A panel session—AI in medicine

SESSION CHAIRMAN—SAUL AMAREL Rutgers University

Panel Members

Peter Szolovits—Massachusetts Institute of Technology
Harry Pople—University of Pittsburgh
Casimir A. Kulikowski—Rutgers University
Sholom M. Weiss—Rutgers University
J. C. Kunz—The Institutes of Medical Science, Pacific Medical Center
L. M. Fagan—Stanford University
R. J. Fallat—The Institutes of Medical Science, Pacific Medical Center
D. H. McClung—The Institutes of Medical Science, Pacific Medical Center
J. S. Aikins—Stanford University
H. P. Nii—Stanford University
E. A. Feigenbaum—Stanford University
J. J. Osborn—The Institutes of Medical Science, Pacific Medical Center
Mark Stefik—Stanford University
Peter Friedland—Stanford University
R. O. Duda—SRI International
P. E. Hart—SRI International
R. Reboh—SRI International

PANEL OVERVIEW—Saul Amarel

The papers in this session will focus on applications of Artificial Intelligence in the development of knowledge-based systems in medical decision-making situations, in scientific inquiry, and in mineral exploration.

Much of the work which will be reported in this session is being done within the AIM community. [AIM stands for Artificial Intelligence in Medicine.] The AIM project is a national resource sharing activity supported by the biotechnology program (BRP) of DRR, NIH, whose principal objective is to promote applications of AI to medicine and the life sciences. The main focus of AIM activity is at the Stanford University SUMEX-AIM project, which provides computer shared resources to the AIM community via national networks. The Rutgers research resource on computers in biomedicine is one of the (BRP-supported) projects in the AIM community. Included in its functions is the organization of AIM workshops. The objective of the workshops is to strengthen scientific interactions within the national AIM community and to disseminate AI-based methodologies, tools, and specific systems that are relevant to AIM. The next AIM workshop (this is the fourth) will be held at Rutgers on June 25-28, 1978. All the participants in this session are expected to attend the Rutgers workshop and to discuss in depth various aspects of their research on knowledge-based systems. The present session will provide an opportunity to bring summary accounts of this research to a wide national audience of computer professionals; a similar session was held last year at the Fifth International Joint Conference on Artificial Intelligence.  

Experience with work in AI applications to date shows the following:

(a) The problem of acquiring specific knowledge in a domain, managing it in an AI system, modifying it, and using it appropriately is fundamental. The approach to most designs is incremental and responsive to the fact that the knowledge base in the domain is not stationary. Initially, a relatively low-performance AI system is created to provide the basis for subsequent stages of knowledge acquisition and improvement, which eventually leads to high performance, and expert systems.

(b) Work on applications requires very close collaboration between AI experts and experts in the problem domain. Furthermore, special technical support facilities (e.g., computer networks) can play a significant role in establishing these collaborations and in sustaining their effectiveness. This is a key point which has important implications for organizational and shared resource aspects of applied AI projects.

(c) The development of an expert system within a reasonable time span requires more powerful technologies than those in use today—especially when the knowledge bases will grow from the present $10^{\ast}2$ to $10^{\ast}3$ “facts” to more realistic situations with $10^{\ast}4$ to $10^{\ast}5$ “facts”. So far, system development times (from conception to expert level in a research environment) have been 4 to 8 years. To reduce this time span, or to keep it from growing too much as knowledge bases grow, we need more effective methods of knowledge acquisition and organization and more powerful program design environments. Related to this, we need better techniques for interfacing AI programs with experts and users. At a more basic level, we need better schemes for coordinating multiple knowledge bases and for handling inconsistent and/or uncertain information.

From the collection of the Computer History Museum (www.computerhistory.org)
Experience of several years in the development of clinical decision-making “expert” systems at MIT to be discussed by Peter Szolovits illustrates the ways in which these issues are encountered in specific AI application projects. Another important conclusion obtained from the MIT experience is that the choice of medical tasks to be approached via AI methodologies/techniques should be made after a careful assessment of the nature and complexity of the task. Only tasks requiring complex knowledge bases and difficult reasoning processes should be approached via AI methods; in many situations simpler computer-based methods may be more appropriate.

