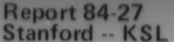

Report 84-27
Stanford -- KSL

Explanation Capabilities for Medical Consultation Systems (Tutorial).

Edward H. Shortliffe,

May 1984

Heuristic Programming Project
Report No. HPP 84-27

May 1984

Explanation Capabilities for Medical Consultation Systems (Tutorial)

Edward H. Shortliffe
Stanford University, Medical School

Published in the Proceedings of AAMSI Congress '84, San Francisco, May 21-23, 1984.
Reprinted by permission of the American Association for Medical Systems and Informatics (AAMSI).

EXPLANATION CAPABILITIES FOR MEDICAL CONSULTATION SYSTEMS (Tutorial)

Edward H. Shortliffe

Heuristic Programming Project
Departments of Medicine and Computer Science
Stanford University, Stanford, California 94305

The role of explanation capabilities in enhancing the transparency and acceptance of medical advice systems has become increasingly recognized since the early 1970's. The tutorial presentation will review the state of the art in computer-generated explanations, survey several representative systems, and suggest directions for future research on the problem. This paper provides an overview of the issues and includes a bibliography of several key pieces of work in the area.

Introduction

Researchers in the development of medical expert systems have increasingly recognized the importance of explanation capabilities in encouraging the acceptance of their programs. One survey of potential users of medical advice systems has suggested that explanation may be the single most important capability of an acceptable clinical decision tool (16). Good explanations serve four functions in a consultation system: [1] they provide a method for examining the program's reasoning if errors arise when the system is being built; [2] they assure users that the reasoning is logical, thereby increasing user acceptance of the system; [3] they may persuade users that unexpected advice is appropriate; and [4] they can educate users in areas where their knowledge may be weak. These diverse roles impose several requirements upon the system. For example, the explanations must adequately represent the reasoning processes of the program, and they should allow the user to examine the reasoning history or underlying knowledge at various levels of detail. In addition, although the program's approach to a problem need not be identical to the expert's approach, its overall strategy and reasoning steps must be understandable and seem logical, regardless of the user's level of expertise. This means that the system must ideally have the capability to tailor its explanations to the varying needs and characteristics of its users.

Although a few experimental consultation systems have incorporated explanation capabilities, none has the full range of functionality suggested in the preceding paragraph. Much basic research in knowledge representation, causal reasoning, user modeling, deep structures, and the human-computer interface will be required before "ideal" explanation capabilities are possible. I will return to some of these issues, but will first review the current capabilities of medical advice systems and mention some of the key research efforts to date.

Review of Past Work

Early advice systems based on Bayes' Theorem and other statistical techniques generally made no effort to justify their diagnostic conclusions. The programs simply offered probabilistic statements about the most likely disease(s). They could defend the analysis only by pointing to Bayes' formula and to the conditional probabilities and prevalence data used in the calculation.

When algorithmic decision logic was used in systems, however, it was possible to associate "canned text" explanations with decision points and to print them out as part of the final result. Perhaps the best early example of this approach was Bleich's well-known program to advise on the management of acid-base and electrolyte disorders (1). The output from a consultation with that system included a paragraph or two describing the details of the case and the rationale for a specific management approach. However, the system did not permit user queries about specific points of confusion, nor could it defend the structure of the underlying algorithm that guided its advice generation.

I believe it was Gorry who first wrote about the need for explanation capabilities in medical advice systems (5). He had been working on a program that used decision analytic techniques for the assessment of renal disease and became convinced that one of the reasons for its limited appeal to clinicians was the inability for them to gain insight into the basis for the program's recommendations.

The explanation system of the MYCIN program provided dynamic generation of explanations from its inference rules -- the same structures that were used to generate advice (12). This program assisted in the selection of antimicrobial therapy for patients with bacteremia or meningitis (13), and it was able to answer questions about how it reached a particular conclusion (i.e., what rules led to the pertinent inference) and about why it had asked a particular question (i.e., which rules could use the requested information). The capability was available both for a specific run of the program or for general querying of its knowledge base. Although the program's responses provided an accurate description of a portion of MYCIN's reasoning, to understand the overall reasoning scheme a user needed to request a display of all the rules that were used. In addition, the rules were generally designed for efficiency and therefore frequently omitted underlying causal mechanisms that are sometimes needed for effective explanations, especially when the user is a novice. Finally, MYCIN's explanations were customized to neither the questioner's objectives nor characteristics.

