Expert System Profile

GENERAL

Domain: Medical
Main General Function: Diagnosis
System Name: INTERNIST-I/ CADUCEUS (or INTERNIST-II)

Dates: 1970's –1980's

Researchers: Ph.D. Harry Pople, M.D. Jack D. Myers & Randolph Miller
Location: University of Pittsburgh

Language: InterLisp
Machine: <UNKNOWN>

Brief Summary:

When this program was created all other expert medical diagnosis program focused their attention on one small subsection of the medical domain. INTERNIST was first program that attempted to become an expert assistant for a large chunk of the internal medical domain. The program was created to help medical professionals do their job.

Related Systems:

Present Illness Program (PIP) by Szolovits and Pauker was created in 1976. Also deals with large set of data with many possible separate hypothesis and findings. It uses categorical and probabilistic reasoning mechanisms that INTERNIST-I and CADUCEUS don't.

CATEGORY TWO

Characterization of Givens:

The information accepted by the system comes in two types:

1. *Database* – medical database constructed by the authors from papers and years of watching and talking with medical consultants. (Consistent through all instantiation of the program)
2. *Patient's Condition* – Condition of the patient both positive and negative findings. (Changes with each new user and task)

Characterization of Output:

The system produces a list of diseases that that patient has along with which symptoms are present for each of the diseases.

Characterization of Data:

The data is reliable but not complete. Consists of 500 diseases, with 3550 different manifestation. This covers 70-75% of the internal medical domain. The data is organized in different manners for INTERNEST and CADUCUES. Through both keep the data in a hierarchal format.

Generic Tasks:

These systems do classification of diseases and synthesis of hypotheses that are used to make the end classification.

Theoretical Commitment:

INTERNIST and CADUCEUS use a very *ad hoc* method of diagnosis and synthesis. In the attempt to model the way in which doctors actually do diagnosis. Their scoring method of competing hypothesis is not based any confidence factors or probabilistic model but more likelihood.

Reality:

This system was designed to emulate the way in which doctors actually do diagnosis. Doctors keep a hierarchy of interrelated diseases and their symptoms. They also keep track of which diseases and manifestation commonly occur and how these different diseases interact (one causing another, or knowing that it is unlikely that they occur in parallel).

CATEGORY THREE

Completeness:

Both INTERNIST–I and CADUCEUS (INTERNIST-II) have been implemented.

Use:

The system has been used by real users from outside the original development team but has not been used in a real working environment.

Performance:

There are performance evaluations for INTERNIST-I but not for CADUCEUS. The data for this test was taken from cases in the New England Journal of Medicine in 1969 and were all from the Massachusetts General Hospital.

	No. of Instances		
Category	**INTERNIST-I**	**Clinicians**	**Discussants**
Definitive, correct	17	23	29
Tentative, correct	8	5	6
Failed to make correct diagnosis	18	15	8
Definitive, incorrect	5	8	11
Tentative, incorrect	6	5	2
Total incorrect	11	13	13
Total no. Errors in diagnosis	29	28	21
Total possible diagnosis	43	43	43

The clinicians are the actual doctors in the hospital, the discussants are a panel of doctors that later went through the cases and preformed their own diagnosis on the files that the clinicians wrote. INTERNIST-I and the discussants got the same information that was provided by the clinicians. INTERNIST-I did ok getting the best results in 3 out of the 7 categories it was defiantly completive with the clinicians and the discussants. The authors saw two main problems with the INTERNIST-I system, the first was with the knowledge base and its representation and the other was with the actual implementation of the program. The authors decided to tackle some of these problems with CADUCEUS, they refined the data structure and modified the search procedure.

CATEGORY FOUR

Phases:

Although the system works with one main task in mind, there are some partial sub-phases of the system. There is the hypothesis formulation phase where the system looks at all of the manifestation that the patient has and then tries to formulate different hypothesis as to why they are present. In the next phase of the system the different competing hypothesis are explored in-order to select the best set of hypothesis that fully describe the patient. In this phase the system can as question of the user in-order to distinguish between a set of competing hypothesis.

Sub-functions:

Besides from classification the system does synthesis possible explanations of the cause of the patient's illness(es).

Use of Simulation or Analysis:

The system does not really simulate any procedure or analysis.

System/Control Implementation Architecture:

The overall architecture is a hierarchy of knowledge that is used to diagnoses from manifestations. Each node in the hierarchy had numeric values that describe its importance and interactions with its surrounding nodes. (see more about this later)

CATEGORY FIVE

Characterization of Structure Knowledge:

There are two main groups of knowledge pathophysiological(causal) and nosological . The causal knowledge stores information about disease interaction with manifestation (symptoms, signs, history) of the patient. This information is used in INTERNIST-I. While the nosological information is a hierarchy of body, how different body parts interact with each other. For CADUCEUS a hierarchy is made that combines the nosological structure with the causal.

