

THE ROLE OF HYPOTHESES IN MEDICAL DIAGNOSIS

Andee Rubin
MIT Artificial Intelligence Laboratory
Cambridge, Massachusetts 02139

Abstract

This paper describes some aspects of a computer system for doing medical diagnosis in the specialized field of kidney disease. A four-step process which consists of disposing of findings, activating hypotheses, evaluating hypotheses locally and combining hypotheses globally is examined for its heuristic implications. The work attempts to fit the problem of medical diagnosis into the framework of other Artificial Intelligence problems and paradigms and in particular explores the notions of pure search vs. heuristic methods, linearity and interaction, local vs. global knowledge and the structure of hypotheses within the world of kidney disease.

Introduction

The practical importance of studying and developing computer aids for medical diagnosis is obvious. Recently, cognitive psychologists and A.I. researchers have also become interested in the structure of medical knowledge and the processes by which it is manipulated as examples of general knowledge structures and problem-solving processes. This paper summarizes work done investigating these concepts while studying the process of diagnosis within a limited set of diseases: those whose presenting symptom is hematuria, or blood in the urine. This study was done in conjunction with Tony Gorry's group at M.I.T. <Gorry 74> and draws on previous and present work by Tony Gorry, Steve Pauker and Gerry Sussman.

Description of the Problem

We can conceptualize the diagnosis problem as one of a class of recognition problems <Fahlman>. In particular, a medical diagnosis system is presented with a group of symptoms, signs, facts, test results etc. and its job is to come up with a diagnosis, an identification of a disease or several diseases whose manifestations most closely match the condition of the patient.

Because of an interest in process, the model of diagnosis which will be used here is one of the serial acquisition <Gorry 67> of facts about the patient. Thus, we require a diagnosis system to have hypotheses about the patient's disease(s) at each moment and expect that these hypotheses will change after the addition of each new piece of information. In this paper, I will concentrate on the hypothesis-generation and evaluation aspects of the diagnostic process, identifying several mechanisms which serve

to minimize the number of hypotheses being actively considered at any one time. These heuristic methods may be new answers to the problem of limiting hypotheses in a potentially exponentially-exploding search-space.

The following sections present these aspects of the theory: a representational structure for signs, symptoms and diseases, the local evaluation of hypotheses, interaction between symptoms which affects evaluation and global assembly of complex hypotheses, which includes a theory of coherence designed to provide a way of combining local hypotheses into global ones

Basic Concepts and Mechanisms of the Theory

Nodes in the knowledge net necessary to do medical diagnosis primarily represent causes and effects. I have called a node which is primarily an explanation or cause an elementary hypothesis. Those nodes which are basically not causes, but rather raw data, are called findings; several types of findings will be detailed later. Elementary hypotheses are differentiated from findings by the fact that they are subject to a local evaluation process which determines how likely they are to be the diagnosis in light of the current data. Each elementary hypothesis is part of a slice which contains abnormalities which are caused by the disease it represents, as well as facts about a patient's age, sex, race or family history.

Take, for example, the two slices represented in Diagram 1, URINARY-TRACT-INFECTION and PYELONEPHRITIS. Each of them is an elementary hypothesis. The findings associated with URINARY-TRACT-INFECTION (UTI) include FEVER, HEMATURIA, FREQUENCY etc. Those associated with PYELONEPHRITIS include PUS-CASTS, PAIN (LOCATION FLANK) and IVP (RESULT SCARRING). There are several other elementary hypotheses on the page: GLOMERULITIS and IRRITATION are included to illustrate that a single finding may exist in many slices. The symptoms of a disease (relevant findings) are connected to its elementary hypothesis by pointers which have been left unlabelled, and should be read "is evidence for." The designations of EVIDENCE and EXPECTATION strengths, whose derivation is explained below, are included with the diagram.

Some elementary hypotheses are examples of others - more specific designations of etiology. STREP-PHARYNGITIS, SCARLET-FEVER and STREP-SKIN-INFECTION are all examples of STREP-INFECTION. An ISA link goes from the more specific example to the general category; the CHOICE-SET of a category, its set of examples, is also indicated.