MYCIN's explanation capabilities were expanded by Clancey in his work on the tutorial system named GUIDON (3). In order to use MYCIN's knowledge base and patient cases for tutorial purposes, Clancey found it necessary to incorporate knowledge about teaching. This knowledge, expressed as "tutorial rules", enhanced the ability of a student to learn efficiently from MYCIN's knowledge base. Clancey also noted problems arising from the frequent lack of underlying mechanistic or "support" knowledge which is needed to explain the relevance and utility of a domain rule (4).

Swartout developed a system that generated explanations from a record of the development decisions made during the writing of a consultation program to advise on digitalis dosing (15). The domain expert provided information to a "Writer" subprogram, which in turn constructed the advising system. The traces left by the writer, a set of domain principles, and a domain model were then used to produce explanations. Thus both the knowledge acquisition process and automated programming techniques were intrinsic to the explanations generated by Swartout's system. Responses to questions were customized for different kinds of users by keeping track of what class of user is likely to be interested in a given piece of code.

Whereas MYCIN generated explanations that were usually based on a single rule, Weiner described a system named BLAH that could summarize an entire reasoning chain in a single explanatory statement (18). The approach developed for BLAH was based on a series of psycholinguistic studies that analyzed the ways in which human beings explain decisions, choices, and plans to one another (8). For example, BLAH structures an explanation so that the differences between alternatives are given before the similarities (a practice that was noted during the analysis of human explanations).

The tasks of interpreting questions and generating explanations are confounded by the problems inherent in natural language understanding and text generation (11). A consultation program must be able to distinguish general questions from case-specific ones, and questions relating to specific reasoning steps from those involving the overall reasoning strategy (6). As previously mentioned, it is also important to tailor the explanation to the user, giving appropriate supporting causal and empiric relationships. For a system to produce customized explanations, it must be able to model the user's knowledge and motivation for using the system. Wallis described one approach to achieving this behavior in a system that generated customized explanations from detailed causal reasoning chains (17). The role of layered causal domain models, and the differences in explanations that can be generated at varying levels of detail, have also been explored in the ABEL work of Patil (10).

There has been related recent work on the subject of critiquing systems, variants of consultation programs for which explanation is an inherent issue. A critiquing system does not routinely offer advice but, rather, assesses the user's patient assessment or management plan. The ATTENDING system of Miller was perhaps the first large program of this type (9), although Clancey had earlier incorporated related capabilities into MYCIN's therapy selection routines (2). Miller's program assesses an anesthesiologist's management plan for a patient, using a variant of augmented transition networks (which he calls a PROSENET) to generate paragraphs that weigh the risks and benefits of the user's plan and potential alternative approaches. Langlotz has also described a critiquing program and a technique known as hierarchical plan analysis (7). His system provides an alternate mode of use for the cancer chemotherapy advice system known as ONCOCIN (14).

An Organizational Structure For Future Work

One lesson of the last decade's work on explanation is the observation that a variety of knowledge forms are necessary if an intelligent system is to customize its explanations to the individual who is using the program. Underlying structural and causal relationships are generally required in

addition to the high level judgmental "rules" or associations that characterized the knowledge bases of early medical expert systems.

The problems of explanation generation can be divided into at least four categories:

1. Modeling the User's Knowledge:

   GUIDON and other intelligent CAI systems have recognized the need to keep an internal model of the student, i.e., what he has shown he knows, what you already told him, and perhaps a record of where his greatest weaknesses lie. Similarly, it is clear that an expert human consultant customizes his explanations so that they can be understood by the person requesting consultation (and are thereby maximally convincing). The expert starts with certain suppositions about his client's knowledge (e.g., a teacher may presume his student is starting from scratch, but a cardiologist will assume that another physician requesting advice probably already knows a fair amount of cardiology). The default presumption is modulated, however, as the interaction proceeds and the client demonstrates his strengths or weaknesses.

2. Selecting A Response Strategy:

   Most explanation efforts have tended to be simple reiterations of individual reasoning steps, but it is clear that experts and teachers use several alternate strategies for conveying their ideas or key facts. Many of these techniques draw upon common sense world knowledge (e.g., analogies with familiar concepts outside the domain), but advice systems have thus far failed to capitalize on these teaching strategies in their explanation capabilities. Thus another goal of the work that lies ahead should be to develop structures for drawing parallels or otherwise representing the strategies used by good "explainers".

3. Modeling Contextual Information Regarding the Interaction:

   I have already mentioned some of the ways in which contextual information may be useful in determining the best way to answer a question. For example, a more accurate model of the user's knowledge can be developed over time, and the extent to which a given conversation is focused on a particular local topic can be assessed. Note that this emphasizes issues other than those related to natural language understanding; computational linguists also often cite the need to record contextual dialogue information in order to handle problems such as anaphora. An understanding of the "flow" of a dialog is also important in understanding the meaning of subsequent questions.

