

EXPERT: A SYSTEM FOR DEVELOPING CONSULTATION MODELS

Sholom M. Weiss and Casimir A. Kulikowski
Department of Computer Science
Rutgers University
New Brunswick, N.J. 08903

ABSTRACT

EXPERT is a system for designing and building models for consultation. An EXPERT model consists of hypotheses (which can be structured into causal and taxonomic networks); findings or observations; and decision rules for logically relating these components. A relatively simple language for describing models is employed. Logical and probabilistic rules are restricted to particular types which implicitly order the tasks performed by the control strategies: classification, question selection, and explanation. Explicit representations of decision rules are emphasized, as opposed to suboptimal scoring functions. This results in more easily predictable and correctable performance for a model. The system is currently being used to develop consultation models in domains such as rheumatology, ophthalmology, and endocrinology.

INTRODUCTION

The development of knowledge-based consultation programs that exhibit near-expert performance has progressed rapidly in recent years [1-4]. These efforts have focused on specific medical application areas and have relied on one or more experts to build a knowledge base. The hallmark of these consultation systems has been their reliance on large amounts of structured expert knowledge in the application domain [5]. In each system, the choice of knowledge representation has come first, and been a primary determinant of the type of reasoning procedures used. In contrast to traditional methods of automated diagnosis, little emphasis has been placed on the optimality of decision-making under general criteria (utility, information entropy, etc). Rather, these systems have attempted to develop representations based on simulations of the experts' conceptual structures and reasoning [6]. The various

systems have been reviewed and contrasted in [7] and [8]. Szolovits and Pauker [8] point out that there is frequent and general use of categorical (explicit) decision rules in all the systems, although they all also use various kinds of heuristic scoring methods to capture the uncertainty associated with medical decisions. Unfortunately, these methods often have side effects of propagation that make the results of a consultation less clearly predictable. Despite the specific applications for which they were first developed, some of the systems, such as CASNET and MYCIN have been generalized to accept knowledge bases in other domains. After several years of developing the CASNET representation [9], enough new ideas have emerged from our own experience and that of the other AIM (Artificial Intelligence in Medicine) investigators to lead us to develop a new and more general representation of knowledge for use in consultation systems.

The new consultation scheme EXPERT, consists of a general facility for developing and testing consultation models. It was designed without being tied to a specific application. EXPERT has been used to develop models in ophthalmology, endocrinology and rheumatology. Some of the major themes in the design of EXPERT include:

- 1) a relatively simple language and notation to represent expert knowledge. The representation is consistent with the traditional two-level view of a diagnostic problem: selecting appropriate hypotheses or conclusions by interpreting a set of findings or observations.

- 2) an emphasis on categorical reasoning instead of on suboptimal scoring functions. A form of probabilistic production rules is available, which can be compatible with statistically accurate classification when appropriate data is available.

3) descriptive knowledge as a partially-ordered conceptual network, a special type of a semantic net, which, as in the CASNET system, can be pre-compiled for the interpretation of individual cases, hence leading to efficient performance.

4) an emphasis on decision methods which tend to yield predictable and correctable results. This may in some cases require that the expert provide more explicit statements of correlations among findings and a greater number of decision rules. The representation and control strategy simplify the task of correcting erroneous conclusions. Since the interaction among rules and the associated classification strategies is consistently predictable, it is not difficult to trace the changes in program behavior that will result from modification of individual rules. A novel feature of EXPERT is that it automatically detects changes in the reasoning about cases stored in its data base that will arise from modification of the decision rules. This is an important tool for incrementally generating and testing an expert model.

The process of creating and running an EXPERT decision-making model is somewhat similar to writing and running a computer program. An editor is used to create a file which will contain statements describing a model. A model is translated into an efficient internal representation by the compiler program, XP. The model may then be executed, and cases may be entered for consultation, using the program EXPERT.

MODEL DESCRIPTION

An EXPERT consultation model consists of 3 sections:

- a) hypotheses,
- b) findings,
- c) decision rules

Findings are the facts about a patient elicited during a consultation. In a medical domain, they are the history, symptoms, signs, and laboratory test results. Findings are reported in the form of true, false, numerical, or unavailable responses to questions from EXPERT. Hypotheses are the conclusions that may be inferred by the system. They include diagnostic and prognostic decision categories, therapy recommendations, and intermediate hypotheses about pathophysiological states, expected causes of illness, or typical aggregates of findings. A measure of uncertainty is usually associated with a hypothesis.