Elementary hypotheses can sometimes themselves be considered symptoms, so that EVIDENCE and EXPECTATION pointers may connect them. In Diagram 1, UTI and PYELONEPHRITIS, both elementary hypotheses, are so related; the symptoms of UTI are a subset of those of PYELONEPHRITIS. There are also clearcut CAUSE relations between elementary hypotheses where the symptoms of the two diseases concerned are not in a subset/superset relation. In this case, the CAUSE relation is stated explicitly, as in STREP-INFECTION CAUSES AGN. Similar to CAUSE links are COMPLICATION links, as in PYELONEPHRITIS is a COMPLICATION of STONE. In addition, elementary hypotheses may or may not be ULTIMATE-ETIOLOGIES. An elementary hypothesis which is an ULTIMATE-ETIOLOGY is one which could stand alone as a diagnosis, for which a more basic cause does not have to be sought or is not known.

The system knows about several different types of findings: LAB-DATA, PHYSICAL-EXAM, SYMPTOM, FACT and FAMILY-HISTORY. The process of deciding whether or not a particular patient finding is relevant to the symptom description in the knowledge network and thus relevant to the disease hypothesis is called fitting: (see <Winograd> and <Minsky> for earlier uses of this term in frame theory) It requires trying to fit a particular finding-description into a sometimes more general specification. The result of this fitting process is that the finding-specification is either confirmed or disconfirmed, if the finding contains enough detail; if more information is needed to see if the finding fits the specification, the doctor often asks more questions.

The time of occurrence of the symptoms and the time relationships indicated in the knowledge net are also taken into account in fitting, as described in <Rubin>.

During the course of a diagnostic session, nodes of the data network change state with the addition of new information about the patient. While finding-specifications may be either confirmed or disconfirmed, elementary hypotheses, because they are not directly confirmable, have a more complicated set of alternative states. When a diagnostic session starts, all elementary hypotheses are inactive; that is, no particular disease has been suggested by the patient's symptoms. As more data is presented, certain hypotheses become active by virtue of their correlation with and ability to account for the findings present. Once a hypothesis is active, it is evaluated after the addition of every finding to see how well it fits the data so far and some score is produced which represents the likelihood of that disease's being present. On the basis of

this process, a hypothesis may be accepted or rejected; in most cases, however, no definite decision will be made, but its score will be modified to reflect the effect of the new data. An accepted elementary hypothesis is one for which the evidence is sufficiently specific to rule out any other cause for the symptoms present. For example, the presence of RED-BLOOD-CELL-CASTS confirms the diagnosis of GLOMERULITIS, making it an accepted hypothesis, but the very same finding makes SICKLE-CELL-TRAIT a rejected hypothesis. Finally, consideration of an elementary hypothesis may be deferred until more supporting symptoms are known. This process helps reduce the number of concurrently-active hypotheses.

How does the magic transformation from a bunch of symptoms to a final diagnosis take place? The process seems to be divided into four steps: disposing, triggering, local evaluation and global assembling. This series of four steps is performed after the addition of each finding. A brief description of the four stages follows; the third and fourth will be discussed in more detail below.

1. Sometimes the cause of a finding is clear when the finding is encountered; this is most often the case when the explanation is a FACT. In such a circumstance, the doctor disposes of the finding as a result of some already-accepted etiology rather than trying to find a new explanation. For example, suppose a patient is brought into the emergency room of a city hospital after an automobile accident; if his urine contains blood, the doctor should surely attribute it to abdominal trauma, rather than considering GLOMERULITIS.

2. Triggering is a process by which an elementary hypothesis makes the transition from the inactive to the active state. A subset of the symptoms which are relevant to a disease are marked as triggers. When a symptom is asserted to be present in the current case, it activates all those elementary hypotheses for which it has been designated a trigger. For example, DYSURIA (painful urination) triggers URINARY-TRACT-INFECTION; NAUSEA by itself triggers nothing, as it is a common finding in many disorders.

3. Each elementary hypothesis has an associated local evaluation function which produces a value representative of how likely the disease is to be present given the data. Each of the hypotheses currently active is evaluated, taking the new finding into account. The evaluation done at this stage is local in that the functions do not ask questions about the status of other

elementary hypotheses, or consider symptoms other than those relevant to the disease hypothesis being evaluated.

