. User responses are italicized.

ACKNOWLEDGEMENTS

We thank Moshe Ben-Bassat, Benjamin Grosof, Curt Langlotz, and Peter Szolovits for interesting conversations about unnatural reasoning strategies. We thank Costan Berard for discussing with us his insights about reasoning in pathology. We finally thank Ted Shortliffe for his advice and enthusiastic support.

REFERENCES

[1] Teach, R. L. and Shortliffe, E. H., An Analysis of Physician Attitudes Regarding Computer-Based Clinical Consultation Systems, *Computers and Biomedical Research*, 14 (1981), pp. 542-558.

[2] Swartout, W. R., Explaining and Justifying Expert Consulting Programs, *In* "Proceedings of the Seventh International Joint Conference on Artificial Intelligence", International Joint Conference on Artificial Intelligence, Vancouver, British Columbia, August 1981, pp. 815-822.

[3] R.S. Patil, "Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis," PhD dissertation, M.I.T., 1981.

[4] W.W. Bledsoe, Non-resolution theorem proving, *Artificial Intelligence*, 9 (1977), pp. 1-35.

[5] E. Horvitz, D. Heckerman, "Unnatural reasoning strategies and explanation in expert systems", Knowledge Systems Lab Working Paper, Stanford University, 1986.

[6] J.S. Bruner, J.J. Goodnow, G.A. Austin, "A study of thinking", Wiley, 1956.

[7] G.A. Miller, The magical number seven, plus or minus two, *Psychological Review*, 63 (1956), pp. 81-97.

[8] N.C. Waugh, D.A. Norman, Primary Memory, *Psychological Review*, 72 (1965), pp. 89-104.

[9] M. Ben-Bassat, D. Teeni, Human-oriented information acquistion in sequential pattern classification: Part 1 - single membership classification, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-14 (1), (1984), pp. 131-138.

[10] Horvitz, E. J., Heckerman D. E., Nathwani, B. N., and Fagan, L. M., Diagnostic Strategies in the Hypothesis-Directed PATHFINDER System, *In* "Proceedings of the First Conference on Artificial Intelligence Applications", Denver, CO, December 1984, pp. 8.

[11] Gorry, G. A., and Barnett, G. O., Experience with a model of sequential diagnosis, *Computers and Biomedical Research*, 1 (1968), pp. 490-507.

[12] Gorry, G. A., Kassirer, J. P., Essig, A., and Schwartz, W. B., Decision analysis as the basis for computer-aided management of acute renal failure, *AJM*, 55 (1973), pp. 473-484.

[13] Miller, R. A., Pople, H. E., and Myers, J. D., INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine, *New England Journal of Medicine*, 307 (8), (1982), pp. 468-476.

[14] Ben-Bassat, M., et. al., Pattern-based Interactive Diagnosis of Multiple Disorders: the MEDAS System, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2 (1980), pp. 148-160.

[15] de Finetti, B., "Theory of Probability", Wiley, New York, 1970.

[16] Shore, J.E., Relative entropy, probabilisitic inference, and AI, *In* "Uncertainty in Artificial Intelligence", North Holland, AAAI/IEEE, 1986.

[17] Elstein, A. S., Shulman, L. S., and Sprafka, S. A., "Medical problem solving: An analysis of clinical reasoning", Harvard University Press, Cambridge, Mass., 1978.