There are two different data types:

1. Manifestations:
    - Are the symptoms, signs, history and lab values of a patient, these are used to help identify what condition the patient is in.
    - The manifestation representation in the data structure stores two types of information the **IMPORT** value and the **TYPE**. The **IMPORT** value, this value is responsible for showing, independent of a disease, how important a specific manifestation is, on a scale from 1 to 5. Zero meaning unimportant and five meaning very important. The higher the IMPORT number the more the system will try to find the cause of this manifestation. Manifestations can have one of several **TYPEs**: patient-history, symptoms, signs, and lab values. These values help the system determine what type of questions to ask the user because expensive and evasive procedures should be avoided at all costs. This is implemented into the system to the point that depending on how many competing hypotheses there are for a specific set of manifestations the system is only allowed to ask certain types of questions. For example if there are more then 5 competing hypothesis the system can only ask question about the patients history and easily identifiable symptoms, where as when there is only 1 hypothesis left to prove the system can request lab experiments to be preformed only if all other questions have been explored.
2. Diseases:
    - Have two describing factors **FREQUENCY** and **EVOKING STRENGTH.**
    - **FREQUENCY** – represents on a scale of 1 to 5 how often this specific manifestation is the cause the disease. 1 is occurs rarely and 5 is occurrence is essential.

- **ENVOKING STRENGTH** – relates how strongly a particular manifestation relates to the disease. "Given a patient with this findings how strongly should I consider this diagnosis to be its explanation?" This is on a scale of 0 to 5. Where 0 is nonspecific – manifestations occurs too commonly to be used to construct a differential diagnosis to 5 manifestations is pathognomonic for the diagnosis.

Characterization of Process Knowledge:

The knowledge plays an important role to the point of TYPE, IMPORT, FREQUENCY and ENVOKING STRENGTH. Also the links between these nodes are important because they represent how and why one thing is caused by another.

Deep or Surface:

The data structure and the information that it is describing are quite complex. The representation of how these two different types of node interact through caused-by links is arguably deep knowledge. But because the system is using the numeric range values to determine how important the next node is it almost seems like surface knowledge.

CATEGORY SIX

Search Space:

INTERNIST-I searches through the space of the causal hierarchy, while CADUCEUS searches through the space of the combine nosological and causal hierarchy. The search space is explicit and so are the states. The states represent manifestations and diseases (described above). The space is huge with 500 diseases and 3500 manifestations each of which can me connected to any number of nodes. The systems do a process of hypothesis and refinement in order to classify the disease(s) the patient has.

Space Traversal:

The space is traversed from the manifestations (that the patient has) to the disease(s). A link is followed when there is enough information to distinguish it from its siblings. This is a gradual refinement process that allows the system to determine exactly what disease(s) the patient may have.

Search Control Strategy:

State space search with the ability to compare states and generate & test. The state space search is just and AI technique that seems to fit, while generate & test is more like what an actual doctor would do.

Standard Search Strategies:

State space search with the ability to compare states and Generate & Test methods are used.

Search Control Characterization:

Knowledge-base

Sub-problems:

It is imperative for the system to be able to evaluate partial solutions and compare it with other possibilities. Because states represents how well a particular path represents the true cause of the persons illness with out this continual partial evaluation the system would never be able to find the solution. Besides from a comparison between two competing hypothesizes that explain the same set of manifestations the system also has the problem of creating multiple hypothesis so that in the end all positive manifestation of the patient will be explained. In INTERNIST-I one hypothesis is independent of any others until the end when interactions are then tried to be determined. In CADUCEUS interaction between already proven hypothesis and the current hypothesis are worked into the scoring method of the system, because it is assumed more likely that a patient will have manifestations do to one cause as opposed to two or more. Meaning that a patient is assumed to have one disease that may or may not cause other diseases instead of having many non-interacting diseases.

Search Control Representation:

The search control knowledge is expressed explicitly by the IMPORT, TYPE, FREQUENCY, and EVOKING STRENGTH values that are stored in each node.

Search Control Strength:

Strong: Very domain dependent and knowledge-full method.

CATEGORY SEVEN

Failure Method:

The system really cannot fail, it can reach an incorrect conclusion for the given manifestations or it can make a tentative answer if it cannot prove for certain that a specific disease is the cause of the patient's illness. It does not have any really situation in which recovery is necessary.

Uncertainty:

The system attempts to prove its uncertainty about hypotheses in order to make a diagnosis. As CAUDCEUS proves that one disease a definite illness of the patient this information is then used to help guide future diseases that also might be present.

