Diagnostic Strategies in the Hypothesis-Directed PATHFINDER System

Eric J. Horvitz, David E. Heckerman, Bharat N. Nathwani*, Lawrence M. Fagan

Medical Computer Science Research Group
Heuristic Programming Project
Stanford University
Stanford, California 94305

*Section on Hematopathology
Department of Pathology
University of Southern California
Los Angeles, California

Abstract

PATHFINDER is a developing expert system to assist pathologists in the interpretation of findings noted on microscopic examination of lymph node tissue. We describe PATHFINDER's hypothesis-directed reasoning approach with an emphasis on intelligent question selection strategies, techniques for managing data inaccuracy, and explanation methods for justifying questions asked of the pathologist. Although this work was originally inspired by the INTERNIST-1 approach to hypothesis scoring and question selection, we have made several modifications to the INTERNIST-1 approach in building an expert system for pathology.

Introduction

PATHFINDER is a hypothesis-directed expert system for the diagnosis of lymph node pathology based upon the appearance of microscopic features in lymph node tissue. Major questions addressed in our research include the investigation of (a) alternative methods for combining evidence in support of possible hypotheses, (b) useful diagnostic problem-solving strategies, and (c) appropriate question justification schemes. In this paper, we discuss our research on strategies for refining a list of plausible disease hypotheses through the generation of appropriate questions and the management of potential data inaccuracy. We also describe our current approach to question justification.

Motivation for the expert system

A working expert system that could give general pathologists ready access to the diagnostic reasoning capabilities of experts within the lymph node pathology domain would be a useful clinical innovation. Over 30,000 new cases of lymphoma (malignancies of the lymphatic system) are reported each year in the United States. As most lymphomas have a distinct natural history and specific therapy, precise diagnosis is crucial. Unfortunately, the diagnosis of lymph node disease is often error-prone and is considered one of the most difficult tasks in pathology. Experts in the diagnosis of lymph node pathology (expert hematopathologists) make more accurate diagnoses than community-based general pathologists. Several studies have shown that experts generally agree with one another whereas the diagnoses offered by general pathologists may be changed in as many as 50% of the cases receiving secondary review by experts[1].

This work was supported by gifts from the Henry J. Kaiser Family Foundation and the Josiah Macy, Jr. Foundation. Computing resources were provided by the Hewlett Packard Company, through a gift to the Stanford Medical Information Sciences Training Program, and by the SUMEX-AIM Resource under NIH Grant RR-00785.

The problem

As in other areas of medicine, a central problem-solving task within the lymph node pathology specialty is the classification of sets of symptoms into disease categories. Expert hematopathologists estimate that approximately 150 low- and high-magnification microscopic features are used in reasoning about lymph node diseases. There are over eighty lymph node diseases, fifty of which are malignant. The malignant diseases include primary malignancies (lymphomas) that arise from cells of lymphatic origin and secondary malignancies caused by metastatic invasion of the lymph nodes. The great majority of abnormal lymph node sections are attributed to thirty different benign diseases ranging from mononucleosis to rheumatoid arthritis. A source of difficulty in this domain is that many of the benign lymph node diseases closely resemble the malignant lymphomas in appearance.

The choice of methodology

Two symbolic reasoning approaches were considered in the initial design stages of an expert system for lymph node pathology: the INTERNIST-1 hypothetico-deductive approach[2] and the MYCIN rule-based production system approach[3]. Early informal "process tracing"[4] involving discussions with experts and observation of expert diagnostic protocols suggested that the diagnosis of lymph node pathology often involves iterative hypothesis refinement. The hypothesis-directed model of physician problem-solving explicit in INTERNIST-1 attracted our attention as a potentially useful way to simulate diagnostic strategies in pathology. INTERNIST-1 is an expert system for internal medicine initiated at the University of Pittsburg 10 years ago[2]. It is the core of a continuing research program, CADUCEUS[5].