4. Understanding The Question:

   This issue interfaces with the problem of natural language understanding, but can be viewed in a somewhat different light if we emphasize instead the ways in which the model of the user and contextual information may allow us to disambiguate questions. To draw from a medical example, consider the following scenario. A reasoning program for pharyngitis diagnosis and management has just diagnosed strep throat and recommended penicillin and the user asks the question "why would you give penicillin?" In the most obvious case, one might imagine a response that itemizes the risks of streptococcal infections and the reasons for treating early with penicillin. Similarly, one might expect a more detailed response for a student and a quick summary for a physician using the system.

   However, an alternate interpretation is that *every* physician knows the theoretical reasons for giving penicillin in strep pharyngitis, and that if the user is a physician and is asking the question then he must be asking something different from the simple informational question. In this case the query might be interpreted as a challenge (one that might have been conveyed by tone of voice if it had been asked of a human consultant). Apparently the user has reason to doubt that penicillin was the appropriate agent in this case, or thinks that no drug was required. Other background information and contextual knowledge should also help, and an intelligent program might thereby answer the question in a given case in any of the following ways:

- Because the patient has pre-existing rheumatic heart disease.
- Because I doubt that he is allergic to penicillin, even though he reported that he is.
- Because I tend to treat conservatively and give penicillin for strep throat even though I know there hasn't been a case of rheumatic heart disease in California in over 10 years.

The ideal intelligent assistant should be able to determine from knowledge of the user, the domain, the individual case, and the context of the dialog, which of the preceding responses is most appropriate. It will be a challenge to develop expert advice systems with explanatory capabilities that can even begin to approach these goals.

Conclusion

In addition to the knowledge representation and problem solving issues I have emphasized in this discussion, the explanation capabilities of future medical advice systems are likely to reflect ongoing technological advances. Particularly relevant in this regard are the graphics capabilities of modern computer terminals and workstations, as well as alternate interactive devices such as touch screens and mice. Many of the explanations offered by the systems of tomorrow will be graphical in nature, or draw directly on pictorial material retrieved from associated storage devices such as video discs. It should accordingly be clear that there is ample opportunity for a wide range of innovative research activities in the area of explanatory capabilities for medical advice systems.

Acknowledgment

Portions of this paper were adapted from an article written by the author in collaboration with Dr. Jerry Wallis (17). Dr. Shortliffe is recipient of award LM00048 from the National Library of Medicine and is a Henry J. Kaiser Family Foundation Faculty Scholar in General Internal Medicine.

References

1. Bleich, H.L. Computer-based consultation: electrolyte and acid-base disorders. Amer. J. Med. 53;1982:285-291.

2. Clancey, W.J. An antibiotic therapy selector which provides for explanations. Proceedings of the 5th IJCAI, American Association for Artificial Intelligence, 1977, pp. 858. (To appear as Chapter 6 in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B.G. Buchanan and E.H. Shortliffe, eds., Reading, MA: Addison-Wesley, 1984.)

3. Clancey, W.J. Tutoring rules for guiding a case method dialogue. Int. J. Man-Machine Studies 11;1979:25-49.

4. Clancey, W.J. The epistemology of a rule-based expert system: a framework for explanation. Artificial Intelligence 20;1983:215-251.

5. Gorry, G.A. Computer-assisted clinical decision making. Meth. Info. Med. 12;1973:45-51. (To appear as Chapter 2 in *Readings in Medical Artificial Intelligence*, W.J. Clancey and E.H. Shortliffe, eds., Reading, MA: Addison-Wesley, 1984.)

6. Hasling, D.W., Clancey, W.J., and Rennels, G. Strategic explanations for a diagnostic consultation system. Int. J. Man-Machine Studies 20;1984. (in press)

7. Langlotz, C.P. and Shortliffe, E.H. Adapting a consultation system to critique user plans. Int. J. Man-Machine Studies 19;1983:479-496.

8. Linde, C. and Goguen, J.A. Structure of planning discourse. J. Soc. Biol. Struct. 1;1978:219-251.

9. Miller, P.L. ATTENDING: critiquing a physician's management plan. IEEE Trans. Patt. Anal. & Mach. Intell. 5(5):1983:449-461.

10. Patil, R.S., Szolovits, P., and Schwartz, W.B. Causal understanding of patient illness in medical diagnosis. Proceedings of the 7th IJCAI, American Association for Artificial Intelligence, 1981, pp. 893-899.

