An Expert System for Assigning Patients into Clinical Trials Based on Bayesian Networks

Costantinos Papaconstantinou,[*] Georgios Theocharous,[†] Sridhar Mahadevan [‡]
H. Lee Moffitt Cancer Center & Research Institute
University of South Florida
Tampa, Florida 33612

November 24, 1997

Abstract

Assigning patients into clinical trials is a knowledge and data intensive task. Eligibility determination for entry into a clinical trial is based upon specific inclusion and exclusion criteria. This paper investigates the use of an expert system to assist the physician through this task. This expert system uses Bayesian networks, a probabilistic method that can take advantage of pre-existing statistical knowledge. The paper also describes the feasibility of such a system by presenting the implementation of three clinical protocols. The experimental results reveal that the approach is feasible. The system gives correct eligibility scores when all evidence is available but also predicts eligibility when there is missing evidence. The system directs the physician to the protocols the patient is most eligible for, according to the current evidence. The system has the ability of learning its prior and conditional probabilities (expert knowledge) from training examples.

[*]From the H. Lee Moffitt Cancer Center & Research Institute, University of South Florida, Tampa Florida
[†]From the H. Lee Moffitt Cancer Center & Research Institute, University of South Florida, Tampa Florida
[‡]From the Computer Science Department Of Michigan State University, East Lansing, Michigan

1 Introduction

A clinical trial is a method used to evaluate new procedures and treatments for a specific population of patients.[1] Each trial is designed to answer questions about the validity and effectiveness of new treatments and how they could be improved.

Recruitment of patients into clinical trials involves examining their medical records and matching their characteristics to specified eligibility criteria. Eligibility criteria are typically described in a protocol document and include demographic information, laboratory test results, subjective symptoms, physical findings, current or prior medications and therapies, and drug allergies. An example of some inclusion/exclusion criteria for a particular breast cancer protocol is given in Figure 1.

Achieving adequate rates of accrual to a protocol is critical to the successful completion of a clinical trial. The eligibility screening process however, can be inefficient and extremely time consuming. Currently, the H. Lee Moffiitt Cancer Center & Research Institute participates in approximately 150 open trials; each trial has 10 to 25 inclusion and exclusion criteria. As such, it is difficult for a physician to determine the trials to which a patient may appropriately be assigned. Due to the complexity of this task, it is also very likely that a patient may not be considered for all protocols for which s/he is eligible; therefore, theoretical accrual rates are not usually achieved. To further complicate matters, the information needed to determine eligibility may not always be readily available.

The primary motivation for this research was to establish a more efficient method of assigning patients into clinical trials using an expert system.[2] Assigning patients into clinical trials is a knowledge and data intensive task that involves uncertain domain knowledge.[3] Eligibility determination for entry into a clinical trial is based upon specific inclusion and exclusion criteria. This paper describes an attempt to solve the problem of assigning patients into clinical trials using a Bayesian-based expert system.[4] The Bayesian approach was selected because it is a probabilistic method that takes advantage of pre-existing statistical knowledge. Our expert system utilizes Bayesian propagation and can also learn the conditional and prior probabilities of a Bayesian structure. A World Wide Web (WWW) interface to our expert system was developed. The expert system was tested with actual protocols that contained the inclusion and exclusion criteria for entrance into clinical trials.

The test results reveal that the approach is feasible. The system gives correct eligibility scores when all evidence is available, but also tries to predict eligibility when there is missing evidence. The system directs the physician to the protocols the patient is most eligible for, given the current evidence. The system has the ability to learn its prior and conditional probabilities from training examples.[5]

In the past, few researchers have dealt with the problem of efficient assignment of patients into clinical trials. A qualitative and a probabilistic approach to compute and summarize the eligibility status of potentially eligible patients for clinical trials is described in a paper by Tu et. al.[3] These two approaches are compared, and a synthesis that draws on the strength of each approach is proposed. The work presented however, implements only the qualitative approach.

Another system that deals with the problem of assigning patients into clinical trials is

Criteria for Eligibility (All responses must be yes)

Yes No

1. ___ ___ Has the patient been diagnosed with primary invasisve adenocarcinoma of the breast ?

2. ___ ___ Was an axillary dissection performed and at least 6 axillary lymph nodes removed and examined ?

3. ___ ___ Were 0-3 axillary lymph nodes positive for tumor?

4. ___ ___ Is the disease high-risk per at least one of the following? (1). Both ER and PgR negative and tumor greater than 1.0 cm in greatest diameter. (2). Tumor greater than 2.0 cm in greatest diameter (any receptor status). (3). Tumor involves 1, 2, or 3 axillary lymph nodes (any diameter of primary)

5. ___ ___ Will a minimum of one block of primary tumor or a letter from the pathologist explaining that sufficient tissue is not available be submitted per section 15.0 within 18 months of registration?

6. ___ ___ Is the patient currently free of breast cancer?