The method of sequential diagnosis

PATHFINDER and INTERNIST-1 are based on the method of *sequential diagnosis*[6]. With this approach, a set of salient disease manifestations are initially presented to the program. A list of plausible disease hypotheses (a differential diagnosis) is then formulated based on these manifestations and questions are selected that can that can help narrow the number of diseases under consideration. After the user answers these questions, a new set of hypotheses is formulated and the process is repeated until a diagnosis is reached. The method of sequential diagnosis is hypothesis-directed in that the questions are selected by strategies or *modes* that consider a current list of hypotheses. The INTERNIST-1 approach to the method of sequential diagnosis uses several diagnostic strategies for selecting questions in conjunction with a set of heuristics for making decisions about using the alternative strategies.

The method of sequential diagnosis is an advancement on older Bayesian statistics programs that require all relevant

CH2107-1/84/0000/0630$01.00 © 1984 IEEE

findings in a patient case at once to make an accurate diagnosis. As there are often dozens of relevant clinical findings, the Bayesian approach has been considered less suitable for application in a clinical setting than systems using the method of sequential diagnosis[6].

In PATHFINDER, a crucial task in the formulation of differential diagnoses is the assignment of a numerical score to each disease based on the given disease manifestations. We will not discuss details of scoring or formation of the differential in this paper. The scoring procedures make use of expert estimates of the associations between diseases and disease manifestations. In PATHFINDER and INTERNIST-1, there are two estimates attached to each disease-manifestation pair: an ***evoking strength*** and a ***frequency***. The ***evoking strength*** for a disease-manifestation pair answers the question: "If I see a case with this finding, how strongly should I consider this disease to be its explanation?" The ***frequency*** is an estimate of how often the disease under consideration is associated with the finding. The evoking strength and frequency estimates are similar to the formal statistical concepts of predictive value and sensitivity.

Disease manifestations in lymph node pathology are microscopic features. In PATHFINDER, ***features*** are each subdivided into a mutually exclusive and exhaustive list of ***values***. For example, the feature ***pseudofollicularity*** can take on any one of the values ***absent***, ***slight***, ***moderate***, or ***prominent***. Features are evaluated by the selection of a value that reflects the severity of the feature. We refer to a particular feature and value as a ***feature-value***. Every feature-value in the PATHFINDER knowledge base is associated with an evoking-strength and frequency for each disease.

Problems with the method of sequential diagnosis

The INTERNIST-1 effort demonstrated the utility as well as the limitations and problems with the method of sequential diagnosis in internal medicine. After summarizing some problems with INTERNIST-1, we will discuss each in the context of lymph node pathology.

While INTERNIST-1 was shown to reason admirably within defined problem areas using general diagnostic strategies, deficiencies were identified in the system's inability to consistently focus on an accurate set of diseases[5]. Specifically, INTERNIST-1 researchers have identified four problems:

1. The program performs poorly when several disease processes coexist and single manifestations can be explained by more than one disease.
2. Inappropriate questions are sometimes asked because of a lack of explicit knowledge about problem-solving strategies.
3. Diagnoses are sometimes incorrect because of a lack of deep pathophysiological and anatomical knowledge.
4. Diagnoses are sometimes incorrect because important information may be discarded when there is a preponderance of unimportant information.

Applying the method of sequential diagnosis to lymph node pathology

We believe that the INTERNIST-1 approach is in several aspects more appropriate for reasoning about lymph node diseases than it is for the domain of internal medicine. We will now describe PATHFINDER's relative immunity to the first three problems listed above. Later, in our discussion of PATHFINDER's ***confirmation mode*** strategy for selecting questions, we will address the fourth issue.

1. **The multiple disease problem:** Within internal medicine, the presence of concurrent diseases in a single patient is not uncommon. Two or more diseases can co-exist in lymph node pathology as well. However, they are by definition located in ***spatially discrete*** areas in one or more lymph node specimens from the same patient. Although it is understood that co-existing lymph node diseases could grow to become adjacent and continue to spatially admix, there is no concept of spatially superimposed diseases. Experts believe that the resulting morphological patterns are too difficult to classify. Since PATHFINDER can process each area of lymph node tissue separately, it is free from the multiple disease deficiencies of INTERNIST-1.
2. **Diagnostic problem solving strategy:** INTERNIST's developers attribute the generation of poor questions in part to the system's deficiency of explicit knowledge about problem-solving strategies in internal medicine. Diagnostic strategies in internal medicine are complex. There exists no simple diagnostic strategy that can be used to guide problem formulation. Unlike internal medicine, the domain of lymph node pathology appears to have a diagnostic strategy based on a simple disease nosology. The current version of PATHFINDER uses this nosology in question selection. This will be discussed in detail below in the section on PATHFINDER methods for question selection.
3. **Deep models:** We mentioned above that INTERNIST-1 does not use deep models of disease processes in clinical problem-solving. Evaluators of INTERNIST-1 have speculated that use of more complete knowledge about causal relations could raise the competence of later versions of INTERNIST-1[5]. While this is undoubtedly true in the domain of internal medicine, lymph node pathology apparently makes little use of deep causal or structural models; knowledge relating the complexity of observed lymph node patterns to simplifying fundamental principles is sparse. As deep models seem to have little importance in reasoning about lymph node pathology, the addition of deep models would likely yield little or no improvement in PATHFINDER's performance.