Within the 3 sections of the model, several subsections are possible. The

representation used in building a model divides the sections as follows:

****HYPOTHESES**

***TAXONOMY**

[*CAUSAL AND INTERMEDIATE HYPOTHESES]

[*TREATMENTS]

****FINDINGS**

****RULES**

[*FF RULES]

*FH RULES

[*HH RULES]

Two asterisks indicate one of the three major sections. A single asterisk indicates a subsection, and brackets indicate optional statements.

HYPOTHESES

The major hypotheses are structured into a taxonomic classification scheme. Contained within the *TAXONOMY subsection shown above are the possible diagnostic and prognostic conclusions and the useful set-subset relationships between general diagnostic categories and intermediate interpretations. As an example, we show part of a thyroid disease classification:

***TAXONOMY**

EU .Euthyroid (.75)

ThD .Thyroid Dysfunction (.25)

HYPER ..Hyperthyroidism (.05)

HYPO ..Hypothyroidism (.20)

NOP .No Pathology (.70)

GRAV .Graves' Disease (.25)

The mnemonics for each hypothesis become a shorthand for specifying its place in the production rules. The optional weight associated with some hypotheses indicates their frequency of occurrence relative to the higher level hypotheses in which they are included. The indentation indicates the set-subset relationship among hypotheses.

The optional subsection of "CAUSAL AND INTERMEDIATE HYPOTHESES" is used to define hypotheses that may not belong in the taxonomy because they are not explicit categories of disease. However, these intermediate hypotheses can be valuable diagnostic summaries of important combinations of events. They can stand for pathophysiological states, such as "elevated thyroid uptake," or they may abstract many findings such as "ocular symptoms of Graves' disease." It is usually much simpler to reason with a small set of such hypotheses than with the much larger set of original findings. Hypotheses which appear in the taxonomy may reappear in this section indicating a causal or superset-subset relationship with the intermediate hypotheses.

Treatments are also described within

the general notational form of hypotheses. Their interpretation is somewhat different. They may stand as a taxonomy of therapies, or they may be organized into a hierarchy of treatment plans. The top level treatment hypothesis is the general treatment plan for a particular diagnostic category and lower level nodes are subsets of possible treatments. In this case, the weights indicate the prior frequency with which it is believed that a specific treatment will be successful, given that the treatment category is indicated for the patient. An example of a treatment category is:

***TREATMENTS**

RXTHY Thyroid medication
 RXUSP .Thyroid USP
 RXU1 ..Thyroid USP, 1 gram daily
 RXU2 ..Thyroid USP, 2 grams daily

FINDINGS

Findings are represented as attributes that can be present, absent, or undetermined in the patient, or as a numerical variable, which when measured adopts a value within a pre-specified range. In those cases where uncertainty is associated with an observation, the uncertainty must be described explicitly in terms of additional modifying findings. For example, the accuracy of a test result can be requested, and decision rules then written to logically modify inferences about the original test. Our experience has been that uncertainty about facts tends to be circumscribed by experts to a few important situations in each medical domain. Rather than impose a generalized heuristic for handling them, the EXPERT representation encourages the model builder to pay attention to the manner in which uncertainty in each important situation affects further inferences.

For purposes of acquiring information about a patient, four question types may be employed: a) multiple choice, b) checklist, c) numerical, and d) yes-no. The checklist question differs from a multiple choice question in that the choices are not mutually exclusive, and more than one may be true. The checklist question combines many questions that could be asked individually as single questions. Although some items in a checklist may be more important than others, placing them together serves to bring up for consideration related topics at the same time. There is a higher level of control over the sequence of questioning by the system. It is possible to group several of the question types (checklist, numerical, etc.) together in the form of a questionnaire. When a member of the

questionnaire is selected to be asked, all the questions must be asked in the order of the questionnaire. The questionnaire is particularly effective when used in conjunction with logical constraints among findings: the finding-to-finding, or FF, production rules.