Criteria for Exclusion (All responses must be no)

Yes No

1. ___ ___ Does the patient have pure tubular, mucinous, or papilarry carcinoma

2. ___ ___ Does the patient have recurrent or invasive cancer

3. ___ ___ Did the patient have locally advanced breast cancer at diagnosis?

4. ___ ___ Were there any gross or microscopically positive surgical margins noted in the final surgery or pathology reports?

5. ___ ___ Has the patient had prior chemotherapy for this breast cancer or any systematic therapy of any type for a previous breast cancer?

Figure 1: A Sample Protocol For Breast Cancer

described in a paper by Ohno M. et.al.[6] This system combines belief networks with a classification of domain knowledge. It uses a method of propagation of missing values utilizing a qualitative approach (similar to the one used by Tu et. al[3]) that combines the specific inclusion and exclusion criteria for a particular protocol and produces an eligibility score for that protocol. The eligibility criteria are then classified into categories that prioritize the order of pursuing missing protocol data.

In our opinion, a simpler approach employs one large Bayesian network. The order in which missing information should be obtained could be implicitly implemented with the correct assertion of relations (arcs) between variables and strengths (conditional probabilities) between them. If we simply choose that the next criterion to be satisfied is the one that will change the eligibility score the most, then we could arrange the variables in an ancestral ordering, such that important variables are closer to the eligibility node. This is the approach presented in this paper.

2 Materials and Methods

2.1 Bayesian Networks

Theory Bayesian theory can be used as a way of representing uncertainty in expert systems. Bayesian theory is based on the primary axioms of probability theory. The cornerstone of Bayesian probability theory is the inversion formula due to Bayes (Equation 1) which provides a convenient way of updating belief in a proposition in light of new evidence.

$$P(B_r|A) = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)} = \frac{P(B_r)P(A|B_r)}{P(A)} \quad for \quad r = 1, 2, \ldots, n. \tag{1}$$

Where $B_1, B_2, \ldots, B_n$ are mutually exclusive events of which one must occur. A is an event that depends on B events.

However, simply using the invertion formula induces conditional independence assumptions between variables and also requires large numbers of subjective estimates of conditional probabilities. These difficulties are simplified by the use of a causal model. Building a causal model between variables (draw arcs as cause and effect) greatly reduces the conditional independence assumptions and the number of probabilities that need to be asserted.

The causal model, or belief network, is a directed acyclic graph where the nodes in the graph correspond to relevant and uncertain variables of the problem. The arcs in the directed acyclic graph represent direct probabilistic dependencies among the uncertain variables.[7] In particular, an arc from node A to node B reflects an assertion that the probability distribution of B depends on the instance of the variable A. We say that A conditions B. Thus, a node has probability distribution for every instance of its conditioning nodes. The lack of arcs in a belief network reflects the assertion of conditional independence (see Figure 2).

Given a Bayesian network we can construct the joint probability distribution of the variables. The joint probability distribution is achieved through the chain rule of probability

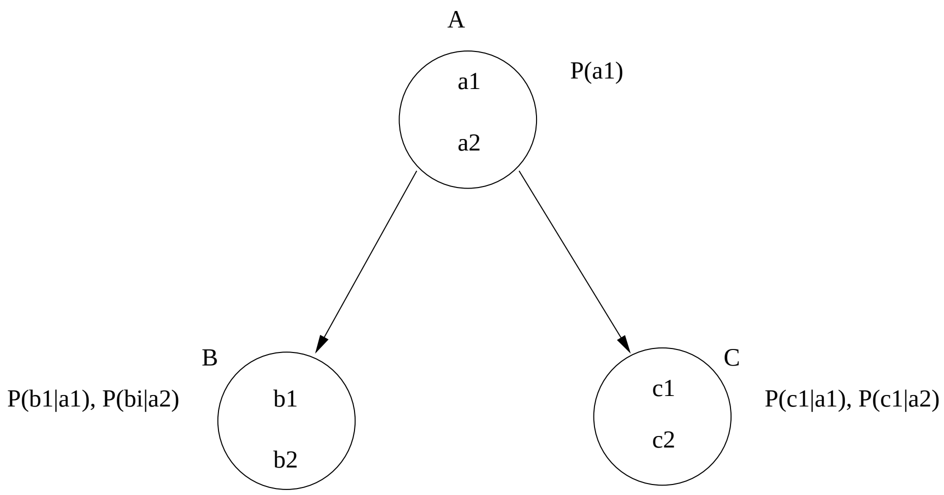

Figure 2: A Simple Bayesian Network: 3 nodes are shown in the figure, where each node has two states. Nodes B and C are depended on node A. The indicated probabilities need to be specified.

given in Equation 2.