In summary, properties of the lymph node pathology application area that minimize the problems attributed to the INTERNIST-1 method of sequential diagnosis are the domain's lack of recognizable superimposed diseases, the existence of a simple globally-applicable problem-solving hierarchy, and the lack of accepted deep models of lymph node diseases.

PATHFINDER hypothesis-directed question selection strategies

As mentioned above, a central feature of sequential diagnostic systems is their ability to pose questions that direct the user to collect useful data. We will now describe PATHFINDER hypothesis-directed strategies for selecting questions that can maximally reduce the uncertainty in the differential diagnosis.

PATHFINDER dynamically applies different modes depending on the nature of the current differential diagnosis list. The current system uses three different modes for question selection. We call these methods ***group-discriminate mode***, ***entropy-discriminate mode***, and ***confirmation mode***. We will describe these modes, the heuristics used in the decision to apply them, and the motivation for their development.

Early PATHFINDER question selection strategies

Earlier versions of PATHFINDER used questioning modes similar to those found in INTERNIST-1. The modes used in INTERNIST-1 are named *pursual mode, discriminate mode,* and *rule-out mode.* The selection of modes in INTERNIST-1 is a function of the number of diseases and corresponding scores of the diseases on the differential diagnosis list. If there is only one disease on the differential diagnosis, INTERNIST-1 immediately concludes the disease as the final diagnosis. Otherwise, INTERNIST-1 determines those diseases which are "close" to the disease with the highest score. Two diseases are "close" if their scores differ by less then some carefully chosen threshold. Depending upon how many diseases are close to the leader, INTERNIST-1 enters one of the three modes:

1. **Pursual mode:** If there are no diseases close to the leader, INTERNIST-1 *pursues* the leading disease. That is, the user is asked to report on patient findings that would confirm the diagnosis. These findings have associated high evoking strengths for the disease being pursued.
2. **Discriminate mode:** If one to four diseases have scores close to the score of the leading disease, INTERNIST-1 enters *discriminate mode.* In discriminate mode, questions are asked which best discriminate between the *two* leading contenders.
3. **Rule-out mode:** If more than three diseases have scores close to the score of the leading disease, INTERNIST-1 goes into *rule-out mode.* In this mode, for each disease on the differential diagnosis, questions are asked that would tend to eliminate further consideration of the disease.

The first several versions of PATHFINDER used the pursual and discriminate modes. We did not implement a rule-out mode because the pathologist on our team believes that hematopathologists very rarely attempt to prune the differential diagnosis list by methodically ruling out unlikely contenders. This is based in the frequent occurrence of very large differential diagnoses in lymph node pathology.

Later PATHFINDER question selection strategies

After some experimentation with pursual and discriminate modes, we decided to eliminate pursual mode in favor of a *confirmation mode* which is described below. Our experimentation also revealed the utility of posing questions that could discriminate among more than two diseases. We thus began to explore new diagnostic strategies that could consider the entire set of hypotheses in the differential diagnosis. We first implemented *focus mode* which selected questions that could rule out the largest number of diseases on a differential diagnosis. Later, focus mode was replaced by the more general *entropy-discriminate mode.* The testing of entropy-discriminate mode uncovered an important hypothesis-specific strategic hierarchy used by experienced pathologists in the diagnosis of lymph node pathology. Integrating the strategic hierarchy into entropy-discriminate mode yielded *group-discriminate mode.* We will discuss the new modes and the reasoning behind their development.