One of the key processes in medical reasoning and in scientific inquiry is the interpretation of empirical data in the light of a given body of theoretical knowledge. Much of the work presented in this session is concerned with these processes. Typically, a problem of interpretation involves reasoning about an individual case. Given evidence (data) about the case and a theory within which the evidence is to be understood/explained, find a hypothesis which explains the case on the basis of the theory. While there has been considerable progress in the development of strategies for solving interpretation problems, there still remain several open problems: under what conditions the interpretation process should be controlled by the specific, “low level” features of the case under consideration or by possible “high level” hypotheses and by expectations derived from these hypotheses; how to best represent and keep track of information about a special case, of general knowledge in terms of which the case is interpreted, and of alternative interpretations of the case as the process evolves with time.

Some of these issues will be discussed in depth by Harry Pople in the context of his research (in collaboration with Dr. Jack Myers) on diagnostic problem solving in internal medicine. These are also important issues in (I) the work of John Kunz, Larry Fagan and their collaborators in the domain of pulmonary disorders, (II) the work of Richard Duda, Peter Hart and their collaborators in the domain of mineral explorations, and (III) the work of Sridharan, which is oriented to the development of an AI system for assisting in the formulation of psychological theories of action interpretation.

Processes of planning are becoming another focus of interest in AI applications to natural systems. The synthesis of plans has been a subject of research since the early days of AI. In this session, the synthesis of treatment plans in a medical context (consultation in glaucoma) will be discussed in the paper by Casimir Kulikowski and Sholom Weiss. AI approaches to the development of plans in the context of scientific experimentation (in molecular genetics) will be discussed in the paper by Mark Steffik. Research on the analysis/interpretation of plans is of a more recent vintage in AI and it has been stimulated by AI work in “real life” situations. Work on this problem will be presented in the context of AI applications to scientific work in the papers by Steffik (molecular genetics) and Sridharan (psychology).

Some of the systems presented in this session have already reached high levels of expertise in limited domains. CASNET/glaucoma at Rutgers and INTERNIST at the University of Pittsburgh and PROSPECTOR at SRI are in this category. There are also several “younger” systems that are still in a process of intense development and growth. In some of them techniques developed/tried in previous knowledge-based systems are being assimilated and adapted (e.g., the structure of MYCIN4 is being used in the PUFF project) and the DENDRAL5 experience as well as ideas from several more recent AI efforts are being used in the MOLGEN project. On the other hand, in the BELIEVER task discussed by Sridharan, a new system, called AIMDS, is being developed to provide an appropriate environment for theoretical work on the task.

We are now at a point where substantial achievements have been made in AI applications and there are several new efforts that are starting to build on top of past experience. In general, there has been good progress in this area and the prospects for the future of high performance in AI expert systems are bright.

The following is a summary of the papers to be presented at the NCC session.

REFERENCES


THE DEVELOPMENT OF CLINICAL EXPERTISE IN THE COMPUTER—THE EVOLUTION OF CLINICAL DECISION-MAKING PROGRAMS AT MIT—Peter Szolovits

We at the Clinical Decision Making Group at MIT’s Laboratory for Computer Science, along with our collaborators at the Tufts-New England Medical Center Hospital, have been engaged for several years in the development of computer programs which embody expert medical knowledge. We have applied techniques developed in the study of Artificial Intelligence (AI) to encode expert clinicians’ approaches to handling a number of important medical problems: (1) prescribing appropriate doses of digitalis glycosides to patients with congestive heart failure or arrhythmias (the DIG program), (2) taking the history of the present illness...
of a patient with kidney disease (the program is called PIP), and (3) diagnosing and treating patients with acid/base and electrolyte disturbances. We have also studied the application of decision analytic techniques to the testing and treatment of Hodgkin’s disease, the differential diagnosis of acute renal failure, the value of coronary surgery to the individual patient, and other problems.