4. The purpose of global assembly, the fourth stage, is to arrange the various local elementary hypotheses into a larger structure which is both coherent and adequate. The rules of coherence have to do with the ways to connect various elementary hypotheses through links like CAUSE, COMPLICATION and EVIDENCE. An adequate hypothesis is one which accounts for all the abnormal findings in a case and is the end goal of a diagnostic process.

Local Evaluation

Any hypothesis-based theory needs a method for evaluating an elementary hypothesis - in this case, a single disease or syndrome - in isolation. Correlations between symptoms and diseases are the major determinants of such evaluation. One such correlation is the conditional probability of a symptom given a disease. I have called such numbers EXPECTATIONS. In Bayesian terms they are $P(S/D)$ (read "the probability of S given D"), where S is the symptom and D the particular disease in question. These correlations, however, are all disease-centered; that is, they spring directly from the description of a disease. More useful diagnostic information is symptom-centered, since a diagnosis proceeds from symptoms to diseases.

This other more sophisticated type of information is what I have termed EVIDENCE; in Bayesian terms, it is the conditional probability of a disease given a symptom, or $P(D/S)$. The complexity of the transformation of information from EXPECTATIONS to EVIDENCE (see <Feller>) makes plausible the idea that part of a doctor's expertise lies in the translation of knowledge from the disease-centered mode to the symptom-centered mode.

A local scoring algorithm must take into consideration both positive and negative contributions to the current hypothesis. In general the presence of relevant symptoms (EVIDENCE) will add to the validity score of an elementary hypothesis, while their absence (VIOLATED EXPECTATIONS) will subtract from it. The presence of FEVER will add to the validity of STREP-INFECTION, while its absence will subtract. The scoring algorithm we developed takes into account both positive and negative evidence for a hypothesis and scores are normalized by being divided by their highest possible total score.

The scoring algorithm also takes into account differing strengths of properties, the inclusion of ISA links and symptoms associated with a hierarchy of diseases, and age and sex of the patient, and operates with only four levels of EVIDENCE (SUFFICIENT, STRONG, MODERATE, WEAK) and four of EXPECTATION (NECESSARY, STRONG, MODERATE, WEAK). Details of the algorithm can be found in <Rubin>, but more important to the present discussion are the aspects

of local evaluation which serve to reduce the number of hypotheses actively being considered at any given time.

A hypothesis may be activated by one of its triggers. A hypothesis which is not active is not being currently considered or evaluated. This selective activation of hypotheses is one way to control the number of diseases being actively considered at any time. Notice that this use of triggers is certainly a heuristic device, since the diagnosis for the particular case on hand may not be one of those triggered.

A second way of minimizing hypotheses is to reject unlikely ones. One method of rejection is to include in a disease's slice rejecting symptoms whose presence precludes that disease's existence. For example, the presence of RED-BLOOD-CELL-CASTS rules out the diagnosis of SICKLE-CELL-TRAIT.

In addition, we can assign a priori probabilities to diseases. The age and sex of a patient affect this probability profoundly. Combining age, sex and disease leads to a useful number representing the probability of the disease occurring in a patient of particular age and sex. If this number is especially low, we may consider it 0 for heuristic purposes and put the hypothesis on the DEFERRED-LIST.

Interaction

The theory presented so far has been a linear one. Such a theory assumes that subparts of a problem can be treated independently and the solutions to those subproblems combined without alteration. Of course, there are interactions among symptoms which contradict the linearity hypothesis. The three examples of non-linearity described below illustrate some ways of dealing with interactions.

Both HEMATURIA and PROTEINURIA are EVIDENCE for GLOMERULITIS and G-U-TRACT-BLEEDING. However, their relative severities differ in these two hypotheses. In G-U-TRACT-BLEEDING, we expect the ratio of HEMATURIA to PROTEINURIA to be near that in whole blood; for HEMATURIA GROSS we expect PROTEINURIA LIGHT. In GLOMERULITIS, on the other hand, there should be relatively more PROTEINURIA than in G-U-TRACT-BLEEDING; for PROTEINURIA MODERATE, we would expect HEMATURIA MICROSCOPIC or LIGHT. The approach I have taken to this interaction is to specify for each disease or state which combinations would rule it out. Thus (AND (HEMATURIA GROSS) (PROTEINURIA LIGHT)) precludes GLOMERULITIS, while (AND (HEMATURIA LIGHT) (PROTEINURIA HEAVY)) precludes G-U-TRACT-BLEEDING.