PATHFINDER *Focus mode* As the testing of early versions of PATHFINDER progressed, deficiencies in the discriminate mode strategy became apparent. For example, if a differential diagnosis contained several small cell, intermediate-sized cell, and large cell diseases, our initial version of PATHFINDER would use discriminate mode to identify questions that only discriminated between the top two diseases rather than proposing that the user answer a more globally discriminating question about the size of the cells that is seen to be predominating. In such cases, concentrating on just two diseases is considered inappropriate by expert hematopathologists. This strategy is viewed as being too narrowly focused at the top of the differential diagnosis. These considerations led to the formulation and implementation of *focus mode.*

Focus mode considers all of the diseases on the differential diagnosis. The mode selects questions about the status of features which will tend to minimize the number of diseases that remain on the differential diagnosis after values for these features are reported. The mode operates by estimating the number of diseases that will remain on the differential diagnosis when the user reports a feature and value. The mode then selects those features which yield the smallest final disease estimates.

The expected number of remaining diseases for a feature is calculated by summing, over all values for that feature, the product of:

- the probability that a value will be reported, and
- the number of diseases that will remain if that value is reported.

The probability that a value will be reported is estimated by taking the average of the frequency of that value for each disease on the current differential diagnosis. This assumes that each disease in the differential diagnosis is equally likely. Although usually not the case, this assumption is used to help select questions that discriminate equally well among the diseases on the differential diagnosis. The pathologist in our group believes that this is an appropriate strategy for this domain.

The exact number of diseases remaining on the differential diagnosis after a feature and value are reported is not calculated explicitly because the necessary computation time is unacceptable. Instead, focus mode uses the approximation that only those diseases on the differential diagnosis that are assigned a feature-value frequency score of 0 by experts will be removed from the differential diagnosis if that feature and value is reported.

The focus mode strategy has a simple decision-theoretic interpretation. If we define the *utility* of a question to be proportional to the number of diseases removed from the differential diagnosis after the question is answered, we are simply looking for the question with the maximum utility. This definition of utility is reasonable in that smaller numbers of diseases on a differential diagnosis reflect more certain knowledge about the diagnosis. Thus, a question with a high utility corresponds to a question that can lead to a large increase in certainty.

The basis and behavior of the mode was well-accepted by expert hematopathologists. However, it had a minor problem that was traced to the use of a threshold on frequency values to determine the number of diseases remaining on the differential diagnosis. If a feature has one or more values that are seen with low (but not zero) frequency in many diseases on the differential, focus mode will overlook this feature as a useful question even though evaluating the feature could highly disfavor many diseases on the differential. This problem with focus mode led us to the formulation of the PATHFINDER *entropy-discriminate mode.*

PATHFINDER *Entropy-discriminate mode* In focus mode, the utility of a question is related to the number of diseases removed from the differential diagnosis. This change in number of diseases on the differential diagnosis reflects a decrease in the uncertainty of the differential diagnosis.

Entropy-discriminate mode makes use of a more general notion of uncertainty used in information theory[7]. In this mode, a quantity called *entropy* is used as the measure of uncertainty. The use of entropy as a measure of uncertainty in the differential diagnosis was inspired by a program developed by Gorry for diagnosis of acute renal failure[8]. In this context, entropy is equal to the sum of $p_i log p_i$ over each disease in the differential diagnosis where p_i is the probability that a disease is the diagnosis. Instead of calculating the number of diseases remaining on the differential diagnosis when a particular value is reported for a feature, we calculate the entropy of the resulting differential diagnosis. For a potential question, we must calculate for each value, the probability that each disease will be on the final differential diagnosis. We use Bayes' theorem to calculate the probabilities and assume that each disease on the initial differential diagnosis is equally likely (as we do in focus mode).

The entropy discriminating mode was judged by experts to be an improvement over focus mode. However, there were still occasions when the questions were not viewed as optimal.

PATHFINDER *Group-discriminate mode* Preliminary testing showed that the rationale for questions selected by the powerful entropy-discriminate mode was often not easily *understood* by experts. Although these questions were undoubtedly the most discriminating among the diseases on the differential diagnosis, we found that they were not *natural* questions to ask in relation to the problem-solving protocol followed by the human expert.