Throughout many of these projects, we have discovered that our initial estimate of the difficulty of the problem has been far from what we finally decided. Naturally, we have often underestimated the magnitude of the effort needed to collect data, generate plausible decision structures, and choose and build computer implementations for the data base and reasoning components of the program. Not so obviously, however, we have also found that some problems succumb to much simpler techniques than what we had originally estimated to be needed.

The key to these observations, we now think, concerns the degree to which limited human-like expertise in a computer program is adequate to cover the great majority of situations which arise in the program’s domain. To the extent that a simple statistical model or a small set of decision rules or protocols can be made to work, any program which uses those models will be straightforward. It is only when the task involves complex reasoning to deal with a wide variety of medical possibilities, when a “deep” understanding of a problem is needed, or when the medical situation can consist of a complex combination of many simultaneous difficulties that the programs we write (and the experts’ knowledge which with models) must be sophisticated.

The presentation will review a number of attempts to build programs for digitalis therapy, to contrast them in terms of the magnitude of medical knowledge involved in each and the corresponding sophistication demanded of the computer implementation.

REFERENCES


THE ROLE OF HYPOTHETICAL REASONING IN DIAGNOSTIC PROBLEM SOLVING—Harry Pople

INTERNIST is a computer-based system designed to deal with complex diagnostic problems in internal medicine. An important aspect of this task domain is the need to synthesize ad hoc diagnostic categories to characterize patients with two or more disease processes at one time. The presentation will review the role of hypothetical reasoning in the INTERNIST approach to this concept-formation component of the diagnostic task.

REFERENCES


STRATEGIES OF GLAUCOMA TREATMENT PLANNING—Casimir A. Kulikowski and Sholom M. Weiss

The ultimate goal of most difficult medical consultations is to produce advice that will help formulate a plan of treatment with the greatest likelihood of success for the patient. Recently, several computer consultation programs using Artificial Intelligence techniques have developed different strategies for treatment planning. The strategies used in the CASNET/Glaucoma program are designed to select plans for the long-term management of complex glaucoma cases over many follow-up visits.

Although many subjective rules of clinical judgment enter into the formulation of treatment plans, the expected outcomes are usually based on an understanding of the physiological effects of the treatment. In CASNET/Glaucoma, a causal model describes the mechanisms of disease as well as their expected changes following treatment. A two-level data structure is used: general treatment plans subsume ordered sequences of specific treatments. The initial step is to select a general plan of treatment on the basis of the patient’s diagnostic categorization. The choice of specific treatment within the plan is determined by many of the individual characteristics of the patient, through the use of scoring functions for the enhancement and inhibition of selections. For example, the status of the fellow eye of a patient becomes a crucial factor when surgery is contemplated: the potential risk may be too great if there is little or no vision in the fellow eye.

For follow-up visits, the effectiveness of any current treatments must be judged and compared with the expected benefits of the alternatives suggested by the selection strategy. The diagnostic status of the patient must also be reassessed, in the case that changes dictate a complete change of treatment plan. In the event that the patient remains within the scope of the original general plan, but the original specific treatment is not sufficient to control the progression of disease, the program selects the next, stronger medication in
the sequence. Potential side-effects or complications are monitored and may result in changing the order in the sequence of specific therapies.

The CASNET/Glaucoma program's recommendations were compared to those of a panel of experts at the 1976 meeting of the American Academy of Ophthalmology and Otolaryngology. The resulting discussion is serving as the basis for testing a variety of new model-based strategies for treatment.

REFERENCES

2. Weiss, S., A System for Model-based Computer-aided Diagnosis and Therapy, Ph.D. Dissertations, Rutgers University, 1974.


Interpretation of physiological measurement data is a central process in medical diagnosis. Biomedical engineering has succeeded in providing a wide variety of quantitative measures of physiological state, but the biomedical community has made only limited effort at systematically attempting to automate the interpretation of the medical significance of these quantitative measurements.

We characterize the problem of medical data interpretation by its medical and its sociological aspects. We choose to highlight the following features of the medical problem:

Judgmental: Deductive and inductive reasoning are both used to conclude a diagnosis given a medical situation;
Uncertain: There is only a loose relation between measured result, physiological state, overall clinical state, and effects of therapy;
Redundant: Measurements have redundancy with each other;
Context-dependent: Interpretation depends upon the clinical situation;
Time-dependent: The value and meaning of information changes.