Some questions may involve different degrees of risk or cost, which affect the general order in which they are usually sought in the course of a consultation. For example, history and symptom questions are almost always asked before a complicated laboratory test is requested. An optional cost declaration can be associated with a finding. It serves to order the acquisition of findings, since the questioning strategy seeks information about the least costly findings first. Examples of the question types are shown below:

****FINDINGS**

***BEGIN QUESTIONNAIRE**

***NUMERICAL/MIN=1/MAX=100**

AGE Age:

***MULTIPLE CHOICE**

Sex:

M Male

F Female

***MULTIPLE CHOICE**

Pregnancy test:

PREGP positive

PREGN negative

***CHECKLIST**

Symptoms Appeared/Intensified, Past Year:

RHP Rapid Heart, Palpitations

***END QUESTIONNAIRE**

RULES

There are three types of rules for describing logical relationships among findings and hypotheses:

- 1) FF - finding to finding rules,
- 2) FH - finding to hypothesis rules,
- 3) HH - hypothesis to hypothesis rules.

The FF rules specify truth values of findings that can be directly deduced from an already established finding. They are processed in the fixed order specified by the model designer, and are crucial for establishing local control over the sequence of questions in a fashion consistent with medical practice. The format of FF rules is:

$F(MNE1, TVAL) \rightarrow F(MNE2, TVAL)$

where MNE1 and MNE2 are mnemonics for findings and TVAL = {T, F, U, or a number}. If the left side of the rule is satisfied, the response to the right side of the rule is known, and question MNE2 will not be asked. A simple example of an FF rule is:

$F(M, T) \rightarrow F(PREGP, F)$

which would rule out seeking pregnancy tests for males. The FF rules are particularly useful when used with questionnaires. They can describe conditional branchings from sections of the questionnaire. This is most advantageous when questions are consecutively ordered from the general to the specific.

FH rules are logical combinations of findings which indicate confidence in the confirmation or denial of hypotheses. The general format for these rules is:

$X_1 \& X_2 \& \dots X_i \rightarrow H(MNE, CF) \quad -1 \leq CF \leq 1$

Where: $X_i = F(MNE, TVAL)$ or

$= [n:F(MNE1, TVAL), F(MNE2, TVAL) \dots]$

If the logical combination of findings on the left side of the rule is satisfied, the hypothesis, MNE, is assigned a confidence value, CF. The selector argument, n in $[n:F(MNE1, \dots)]$, indicates that if n of the listed findings are satisfied, the bracketed condition is evaluated as true. An example of an FH rule is:

$F(RHP, T) \& F(FFT, T) \rightarrow H(HYPER, .5)$

which summarizes the inference:

IF: rapid heart palpitations are reported by the patient to have noticeably increased in the past year and a fine finger tremor is observed in the patient,

THEN: consider the hypothesis of hyperthyroidism with a confidence of 0.5.

Confidence measures are assigned on a scale of -1 to 1, with 1 being complete confirmation and -1 being complete denial, following the CASNET and MYCIN rationale and usage [2,9]. When in the course of a consultation several rules are applicable in inferring a hypothesis, they are compared and the single rule with the MAXIMUM absolute value of confidence in the hypothesis will prevail. The maximum is used as a fuzzy logic selector because of the disjunctive nature of the rules. Because no universal scoring function can adequately handle conjunctions of findings, we have taken the approach that requires the designer to explicitly specify separate rules for significant combinations of correlated findings. In return, the scheme guarantees predictability of weight interpretation for the model designer. This applies to the FH rules. After these rules are processed, inferences may be derived through the HH rules and the taxonomic-causal network. These procedures are described below. The system also assigns a small bonus for hypotheses that have the greatest number of satisfied rules supporting them.

The HH rules allow the model builder to

specify inferences among hypotheses and treatment selections that follow from other (diagnostic and prognostic) hypotheses. Since such higher level inferences may be sometimes modified by the presence or absence of a finding, the left hand side of HH rules may also contain assertions about findings. In addition, each HH rule must have a context defined in terms of a set of findings or hypotheses, which enables the application of efficient rule evaluation strategies over the compiled model. The context specifies a set of necessary conditions among the findings to enable evaluation of the HH rule. HH rules are evaluated in the order of specification in the model. Hypotheses are specified with a range of confidence which must be satisfied for the HH rule to be invoked. The HH rules are implemented in a table of the following form:

*HH RULES

*IF THERE ARE EYE AND THYROID DYSFUNCTIONS

$F(EENO, F) \& F(HXTH, T)$

*THEN CONSIDER GRAVES' DISEASE:

$F(ETHO, T) \& H(EYE, .5:1.) \rightarrow H(GRAV, .9)$

*END

CONTROL STRATEGIES

In any consultation system, the principal control strategy problem is to structure and rank conclusions. In many general situations the control of sequential questioning is also desirable, whereas in more stylized and well-defined applications, a prespecified logical control of questioning may be possible. Question selection is usually closely related to the formulation of tentative hypotheses that best explain the patient's condition.