The pathologist on our team uses a simplifying problem-solving strategy for managing complexity in the domain. While entropy-discriminate mode selects questions that best discriminate among all diseases on a differential diagnosis, we found that our expert reasons about small numbers of diagnostic *categories*. For example, if there are benign and malignant diseases on a differential diagnosis, our pathologist deems most appropriate those questions that best discriminate between the benign and malignant groups rather than questions that might best discriminate among all of the diseases. If there are only primary malignant and metastatic diseases on the differential diagnosis, the pathologist will attempt to discriminate between the primary malignancy and the metastatic categories. Our pathologist finds questions that discriminate among various natural groupings of diseases on the differential diagnosis to be more understandable than the question that could best discriminate among all the diseases.

Our pathologist's diagnostic strategy can be described by traversing a hierarchy of disease categories. The strategic hierarchy, pictured in figure 1, is a binary tree of disease groups. The hierarchy can be used to group the differential diagnosis at various levels of refinement. It is interesting to note that several previous studies of medical problem-solving have identified similar decision trees in other medical domains[4].

The existence of the categorical reasoning strategy led to the formulation of *group-discriminate mode*. For a given differential diagnosis, the group-discriminate mode identifies the most specific grouping possible and then selects questions that best discriminate among the diseases as a group. This mode is analogous to the INTERNIST-1 discriminate mode except that groups of diseases, rather than diseases are discriminated.

The pathologist on our team is pleased by the behavior of group-discriminate mode. According to him, the mode selects *reasonable* questions at each point in the problem solving session. Our experience with group-discriminate mode supports results of previous physician attitude surveys that have demonstrated the importance of understandable behavior in an expert system[9]. Of course, a drawback of group-discriminate mode is that the differential diagnosis refinement process will not always proceed as quickly as with entropy-discriminate mode.

In addition to generating more acceptable questions, this newest mode has other benefits. We have found that it is easy to justify questions offered by group-discriminate mode. This should not be surprising in light of the simplifying restrictions on question selection imposed by this mode. Also, this mode executes much faster than its entropy-discriminating counterpart.

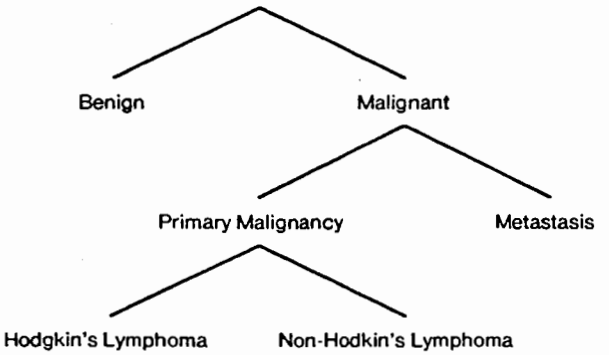

Figure 1: Lymph Node Disease Strategic Hierarchy

PATHFINDER *Confirmation mode* Recall that the INTERNIST-1 pursual mode selects questions that tend to confirm the putative diagnosis. These questions are those with high evoking-strengths for the disease under consideration. In place of pursual mode, we have designed *confirmation mode*. This new mode is similar to pursual mode in that questions are asked when a single disease is being considered as the diagnosis. Confirmation mode differs from pursual mode in that the questions selected are marked by an expert pathologist as being "important" to the disease under consideration. For example, the presence of Sternberg-Reed cells is an important feature of Hodgkin's disease. If Hodgkin's disease is to be concluded by the program and Sternberg-Reed cells has not been evaluated, PATHFINDER will ask the user to evaluate this feature.

Confirmation mode uses expert knowledge as a safeguard that counters the tendency of the method of sequential diagnosis to focus the differential diagnosis. In terms of computational problem solving, this mode will tend to minimize the "foothill problem" associated with hill-climbing algorithms[10]. Such a mode serves a similar function as the traditional *review of systems* used by physicians when interviewing a patient. One objective of the traditional medical review of systems and PATHFINDER confirmation mode is to ensure that the conclusions reached are not completely off target. A confirmation strategy addresses those cases in which a preponderance of unimportant information leads an INTERNIST-like expert system to a inaccurate diagnosis.