In addition, we highlight the following features of the sociological situation in which medicine is practiced:

Experience: Good clinical practice is characterized by experience, in addition to didactic knowledge;
Interpretation: The consultant, either human or automated, should explain the process followed in making a consulting diagnosis.
Expectation: Patient condition is viewed as improving or deteriorating in relation to an expectation of disease course.

We will describe a general architecture for a medical interpretation system which recognizes these features of the problem. We have chosen Artificial Intelligence (AI) techniques for our approach to medical data interpretation. AI offers an approach to symbolic reasoning which is amenable to automation. The symbolic manipulation allows us to represent the knowledge of the clinical problem in a direct form, i.e., as structured relations among medical states and measurements. We report here on our early progress with use of AI to address two medical problems as we see them.

The PUFF system is now in routine use in our hospital for interpreting pulmonary function test measurements. PUFF interprets the medical significance and implications of input quantitative test data and patient history. Using a production rule formalism, PUFF has a set of about 60 physiological rules of the general form "IF condition THEN draw conclusion." The rule formalism has allowed our pulmonary physiologist to define the medical interpretation problem in his terms, using a level of judgment and a compensation for the uncertainty and redundancy of the data which he feels appropriate. The system is implemented in a version of the MYCIN system on the PDP 10 and in a straightforward way on the PDP-11. The system operates in a batch mode, accepting input from the data collection computer and printing out interpretations. In a 144-case evaluation, the system showed 86-93 percent agreement with the physiologist in diagnosing the presence of major pulmonary disease. For comparison, the physiologist changed his initial diagnosis of the presence and severity of major pulmonary disease upon further consideration in 42 of 107 cases during an early validation process.

The VM system is designed to interpret the clinical significance of quantitative measurements from our ICU patient monitoring system. Every 2-10 minutes, VM will accept the 40 measurements comprising the monitoring data and make a context and time dependent interpretation of the significance of data. The system first compares measured data with expected values in order to identify the possible presence of alarm conditions. The system has rules which define criteria for accepting data as physiologically indicative or possibly spurious. Next the system focuses its attention on the measurements which categorize the clinical state of the patient, and then it attempts to identify possibly advantageous therapies. VM now includes the capability to process rules describing optimal procedures for removing patients from mechanical respirators. The program will print suggestions to clinicians concerning optimal control of respirator settings and choice of specific respirator types. Detection of unexpected mechanical failures and potentially significant
changes in the patient's physiological measurements will be interpreted and reported to the clinician.

MACHINE INFERENCE FOR MOLECULAR GENETICS: METHODS AND APPLICATIONS
—Mark Stefik and Peter Friedland

An increasing amount of research in Artificial Intelligence is involved with difficult applications which require representing and using significant amounts of specialized knowledge in a computer. Several researchers 5,6 have suggested that this represents a shift of paradigm in AI—from seeking performance via powerful general methods toward a recognition that many difficult processes are knowledge intensive. This view emphasizes the use of large amounts of specialized knowledge in a fashion that facilitates their effective use and interaction.

The MOLGEN project at Stanford is an example of this trend in research. MOLGEN the joint effort of computer scientists and geneticists at Stanford, is broadly concerned with the application of symbolic reasoning to molecular genetics. Molecular genetics is a vigorous and rapidly developing discipline with many logical and medium-sized problems for potential computer application. The main thrust of the MOLGEN effort has been in the development of programs to assist in the design of laboratory experiments. Experiment design requires information about problem solving goals, available laboratory techniques, and strategies for combining these techniques to achieve the desired goals. Work in computer-assisted design has been concentrated in a limited class of analysis experiments and a limited class of synthesis experiments.