The ASLO (anti-streptolysin-O) titer often rises several weeks after a person has had a STREP-INFECTION, indicating that the body is fighting the infection with antibodies. Taking PENICILLIN to combat the infection, however, often squelches the antibody response. If a doctor were actively considering STREP-INFECTION, ASLO-TITER

(RESULT NORMAL) would represent a violated expectation. An excuse is sometimes available for the absence of an expected finding; in this case PENICILLIN (STATUS TAKEN) would excuse a normal ASLO-TITER. The STREP-INFECTION hypothesis is evaluated as if ASLO-TITER were not a relevant symptom when penicillin has been taken.

Two or more diseases may resemble each other in many of their crucial aspects; it is particularly important to be able to tell them apart. Besides being a possible pitfall in causing misdiagnoses, findings which are shared among diseases can also be used heuristically to avoid activating an undue number of hypotheses. Suppose diseases A and B share findings X, Y and Z, but are differentiated by Q's occurrence in A but not B. If X and Y are present, we can consider B but not A, provided there is also a piece of heuristic information (called a DIFFERENTIAL-DIAGNOSIS) which activates A and rejects B if Q is discovered.

Global Assembly

Most of the diagnoses at which doctors finally arrive are not represented by a single elementary hypothesis. Patients often have more than one related or even totally unrelated diseases. A final diagnosis may be GLOMERULITIS COMPLICATED-BY NEPHROTIC-SYNDROME or FGN and HYPERTENSION ESSENTIAL. Clearly we need some way to discover and specify these more complex hypotheses as well as to combine pathological states which are themselves elementary hypotheses into a larger hypothesis which postulates a single cause for all of them. These concerns are handled by the global assembly stage of processing, which puts together coherent and adequate hypotheses. The processes in this stage are described in terms of matching a pattern (a template) and performing some action on the basis of that match.

A coherent hypothesis consists of two or more elementary hypotheses joined by "coherence links" which include ISA, CAUSE, COMPLICATION-OF, DEVELOPS-INTO and EVIDENCE links. Coherent hypotheses are constructed out of already-active hypotheses and, perhaps, some inactive ones as well, which are activated in the course of constructing the larger hypothesis. Each type of coherent hypothesis can be represented as a "template" which is placed on the patient's data structure; present findings, active hypotheses and necessary links are the structures which must match. The action taken when a template matches consists of joining the matched components together, along with newly-activated hypotheses. Two coherent hypothesis types are described below.

Symptoms are often not connected directly to their diseases, but to intermediately-general pathological states. It is thus important to be able to reconnect the symptoms to the actual disease; this is done via EVIDENCE-chained hypotheses. They are formed when two or

more active elementary hypotheses have EVIDENCE chains which intersect at a single etiology. For example, if SODIUM-RETENTION and ACUTE-RENAL-FAILURE were both active, we would want to unify them into a hypothesis which postulated AGN. See Diagram 2 for an illustration of a template for this type of hypothesis.

Forming a coherent hypothesis with CAUSE, COMPLICATION-OF or DEVELOPS-INTO links may not involve activating any new elementary hypotheses at all. If two active hypotheses are connected by a CAUSE, COMPLICATION-OF or DEVELOPS-INTO link, they may be joined into one composite hypothesis. More interesting is the case where a new elementary hypothesis must be activated. A template for this situation is contained in Diagram 3. CLOTTING-DISORDER is activated in order to provide the link between HEMATURIA and PREGNANCY. In words, we can activate an intermediate hypothesis which has one (non-trigger) finding present and is also connected by a CAUSE, COMPLICATION-OF or DEVELOPS-INTO link to an active hypothesis. The newly-activated hypothesis provides a link between the two nodes.