-- Order of Evaluation --

When an assertion or result about a finding is made to the system, the rules in which this result appears are evaluated. The results are usually received in batches, as in the form of responses to a multiple choice or checklist question. The rules are evaluated in the following order to produce weights which rank the hypotheses.

1. FF rules are evaluated first. They take simple true, false, or unknown responses, and merely enlarge the set of new finding results. They are handled in the same way as results received directly in response to questions.

2. FH rules are evaluated next. Only those FH rules in which new results of findings appear need be evaluated. FH rules have the property that when the left hand side of a rule is evaluated as true or false, it remains true or false for the

remainder of the consultation session, unless an erroneous response is received.

3. HH rules are evaluated last. The IF part of each HH rule table is evaluated at the same time as the FH rules. Only HH rules found in tables which have their IF part evaluated as true are considered. All such HH rules must be reevaluated sequentially. The premises and consequents of HH rules may include hypotheses and associated intervals of confidence. Unlike findings which remain true or false, these intervals can change not only directly from finding results, but also indirectly from other rules (both FH and HH) which affect the confidence measure of a hypothesis. HH rules are evaluated in the order of their appearance in the model. There is no backwards chaining, because the order of evaluation is determined beforehand by the model builder. Self-referencing rules are therefore acceptable. Because HH rules and tables are evaluated in sequence, by their order of appearance in the model file, it is important to order the HH rules carefully. An HH rule implying a hypothesis H1, which is later needed to establish another hypothesis H2, should appear in the model file before H2 is referenced in the left side of another HH rule. The confidence ranges used in the HH rules are those which have been directly set by other FH or HH rules. They are not the adjusted weights implied by the taxonomy.

4. The above procedures result in the assignment of those confidence measures, $CF(i)$, which can be directly determined from the rules of evidence and hypothesis weight propagation: the FH and HH rules. When more than one rule is applicable, the maximum absolute value of confidence is used. Another procedure is invoked that is helpful both in question selection and as a simple heuristic to adjust weights slightly. Each hypothesis which has some positive evidence, in the form of a satisfied rule or a partially satisfied rule (with unknown truth value) is marked. The count of such rules which apply to each hypothesis is kept. This corresponds approximately to the number of positive indicators of a hypothesis.

5. A different set of weights is derived from the taxonomy and causal network structure. Forward weights are propagated from predecessor to successor and inverse weights are propagated from successor to predecessor. A taxonomy contains implied relationships between hypotheses that can be treated similarly to causal connections.

The procedures used to generate weights are similar, but not identical, to those used in CASNET [9]. After all FH and HH rules have been evaluated, the measures of confidence, $CF(i)$, may be modified or propagated according to their taxonomic or causal relationships. Because hypotheses which are related by causal or taxonomic connections can be topologically sorted, efficient algorithms can be used to compute the weights implied by this partially ordered network.

6. A final weight is derived from both the rule-based and taxonomic-causal net weights. It is taken as the maximum absolute value from all the indicated directions (with the appropriate sign). A small bonus (no greater than .33) may be awarded to the final weights. The bonus is given on the basis of the percentage and number of rules (derived directly from findings) that are covered by any single hypothesis. The largest bonus is given to the hypothesis which can potentially cover the most rules. The bonus effect is slight and has its greatest effect when only few results have been received, and not many high confidence rules are satisfied.

-- Sequential Test Selection --

The question selection procedures are relatively exhaustive, but straightforward and efficient. Most of the tasks performed by the selection procedure are simple and follow directly from the inference procedures. Once a single finding is selected to be asked as a question, the explicit question structure, such as a multiple choice or checklist question is invoked. Rules are examined that have been evaluated as unknown and their effect on the most likely current hypotheses is considered. Question structures temporarily override the sequential question selection procedures; the system proceeds with the pre-specified questioning order, subject to the FF rules for branching. The general sequential test selection procedure may be specified as follows:

I. Consider those rules (FH or HH) such that:

a) They set confidence measures for marked hypotheses - they are related to current evidence.

b) The confidence of the rule is greater than the current confidence in $H(j)$ - it provides useful additional information.