PATHFINDER heuristics for mode application

In reasoning about a case, PATHFINDER first attempts to classify diseases on the differential diagnosis into two groups at the most specific level of the strategic hierarchy for diagnosis of lymph node diseases. If two disease groups can be ascertained, group-discriminate mode is applied to the differential diagnosis. If there are two or more diseases on the differential diagnosis and all the diseases can be classified in a group at one leaf of the strategic hierarchy, entropy-discriminate mode is applied. Finally, if one disease remains on the differential, confirmation mode is applied. If the same single disease remains at the completion of confirmation mode, the disease is concluded as

the diagnosis.

While a pathologist in training or an expert interested in PATHFINDER's reasoning might desire understandable questions selected by group-discriminate mode, others may be more interested in reaching a diagnosis as quickly as possible. For this reason, we allow the user the option of disabling group-discriminate mode. When group-discriminate mode is disabled, entropy-discriminate mode is applied in its place. Although, the questions generated in many cases by the entropy-discriminate mode may not be understandable, they are often superior in their ability to discriminate among all the disease hypotheses. Such questions usually lead the user to a diagnosis more quickly than those selected by group discriminate mode.

Managing the collection of unreliable data

The discriminating power of questions is not the only important factor in question selection. INTERNIST-1, for example, considers the cost and invasiveness of tests. We do not consider invasiveness; in lymph node pathology, invasion occurs only at biopsy. Within the lymph node pathology domain, we found it useful to consider the ***reliability***, ***tediousness***, and ***expense*** associated with the evaluation of a particular feature. PATHFINDER uses these factors in combination with the utilities assigned to potential features by the question selection modes to generate a final utility for each question. The final utilities are used for selecting the best questions to ask. The details of the PATHFINDER cost-benefit utility equation will not be discussed here. We will only mention the reliability factor as it relates to a technique for minimizing the collection of unreliable data.

The collection of accurate data is especially important in reasoning about lymphatic pathology. Discussions with expert hematopathologists about the expertise-dependent problems with the recognition and quantification of lymph node features suggested that the questions selected by PATHFINDER at each point in a case should be tailored to different levels of user expertise. We have found that the need for customizing question-generation and inference to a user's expertise to minimize the collection of inaccurate data is an important yet relatively unexplored issue. Such methods could enable medical expert systems to adapt the selection of questions to maximize the accuracy of the inference process.

PATHFINDER considers the level of expertise of the system user in generating questions. After ascertaining the user's level of expertise, PATHFINDER uses three sets of expert estimates of ***expected reliability*** to modify the questions asked of users having different levels of expertise. The ***expected reliability*** is the amount of ***trust*** placed by an expert in the values reported for each feature by system users in different expertise groups. We believe that the expert's reliability estimate is a conglomeration of the relative frequencies of false positives and false negatives for both the recognition of a feature and for the selection of a feature-value pair. Sets of expected reliability estimates for the expert hematopathologist, non-expert pathologist, and pathologist in training are used by PATHFINDER's utility equation. The different sets of estimates enable PATHFINDER to first offer questions to users that they are most likely to answer reliably. The current version of PATHFINDER prompts the user for his level of expertise and incorporates the appropriate set of feature reliability values in the question utility equation. We are currently exploring techniques to dynamically assess the reliability of data detecting inconsistencies in input data.

Justification of question selection

Surveys of potential users of medical advice systems have suggested that the capability of the system to explain its reasoning strategies may be one of the most important factors determining its eventual clinical acceptance[9]. Unfortunately, explanation systems for frame-based systems like PATHFINDER are uncommon. For example, INTERNIST-1 only informs the user about the current question-generating strategy in progress. More specific reasons as to why a particular finding is being requested are absent.

We have experimented with several question justification schemes. We first implemented a free text system that evaluated the questions selected in terms of discriminating power, tedium, reliability, and cost. Early versions simply reported that a question was "good" or "very good" for discriminating among diseases on the differential diagnosis based on the computed utility. We later decided to offer the system user more specific information about the discriminatory abilities of a question. The present version of PATHFINDER displays information about the relative impact of alternative responses on a particular differential diagnosis in a graphical format.