Development of performance programs in the MOLGEN project has required work on several fronts. Experiment planning involves representation of many kinds of laboratory methods and objects. To avoid creating separate packages for each kind of knowledge, we have developed a schema-based representation package 7 with some fairly general capabilities. This part of our work, which occupied much of our first year of research, was considerably influenced by recent work in representation languages. 8 At the same time, the process of planning experiments and resolving conflicts has drawn on recent work in problem solving. 9 Prior to beginning our work in this area, we did a substantial review of problem solving methods. 10

Projects like MOLGEN foster the art of applying the principles and tools of AI research to bear on specific problems. This usually involves adapting and combining these tools; it always involves comparing and selecting among alternative approaches. One of the motivations for these efforts is their utility as case studies in methodology. We believe that artificial intelligence will mature as a discipline only by confrontation of its basic methods with the challenges of various applications. From the case studies are slowly emerging the bits of a theory, which consists of an increasingly well characterized collection of paradigms for solving problems and representing knowledge. One example of the results of a case study from our own work is in Reference 8.

REFERENCES


COMPUTER-BASED CONSULTANT FOR MINERAL EXPLORATION—R. O. Duda, P. E. Hart, and R. Reboh

For the past two years, we have been developing an interactive system of computer programs called PROSPECTOR that is intended to act as a consultant to field geologists on problems of mineral exploration. The performance of the system is based on an internal set of models of various types of mineral deposits, and on a taxonomy of rocks, minerals, and other geological concepts. The models and taxonomy, termed the knowledge base, are provided to the system over a period of time by a panel of expert geological consultants. The field geologist using the system furnishes a set of observations, often incomplete and uncertain, about the particular prospect under investigation. The principal tasks of PROSPECTOR are to compare the observations of the user against the models, to find the most promising matches, to identify for the user the most important missing observational data, and finally to form an estimate of the likelihood that a deposit in fact occurs on the prospect. Viewed abstractly, PROSPECTOR attempts to aggregate the knowledge of expert geologists in order to apply that knowledge to particular mineral exploration problems.

The models of mineral deposits are represented in a special data structure termed an inference network. A node in the network represents an arbitrary geological situation like "The prospect lies within 200 miles of a subduction zone" or "The ratio of pyrite to bornite on the prospect is less..."
than 1." Each node has an associated probability. A subset of nodes are distinguished as model nodes; that is, a node in this subset represents the situation that a mineral deposit satisfying a model occurs on the prospect.

Nodes in the network are connected by directed arcs. An arc corresponds to either a logical connective or to a probabilistic inference relating the nodes it joins. An important form of reasoning in PROSPECTOR is to propagate probabilities along arcs from one node to the next. The computations for propagation are based on subjective probabilities furnished by the experts, and are described in Reference 1.

The inference network formulation is particularly well-suited to our purpose because it allows models to be built up incrementally over time by defining new situations and relations. Each relation, together with the nodes it joins, is called a rule, and for this reason PROSPECTOR is termed a rule-based inference system.

The situations described by the nodes in the inference network are frequently compounded out of simpler situations; in the previous illustration the occurrence of pyrite and bornite are evidently the primitive geological situations of interest. It is often important to reason about these primitive constituents of a more complex situation. For example, the user may have already supplied the system with the observation that no sulfide minerals occur on the prospect. Since both pyrite and bornite are sulfides, PROSPECTOR should be able to infer that it is meaningless to inquire about their ratio. To support reasoning such as this, we represent each situation in the inference network in a more detailed structure called a semantic network.2,3 The semantic network for the previous example would explicitly articulate a situation in which both pyrite and bornite are present and are in the prescribed ratio. Furthermore, additional arcs denoting the subset relation would link pyrite and bornite to the mineral taxonomy in general and to the sulfide minerals in particular.

PROSPECTOR operates as a mixed-initiative system. At any time the user may take the initiative and supply the system with observational data or may request further information of several sorts. When the user relinquishes initiative, PROSPECTOR asks the user to establish by observation the likelihood that a particular geological situation occurs on the prospect. PROSPECTOR selects situations—that is, nodes in the inference network—that will have the most pronounced effect on the probabilities of the model nodes. Thus, PROSPECTOR in effect pursues a strategy of trying to confirm or refute hypotheses about the occurrence of a mineral deposit on the prospect.

REFERENCES

2. Hendrix, G. O., "Expanding the Utility of Semantic Networks Through Partitioning."