II. Select an unasked finding $F(i)$, belonging to the set of rules found in part I. such that:

a) the cost is minimum;

b) the current weight $W(H)$ is greatest - indicating that the finding pertains to the most likely hypothesis;

c) the potential $CF(H)$ is greatest for $F(i)$ - so that this finding $F(i)$ is one that can yield greater confidence in $H(j)$ than another $F(k)$.

THE XP AND EXPERT PROGRAMS

The system is written in FORTRAN and occupies about 70K on a DEC-20 computer. This results in relatively efficient processing and also provides the potential for implementation on a minicomputer. The program XP compiles a model for use by the EXPERT consultation program, and indicates any errors found in the model.

The consistency checking module considers the effect of modifications to a model. A data base of representative cases with known conclusions is stored, and after the model has been modified and compiled, the XP program proceeds sequentially through the cases, noting any changes in the reasoning for each stored case, as reflected in the confidence weight assignments to the major conclusions. This is a practical means of reviewing consistency in a model, short of more powerful theorem proving capabilities. It is a more empirical approach than that taken in TEIRESIAS [10], where a model of the relationships among rules and control strategy was developed to help the model builder write new rules for correcting a previously misdiagnosed case, without tracing the effect of rule changes on other cases.

The user communicates with EXPERT by responding to questions posed by the program or by using a simple command language. Commands may be issued on a sequential basis, so that the current ranking of conclusions may be determined. Explanatory information is also available, such as reasons for asking a question, or an analysis of why a hypothesis has a particular confidence weight.

RESULTS AND DISCUSSION

Several applications of the EXPERT system are in progress. Consultation models are being developed in endocrinology (thyroid diseases), ophthalmology, and rheumatology. A consultation system in rheumatic diseases appears to have strong potential for acceptance by the medical community. There is a shortage of rheumatologists in the United States, and keeping up-to-date on the interpretation of many new immunological tests can be difficult for non-specialist physicians. In collaboration with investigators at the

University of Missouri we have tested the EXPERT formalism by developing a prototype consultation system for rheumatic diseases. Initially, the set of problems considered has been confined to fewer than 10 important, yet complex diagnostic categories. The model is being expanded to cover additional problem areas.

REFERENCES

1. Weiss, S., Kulikowski C., Safir A., Glaucoma Consultation by Computer, Computers in Biology and Medicine, Vol. 8, pp. 25-40 (1978).
2. Shortliffe, E., Computer-based Medical Consultation: MYCIN, Elsevier, N.Y. (1976).
3. Pople, H., Myers, J., Miller, R., The DIALOG Model of Diagnostic Logic and its use in Internal Medicine, Proceedings 4th International Joint Conference on Artificial Intelligence, pp. 848-855 (1975).
4. Duda, R., Hart, P., et al, Development of the Prospector Consultation System for Mineral Exploration, SRI Technical Report (1978).
5. Feigenbaum, E., The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering, Proceedings 5th Joint Conference on Artificial Intelligence, pp. 1014-1029 (1977).
6. Pauker, S., Gorry, G., Kassirer, J., Schwartz, W., Toward the Simulation of Clinical Cognition: Taking a Present Illness by Computer, The American Journal Of Medicine 60, pp. 981-995 (1976).
7. Pople, H., Artificial Intelligence Approaches to Computer-based Medical Consultation, IEEE Intercon Conference, (1975).
8. Szolovits, P., Pauker, S., Categorical and Probabilistic Reasoning in Medical Diagnosis, Artificial Intelligence, Vol. 11, pp. 115-144 (1978).
9. Weiss, S., Kulikowski, C., Amarel, S., Safir, A., A Model-based Method for Computer-aided Medical Decision-making, Artificial Intelligence, Vol. 11, pp. 145-172 (1978).
10. Davis, R., Interactive Transfer of Expertise: Acquisition of New Inference Rules, Proceedings 5th Joint Conference on Artificial Intelligence, pp. 321-328 (1977).

ACKNOWLEDGEMENTS

This research was supported in part by grant RR-643 of the NIH Biotechnology Resources Program. The EXPERT system has been programmed by Kevin E. Kern.