Figure 2 depicts the justification offered by PATHFINDER when two diseases or disease groups are being considered. In this sample case, the feature ***architecture*** has been recommended by the system as having the ability to refine a particular differential diagnosis (not shown) containing a number of benign and malignant diseases. The positions of a set of asterisks is used to indicate the degree to which each group of diseases is favored by each possible feature value. In the example below, the values ***preserved*** and ***partially obliterated*** strongly support diseases on the differential diagnosis that are in the benign group, while the value ***completely obliterated*** strongly supports the malignant disease hypotheses. The value ***greatly obliterated*** is not very useful for discriminating between the two disease groups. If more than two diseases from one category are on the differential diagnosis, a justification format lists, for each possible value, the diseases that will likely become strong contenders if that value is reported. This format is displayed in the sample PATHFINDER transcript in the appendix. The graphical justification formats have been extremely useful in offering insight about the impact of questions on a differential diagnosis and in facilitating the refinement of the knowledge base.

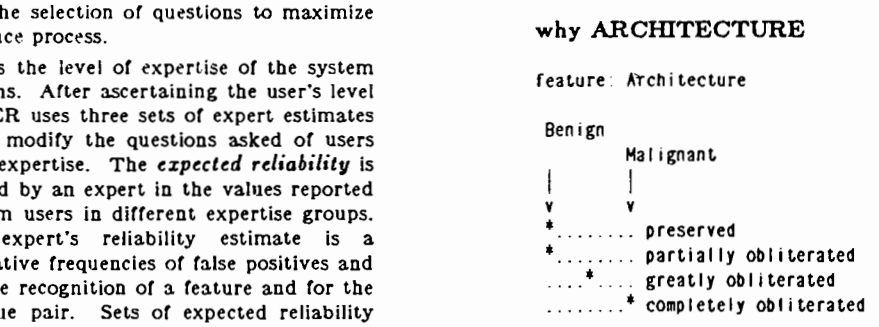

Figure 2: Sample Graphical Justification

Summary

In this paper, we have presented several aspects of our research on the PATHFINDER expert system. Our conclusions are:

1. It is possible to successfully apply the INTERNIST-1 approach to the method of sequential diagnosis to a smaller, more specialized domain. The domain is in many aspects more appropriate than the originally intended application area of internal medicine.

2. When selecting questions to refine the differential diagnosis, it is often necessary to be able to discriminate among more than two diseases.
3. The most discriminating question for refining a differential diagnosis list is not necessarily the "best" question to ask. Such highly discriminating questions are often difficult to understand by system users and difficult to justify. Expert hematopathologists seem to use a simple problem-solving hierarchy. Such a hierarchy can be used to select questions that are easily understood.
4. A strategy for question selection in the spirit of the medical "review of systems" is useful to ensure that a viable disease hypothesis has not been overlooked. This is a safeguard that counters the focusing tendency of the method of sequential diagnosis.
5. It is useful for an expect system to avoid requests for data that can be expected to be unreliable. In this regard, knowledge about the expertise of the user is important in question selection.
6. Useful justification for selected questions is possible in a frame-based system.

Acknowledgements

We are grateful to Costan Berard, Jerome Burke, and Ronald Dorfman for the expert advice they have provided to the PATHFINDER project. We also thank Curt Langlotz, Mark Musen, Glenn Rennels, and Edward Shortliffe for their comments on early versions of this paper. Finally, we thank Edward Shortliffe and Bruce Buchanan for their helpful insight and guidance.

Appendix

Sample PATHFINDER Transcript

Morphologic features as reported from a random slide obtained from the repository archives. Comments are in *italics* and text entered by the user is in **bold face.**

```
Welcome to Pathfinder, Version 22 -- 8/18/84.

Please enter your expertise:
0. RESIDENT
1. NON-EXPERT
2. EXPERT

-> 1
```

The user enters values for salient features seen through the microscope.

```
-> Follicles number

Follicles number
  0. 0
  1. 1-5
  2. 6-50
  3. 51-75
  4. 76-90
  5. >90
-> 5

The number of diseases on the differential is 5.

-> ask
```