Management of Uncertainty:

The system uses an *ad hoc* scoring method to rank the sets of hypotheses so that it knows which set of manifestation that it should try to prove first. To do this it used the IMPORT, FREQUENCY and EVOKING STRENGTH values from section six.

$$Score = \sum_{k \in explained\_Manifestations} EVOKING_k + Bonus - \sum_{j \in Unexplanied\_patientManifestations} IMPORT_j - \sum_{i \in Unexplained\_diseaseManifestations} Frequency_i$$

The score is calculated by adding all the EVOKING STRENGTH values for all manifestations that the patients has and are explained by the disease and subtracting from this all of the values that are not explained, both disease manifestations that the patient does not have and patient manifestations that the disease does not cover. Some bonus points are added if a disease that has already been proven is used to help prove another hypothesis. All IMPORT, FREQUENCY and EVOKING STRENGTH values are weighted as follows:

EVOKING STRENGTHS: 0 = 1pt, 1 = 4pts, 2 = 10pts, 3 = 20pts, 4 = 40pts, 5 = 80pts
FREQUENCY: 1 = 1pts, 2 = 4pts, 3 = 7pts, 4 = 15pts, 5 = 30pts
IMPORT: 1 = 2pts, 2 = 6pts, 3 = 10pts, 4 = 20pts, 5 = 40pts
Bonus: 20pts

When score of a hypothesis is 90 points above its competitors it is considered to by the solution. The search procedure tries to get this spread by asking questions to first narrow down the number of hypothesis and then to widen the spread between two competitors.

The apparent problem is that all of the numbers are made up and do not have any substantial reason for being what they are besides from the fact that is seems to work decently well. Especially the threshold cut off value of 90 points sometimes this value make it hard for the correct diagnosis to be asserted.

Management of Time:

There is no time dependent data. This does affect the system because it cannot take time into account. For example the system cannot understand a progression of a disease over time.

CATEGORY EIGHT

Knowledge Representation Method:

Non-strict hierarchy – causal and nosological

Knowledge Representation Generality:

The system uses InterLisp and does **not** provide a tool for building expert systems.

Knowledge Structure:

Is a non-strict hierarchy, because a manifestation can cause many diseases and a disease can have many manifestations. This is true for nodes in the middle as well, causing a one-way network of information. Because the authors were trying to represent how doctors diagnose, this knowledge structure is representative of how the authors think doctors represent this knowledge.

CATEGORY NINE

Alternative Representations:

Alternative representations of the same piece of data are not allowed but because the data is not stored in a strict hierarchy a node can be linked to many parents.

Alternative Solution Methods:

There are many paths to find the right diagnosis for a patient but because of the structure and ordering of the scoring method and how the system ask questions generally only one path is available. Although the system is robust enough to be able to find alternative paths when one obvious path is not allowable (either because of an expensive/invasive procedure is requested or because information is just not available).

Optimization:

The system is not optimal, but it will always produce the best answer that it has.

Multiple Results:

The system will produce more then one results in two ways. The first because a patient will have more then one disease (so each disease will be presented as a result). Second because the system may not be able to differentiate between two competing diseases in this case all and their explanations will be presented to the doctor.

CATEGORY TEN

Interaction:

There are two parts of the diagnosis that interact with the user. The manifestation input phase, where the doctor will input the positive and negative manifestations that the patient has. The other is during the search for the correct diagnosis. When the system does not have enough information to make a decision it will ask the user to input information about the general area that it is having a problem with, the user can respond or type GO which tells the system to prompt the user with specific questions. The answers to questions are YES, NO or N/A (not available). When responding to the general area question the user can enter in any new manifestation that he knows.

Data Collection:

The system does not require all information at one time, and it does not need complete information to make a diagnosis (this would be infeasible). The system will vary the types of questions that it asks based on the number of hypothesis that it is differentiating between. If there is more then 5 competing hypothesis it will try to ask question that will eliminate groups of similar hypothesis and it is only allowed to ask question that will not involve invasive or expensive procedures to answer. With 2 to 4 hypothesis to differentiate the system will ask question that will find similarities that will quickly eliminate some and improve the results of others. With 2 hypotheses the system will try to ask questions that will greatly differentiate the scores of the two. When only one hypothesis is left the system will try to prove all of the other manifestations of the disease to the best of its ability.

Data format:

The data is in two types, patient data and the hierarchical knowledge base. Patient data has both positive and negative manifestations of the patient inputted in text format into the InterLisp system.

Acquisition:

The system can ask the user about other results not initially given to it that will help determine the exact cause of the disease. It guides the user to enter information about a particular area, if none is given the system will prompt with questions. It can validate its findings by trying to prove for sure all of the positive manifestations of the patient. It has no way of validating information entered by the user.

Learning:

The system does not learn from its own performance.

Explanation:

The system can explain where its results came from, but only to the point of showing which manifestations and sub-disease caused the resultant disease. This is obtained from partial trace of the program during execution.

CATEGORY ELEVEN

Strengths:

The authors of this system did a good job of creating a medical diagnosis assistant expert system that covers more then a very small area of medicine. In fact INTERNIST-I and CADUCEUS cover a large area of internal medicine. Another strength is that the system did relatively well in the test cases that were presented to it. Some problems in INTERNIST-I were solved in CADUCEUS to create a program that would solve this difficult problem with less trouble.

Weaknesses:

The fact that this system does not deal with temporal data is a major weakness. Also Pople was not happy with its ability to explain why a particular diagnosis was selected.

Other: