

CONTROLLING QUESTION ASKING IN A MEDICAL EXPERT SYSTEM

Ruven Brooks
Department of Psychiatry and
Behavioral Sciences
University of Texas Medical Branch
Galveston, Texas 77550 U.S.A.

Jon Heiser
Department of Psychiatry and
Behavioral Sciences
University of Texas Medical Branch
Galveston, Texas 77550 U.S.A.

We have discovered what we believe is an essential design problem for medical expert systems, that of controlling the amount and the type of information which the system requests from the user. This problem is inherent in medical expert systems because of the nature of the distribution of clinical states and the nature of the training and background of physicians. The problem also exists for human consultants, and a complete and general solution for computer systems is probably not achievable. However, several techniques show promise for reducing the magnitude of the problem in various clinical domains.

1.0 Introduction

Over the past two years, we have been engaged in the construction of a computer system to advise primary care physicians and specialists in other areas about the use of psychiatric medications. In the course of our first attempts at building the system, we showed it to a few colleagues and discovered, to our surprise, that some of them regarded responding to 30 questions with one-word answers an overly long interaction and the contemplated 100 item sessions clearly unacceptable. Although a portion of the problem was clumsy use of software on our part, we believe that carefully selecting and controlling the amount and type of information requested from the user is a major aspect of the design of an expert system.

2.0 The Nature of the Problem

Computer based medical expert systems are designed to provide information and advice to physicians or other medical personnel. The information and advice can be on such topics as diagnosis, test interpretation, medication levels, and patient management. All of these systems request information from the user, ranging from yes-no questions to open-ended prompts for free text.

We believe that limiting the number and content of these requests must be a major, primary design goal in the construction of expert systems. In particular, for optimum user acceptance, two conditions must hold:

1. The number of requests must be the minimum that is consistent with maximum utilization of medical inference.

2. The content of the requests must be perceived by the user as important and relevant to the task of the system.

We contend that these two conditions grow out of two characteristics of medical knowledge and medical expertise (though they probably hold for other fields as well).

2.1 Skewed Distribution of Symptoms

The first of these characteristics is that while the space of symptoms, syndromes, and diseases is very large, the distribution of occurrence in clinical populations is markedly skewed; a very few symptoms and diseases account for a large proportion of patients.

This skewed distribution has a profound impact on the way medical experts convey and express knowledge. One of the more subtle, but significant, impacts is on medical terminology. Medical terms are frequently organized in a hierarchical fashion. Terms near the bottom of the hierarchy refer to single signs or symptoms while those further up the hierarchy refer to larger clusters, often with increased etiologic or therapeutic implications. This terminological structure often allows complex medical states to be described concisely, with a few well chosen words. This conciseness, in turn, greatly facilitates medical communication.

A second impact of the skewed distribution of symptom occurrence is that information is often conveyed as much by omission as by statement. For example, the lack of any comment regarding delusions or hallucinations in a brief note by an experienced psychiatrist means that, during the course of the evaluation, no delusions or hallucinations were reported or observed; it does not mean that the psychiatrist did not check for them.

A priori, the terse nature of medical communication and the domain sophistication of the user ought to facilitate the building of expert computer systems. The use of a well-defined terminology, together with the fact that the user understands the underlying medical principles, implies that very simple forms of natural language processing can be used without obviously restricting the range of user responses. To some extent, this expectation has been realized; the MYCIN system [1]. for example, asks open-ended questions such as "What is the Gram stain of Organism-1?" without analyzing the input, other than to check the response against a list of legal values.

Problems may arise, however, from several sources: One is, of course, inappropriate or incomplete representation of the medical knowledge within the system. Systems which ask about the current mood of comatose patients are not likely to be well accepted by users.

Other problems in representing medical knowledge are more subtle. An example from our system involves asking about all medications that a patient received during all prior psychiatric episodes; this is a useful piece of information in selecting a current medication. Unfortunately, when questions were asked for each and every prior episode of psychiatric illness, a user was frequently repeatedly asked about medications given years ago, at other hospitals, concerning which little or nothing was known, and the procedure quickly became more irritating than informative.

2.2 Variance in Physician Knowledge

The second major characteristic of medical knowledge and expertise is that a given physician's level of sophistication may vary greatly depending on the area of medicine. Thus, a second major source of difficulty lies in mismatches between the level of sophistication of the user and that assumed by the system. For example, a system may assume that the user is unable to supply a particular fact or conclusion, and it may, therefore, ask a

series of questions to deduce the conclusion. If the user can, in fact, directly supply the needed information, then he or she may become irritated at how "dumb" the system seems. A mismatch in the opposite direction is equally problematic; if the system prompts for conclusions or inferences which a user is not knowledgeable enough to supply, the reliability and validity of the system may be considerably reduced.

The mismatch problem is not merely relative to the user's knowledge and experience, but is also relative to the user's perception of what the system is doing. For example, a system which has received inconsistent or doubtful input may, quite appropriately, attempt to deduce some fact or conclusion that a user has already supplied. If the user is unaware of why this is being done, the system's "second guessing" may elicit an unpleasant reaction.

3.0 Solution Strategies

While these problems have been discussed in the context of computer systems, similar problems occur with human consultants. Furthermore, the clinical problems, diagnostic techniques, therapeutic measures and language processes used to communicate these phenomena are so different in different areas of medical practice that it is unlikely that these problems will be solved universally and quickly by any single technique or device. Almost certainly, different combinations of techniques will have to be carefully engineered for each clinical application.

As yet, the repertoire of such techniques is small, and there are no guidelines for identifying when particular techniques should be used. Currently, three techniques show some promise as components of such combinations.

3.1 Guidance from Models of the Domain

One device which can be particularly helpful in eliminating unnecessary questions is a model of the dynamics of a medical state. These models are representations of physiological or biological mechanisms and tie together chains of medical events. Several medical systems have been built which rely on such models [2].

As helpful as dynamic models are, they are currently applicable in only a small number of medical systems. In many medical areas, underlying mechanisms are not yet sufficiently well understood to be useful. An alternative which is more widely applicable is the use of

static models of the relationship between observed signs and underlying disease states. [3,4].

The way in which such static models act to reduce the system's requests for

While static models are, they do not provide an absolute guarantee of user acceptance. If the user's knowledge of the domain is too different from that of the system's, he or she might not recognize the critical nature of some of the information being requested and, therefore, interpret the questions as superfluous.

3.3 User Control over Sophistication Level

While both dynamic and static models act to reduce requests for unnecessary information, they may have little or no impact on difficulties caused by a mismatch in knowledge levels between the user and the system. One approach to this problem is for the system to assume a high level of sophistication on the user's part but to provide the user with ways to indicate that a lower level is appropriate. One such mechanism that we are building into the Psychopharmacology Advisor is to permit users to respond with "LEVEL" to a question which she or he feels is at too high a level. The system responds by breaking the question down into several, simpler questions. For example, answering "LEVEL" to a question about complete auditory hallucinations will cause the Advisor to ask a series of questions about whether the patient has auditory hallucinations, what their form is, and what the content of the hallucinations is.

We believe that this device is a useful addition to conventional "help" system displays of documentation and user options. It involves less user reading and interpretation and, therefore, should be less error prone.

3.4 Supplying Rationales

As noted earlier, in determining user acceptance, the perceived rationale for an information request is often as important as the contents of the request itself. Rather than relying on the user to guess the rationale correctly, it may prove more reliable to have the expert system make its rationale available to the user, either automatically or by request. An example of how this may be done is the explanation capabilities of the rule-based EMYCIN software (Shortliffe, 1976) which we are using as the basis for the Psychopharmacology

Advisor. In response to a question from the system, the user may, instead of giving the answer, give commands to obtain information about why a particular question was asked or how a particular result was achieved.

4.0 Conclusion

We believe that controlling the amount and type of information that a medical expert system requests from a user is an essential design problem. This problem is inherent in medical expert systems because of the natures of the distribution of clinical states and of the training and background of physicians. The problem is also present in interaction with human consultants, and a complete solution for computer systems is probably not immediately achievable. However, several techniques show promise for reducing the magnitude of the problem. These include the use of dynamic and static domain models, user control over sophistication level, and user access to the rationales behind information requests.

REFERENCES

- [1] Shortliffe, E. H. Computer Based Medical Consultations: MYCIN. New York: American Elsevier, 1976.
- [2] Gorry, G. A., Silverman, H. & Pauker, S. G. Capturing clinical expertise: A computer program that considers clinical responses to digitalis. American Journal of Medicine, 64(452), 1978.
- [3] Pople, H. E. The formation of composite hypotheses in diagnostic problem solving: An exercise in synthetic reasoning. Proceedings Fifth International Joint Conference on Artificial Intelligence, 1977.
- [4] Aikens, J. Prototypes: An approach to knowledge representation for hypothesis formation. Proceedings this conference.