Typing **ask** *tells the program to select questions that refine the differential diagnosis. Since there are benign and malignant diseases on this differential, group-discriminate mode is used.*

```
discriminating:
 Malignant
  Small cleaved, follicular lymphoma
  Mixed, small cleaved and large cell,
    follicular lymphoma
  Large cell, follicular lymphoma
  Small noncleaved, follicular lymphoma
 Benign
  Florid reactive follicular hyperplasia

I recommend that the following features be evaluated:
  8. Follicles mantle zones
  2. Follicles density
 13. Follicles (majority) predominating cells (>50%)
 18. Comparison of cytology inside and outside the
       follicles

-> why

 Which feature do you want justified?
-> 8

The following table elucidates the discriminating
power of this feature. The position of the asterisk
indicates which of the two groups of diseases is
favored by each value.

 Malignant
         Benign
|        |
v        v
*........ absent
....*.... present
........* prominent
```

The following graph displays the three cost factor estimates (on a 0 to 9 scale). The estimates impact on the final utility assigned to each potential question.

```
Pragmatic considerations in the evaluation of the
 feature, Follicles mantle zones:

 .?....... unreliability
 ...!..... tedium
 $........ cost

-> 8

Follicles mantle zones
  0. absent
  1. present
  2. prominent
-> 0

The number of diseases on the differential is 4.

-> ask
```

Here, entropy discriminate mode is used because the differential consists solely of non-Hodgkin's lymphomas.

```
discriminating:
 Small cleaved, follicular lymphoma
 Mixed, small cleaved and large cell,
   follicular lymphoma
 Large cell, follicular lymphoma
 Small noncleaved, follicular lymphoma

I recommend that the following features be evaluated:
 84  What is the size of the cells that predominate?
 13  Follicles (majority) predominating cells (>50%)
 96  Medium sized lymphoid cells -- number
109  Small lymphocytes -- nuclear shape
 17  Are the follicular center cells monomorphic wrt
       chromatin structure?

-> why

 Which feature do you want justified?

-> 84

The following table elucidates the discriminating
power of this feature.  For each value, the
diseases that will likely become strong contenders
are listed.

feature: What is the size of the cells that
           predominate?

none
 Mixed, small cleaved and large cell,
   follicular lymphoma
small (<12u)
 Small cleaved, follicular lymphoma
medium (12-20u)
 Small noncleaved, follicular lymphoma
large (>20u)
 Large cell, follicular lymphoma

Pragmatic considerations in the evaluation of the
  feature, What is the size of the cells that
  predominate?

  ..?..... unreliability
  ......!.. tedium
  $....... cost

-> 84

What is the size of the cells that predominate?
  0  none
  1  small (<12u)
  2  medium (12-20u)
  3  large (>20u)
-> 1

The only disease remaining on the differential is
 Small cleaved, follicular lymphoma.
```

As there is only one disease left on the differential diagnosis, confirmation mode is applied.

```
confirming:
 Small cleaved, follicular lymphoma
```

References

1. Kim H., et. al., "Pathology Panel for Lymophoma Clinical Studies: A Comprehensive Analysis of Cases Accumulated Since its Inception," *Journal of the National Cancer Institute*, Vol. 68, 1982, pp. 43-60.

2. Miller, R. A., Pople, H. E., and Myers, J. D., "INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine," *New England Journal of Medicine*, Vol. 307, No. 8, 1982, pp. 468-476.

3. Shortliffe, E. H., *Computer-based medical consultations: MYCIN*, American Elsevier, New York, 1976.

4. Elstein, A. S., Shulman, L. S., and Sprafka, S. A., *Medical problem solving: An analysis of clinical reasoning*, Harvard University Press, Cambridge, Mass., 1978.

5. Pople, H., "Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnostics," in *Artificial Intelligence In Medicine*, P. Szolovits, ed., Westview Press, 1982.

6. Gorry, G. A., and Barnett, G. O., "Experience with a model of sequential diagnosis," *Computers and Biomedical Research*, Vol. 1, 1968, pp. 490-507.

7. Shannon, C. E., and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Ill., 1949.

8. Gorry, G. A., Kassirer, J. P., Essig, A., and Schwartz, W. B., "Decision analysis as the basis for computer-aided management of acute renal failure," *AJM*, Vol. 55, 1973, pp. 473-484.

9. Teach, R. L. and Shortliffe, E. H., "An analysis of physician attitudes regarding computer-based clinical consultation systems," *Computers and Biomedical Research*, Vol. 14, 1981, pp. 542-558.

10. Winston, P. H., *Artificial intelligence*, Addison-Wesley, Reading, Mass., 1977.