Ultimately, the global assembly stage must come up with an adequate hypothesis. The primary characteristic of an adequate hypothesis is that it accounts for all the abnormalities noted, while maintaining as much simplicity as possible. An adequate hypothesis consists of several independent parts, each of which is a coherent hypothesis. Each component must also be an ULTIMATE-ETIOLOGY or, in the case of more complex coherent hypotheses, it must contain some ULTIMATE-ETIOLOGY. In addition, all accepted elementary hypotheses must be subsumed in the final diagnosis, either by themselves, or as part of a larger coherent hypothesis. For example, the following is an adequate hypothesis:

LGN
(DURATION (YEARS 10))
HYPERTENSION ESSENTIAL
(DURATION (YEARS 5))
FAMILY-HISTORY NEPHRITIS

Notice that the second component of this hypothesis is HYPERTENSION ESSENTIAL, not HYPERTENSION CHRONIC; this is because only HYPERTENSION ESSENTIAL is marked as an ULTIMATE-ETIOLOGY. HYPERTENSION CHRONIC is a symptom, not an explanation, while HYPERTENSION ESSENTIAL is an explanation (actually the admission that no other explanation has been found!) Essentially, a process which builds adequate hypotheses must partition the symptoms into possibly non-disjoint subsets and account for each subset with some coherent hypothesis. Strategies for this partitioning (see <Rubin>) contribute further to the process of minimizing the number of active hypotheses.

Summary

The area of medical diagnosis has been investigated as an A.I. problem and a

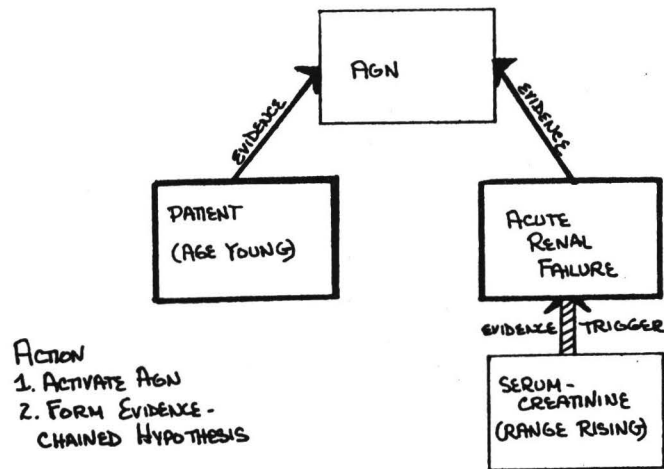
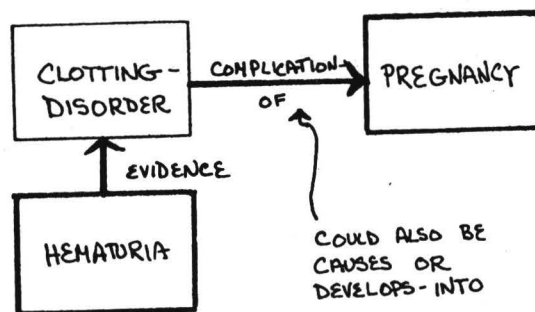


Diagram 2: EVIDENCE-CHAINED TEMPLATE



ACTION:
 1. ACTIVATE CLOTING-DISORDER
 2. FORM COMPLICATION-OF CONNECTED HYPOTHESIS

Diagram 3: GENERAL CAUSES, COMPLICATION-OF, DEVELOPS-INTO TEMPLATE

structure for medical knowledge proposed. The process postulated to act on that structure has as one of its goals the minimization of the number of hypotheses actively considered at any one time. Local vs. global evaluation and linearity vs. interaction have also been studied within the context of the theory.

References

1. Fahlman, Scott, "A Hypothesis-Frame System for Recognition Problems," WP-57, Artificial Intelligence Lab, M.I.T., Cambridge, December, 1973.
2. Feller, William, An Introduction to Probability Theory and Its Applications, Vol. 1, John Wiley and Sons, Inc.: New York, 1968.
3. Gorry, G. A., A System for Computer-Aided Diagnosis, TR-44, Project MAC, M.I.T., Cambridge, September 1967.
4. Gorry, G. A. et al, Computer Laboratory for Clinical Decision-Making, research proposal to NIH, March 1974.
5. "A Framework for Representing Knowledge," in The Psychology of Computer Vision, P.H. Winston, ed., McGraw-Hill: New York 1975.
6. Rubin, Andee, "Hypothesis Formation and Evaluation in Medical Diagnosis," AI-TR-316, Artificial Intelligence Lab, M.I.T., Cambridge, January 1975.
7. Winograd, Terry, "Frame Representations and the Procedural/ Declarative Controversy," in Bobrow and Collins (eds.), forthcoming.