11. Pollack, M.E., Hirschberg, J., and Webber, B.L. User participation in the reasoning processes of expert systems. Tech. Rept. MS CIS-82-9, Department of Computer and Information Science, The Moore School, University of Pennsylvania, July, 1982.

12. Scott, A.C., Clancey, W.J., Davis, R., and Shortliffe, E.H. Explanation capabilities of knowledge-based production systems. Amer. J. Comp. Ling. Microfiche 62;1977. (Also appears as Chapter 18 in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, Reading, MA:Addison-Wesley, 1984.)

13. Shortliffe, E.H.. Computer-Based Medical Consultations: MYCIN. New York: Elsevier/ North Holland, 1976.

14. Shortliffe, E.H., Scott, A.C., Bischoff, M.B., et al. ONCOCIN: an expert system for oncology protocol management. Proceedings of the 7th IJCAI, American Association for Artificial Intelligence, 1981, pp. 876-881.

15. Swartout, W.R. *Producing Explanations and Justifications of Expert Consulting Programs*. Ph.D. Th., Massachusetts Institute of Technology, 1981. Report no. LCS-TR-251

16. Teach, R.L. and Shortliffe, E.H. An analysis of physician attitudes regarding computer-based clinical consultation systems. Comput. Biomed. Res. 14;1981:542-558.

17. Wallis, J.W. and Shortliffe, E.H. Explanatory power for medical expert systems: studies in the representation of causal relationships for clinical consultations. Meth. Info. Med. 21;1982:127-136.

18. Weiner, J.L. BLAH: a system which explains its reasoning. Artificial Intelligence 15;1980:19-48.

STRATEGIES FOR CHOOSING THE NEXT TEST IN AN EXPERT SYSTEM

Ernest Chang MD PhD, Michael McNeely* MD FRCP(C) Ken Gamble BSc

Laboratory for Computer Enhanced Cognition
University of Victoria

*Island Medical Laboratories

The LIVER expert system incorporates techniques from artificial intelligence for the interpretation of laboratory tests in the area of liver disease. It is rule based, uses certainty factors, and implements a sequential strategy for diagnosis. This is done through forward chaining and selecting a best next test(s). A number of metrics for "best" are described, which lead to different selection methods. Two of them have been implemented, and compared against the recommendations for next tests in cases generated by a human expert.

Introduction

Among the applications of techniques in artificial intelligence to medicine [Pople 1981, Shortliffe 1976, Szolovits 1982] are a class of computer programs called expert systems. Briefly, these programs contain the rules of thumb (heuristics) that represent the expert's knowledge in a particular domain; given these rules, simple deductive logic is used to make inferences about facts in the given area. At their best, they perform at a level equal to that of many experts [Feignenbaum 1983]. MYCIN [Shortliffe 1976] is an example of an expert system in antibiotic therapy that has been widely acknowledged to be successful in its clinical performance. Another well-known system is INTERNIST [Pople 1981], which functions in the entire domain of internal medicine, and is capable of making diagnoses of simultaneously occurring diseases.

There are four major advantages to the use of expert systems in computer-assisted medicine. First, heuristics capture the special knowledge and rules of thumb that experts possess, which cannot necessarily be deduced from a simple knowledge of the domain and its mechanisms. This is especially important in medicine, in which knowledge is often uncertain and incomplete [Szolovits 1982]. Secondly, they employ reasoning mechanisms that are explicit and demonstrable. Thus, explanations of the rules and facts used in arriving at conclusions can be provided [Davis 1982], in contrast to statistical methods of inference. Thirdly, expert systems separate the knowledge base from the deductive mechanisms and inference strategies [Davis 1977]. Such systems are much easier to keep current and maintain than comparable programs that include domain-dependent facts with the program logic. Lastly, expert systems can deal with incomplete knowledge, and make inferences based only on what is currently known [Davis 1977], which is again difficult to achieve with techniques like statistical analysis or pattern matching.

Of course, there are a number of problems that are still outstanding in the implementation of good expert systems. The most important of these deal with knowledge acquisition and inference strategy [Davis 1982]. The first is the problem of transferring the expert's knowledge into a database that is consistent and valid, even though experts may not always be able to articulate the processes by which they arrive at certain conclusions. The second is the question of what facts should be considered first, and given a partial set of facts, what next fact should be sought. The research described in this paper deals with this last problem.

Copyright © 1985 by KSL and
Comtex Scientific Corporation

FILMED FROM BEST AVAILABLE COPY