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i|x_1, ..., x_{i-1}) \tag{2}$$

For Figure 2 the joint probability distribution is given in Equation 3, which can be simplified using the conditional independence relations from the network.

$$P(a, b, c) = P(a)P(b|a)P(c|a, b) = P(a)P(b|a)P(c|a) \tag{3}$$

Knowing the joint probability for a domain, we can calculate any probability of interest. For example the probability of $P(a|c)$ is given in Equation 4.

$$P(a|c) = \frac{P(a, c)}{P(c)} = \frac{\sum_b P(a, b, c)}{\sum_{ab} P(a, b, c)} \tag{4}$$

The subscripts of the summation indicate a summation of the entries in the joint probability table for which the particular subscripts change (the maximum value of the subscript is the number of possible discrete values of the variable) as the rest of the variables remain constant (have the values we are looking for). The index of the joint probability table is a factored representation of the possible values of each variable.

Nevertheless, using the whole joint probability distribution of a domain (a large table of all combinations of states of all variables) is not feasible because it involves the summation over an exponential number of terms. If there are n variables in the domain then there are 2^n possible distributions. If we take into account the simplifications due to the conditional independence assumptions (because of the causal structure) the complexity of the problem could greatly be reduced. We can rewrite the sums over multiple variables as the product of sums over a single variable or smaller number of variables (see Equation 5).

$$P(c) = \sum_{ab} P(a,b,c) = \sum_{a} P(a)P(c|a) \sum_{b} P(b|a) \tag{5}$$

There are various methods to inference on Bayesian networks that exploit the conditional independence described above. These methods include message passing,[7] arc reversal,[8,9,10] clustering,[11] and direct manipulations of sums/products.[12] For our system we used the method of clustering. The method of clustering, described by Neapolitan R. E.[13] and by Lauritzen and Spiegelhalter,[11] consists of the following basic steps:

1. Represent the problem domain as a Bayesian network.
2. Specify prior and conditional probabilities.
3. Form the moral graph. That is introduce connections between nodes with common children and drop the directions. The reason for forming the moral graph is that it is convenient to consider an undirected graph when there is the need to keep track of groups of variables (cliques).[11]
4. Triangulate the graph. Make sure that there are no cycles of length four or more without a chord (short-cut). The reason for triangulating the graph is to make sure that term expressing the joint probability distribution are function of cliques that contain complete subgraphs.[13]
5. Find the maximal cliques. Maximal cliques are the maximal complete subsets of the graph.
6. Create a tree of cliques. The tree of cliques is used for message passing among the cliques.
7. Represent the joint probability distribution as a function of cliques.
8. Propagate messages between cliques.
9. Absorb new evidence.
10. Go to step 8.

Learning It has been suggested by various researchers that a beta distribution can be used to quantify prior knowledge.[13,14] The beta function is given in Equation 6

$$\beta(a,b) = \frac{(a+b+1)!}{a!b!} x^a (1-x)^b \text{ where } a,b \geq 0 \tag{6}$$

Using Equation 7 we can calculate $P(v_1)$ if we know a and b. The a and b values can be thought of as a successes in $a+b$ trials. If, for example, V is a propositional variable with two states v_1 and v_2, then x represents the possible values of $P(v_1)$, the probability of variable V being in state v_1. If $f(x)$ is the probability density function for the possible values of $P(v_1)$, for certain a and b, that is, $f(x) = \beta(a, b)$, then the expected value of the distribution is given in Equation 7.

$$P(v_1) = \int_0^1 x f(x) dx = \frac{a+1}{a+b+2} \tag{7}$$

2.2 Knowledge Acquisition

This expert system implemented three breast cancer protocols as one large Bayesian network (see Figure 4). The justification for generating a single Bayesian network was twofold: 1) There were many inclusion/exclusion criteria that were common to more than one protocol; 2) There were criteria in one protocol that were directly related to criteria in another protocol.

With the assistance of our expert (a physician) we established the following knowledge acquisition procedure:

1. Examine all protocols to be implemented and create variable nodes. Most of the time an inclusion/exclusion criterion would generate a node itself. Complex criteria could be broken down to simpler random variables. Superficial variables can also be created that will act as intermediate nodes to combine related variables.

2. Find all the common nodes to protocols and draw arcs between them where there is a cause and effect relationship.

3. Look for nodes in individual protocols (inclusion/exclusion criteria) that can cause the common nodes, or are effects of the common nodes, and draw appropriate arcs between them.

4. Group the criteria using the model shown in Figure 3.

5. Ask the expert to provide as many conditional and prior probabilities as possible.

6. If the Bayesian network takes a long time to propagate data evidence, try to refine the network by decomposing nodes that have many parents. If the decomposition produces large directed cycles, try to rearrange the decomposed nodes as proposed by Heckerman and Breese.[15]

7. Divide the criteria into two classes:

   (a) *Class 1:* Answers to basic eligibility questions like diagnosis, age sex, etc, which do not usually change once specified.

   (b) *Class 2:* Answers to eligibility questions that contain routine laboratory tests, subjective evaluations, clinical symptoms, etc, which can change at a later time.

   Try to get answers for the *class 1* criteria first, since this would eliminate a large number of protocols and allow the system to do an initial propagation and protocol eligibility ordering.

The structure of the resulted network is shown in Figure 4.

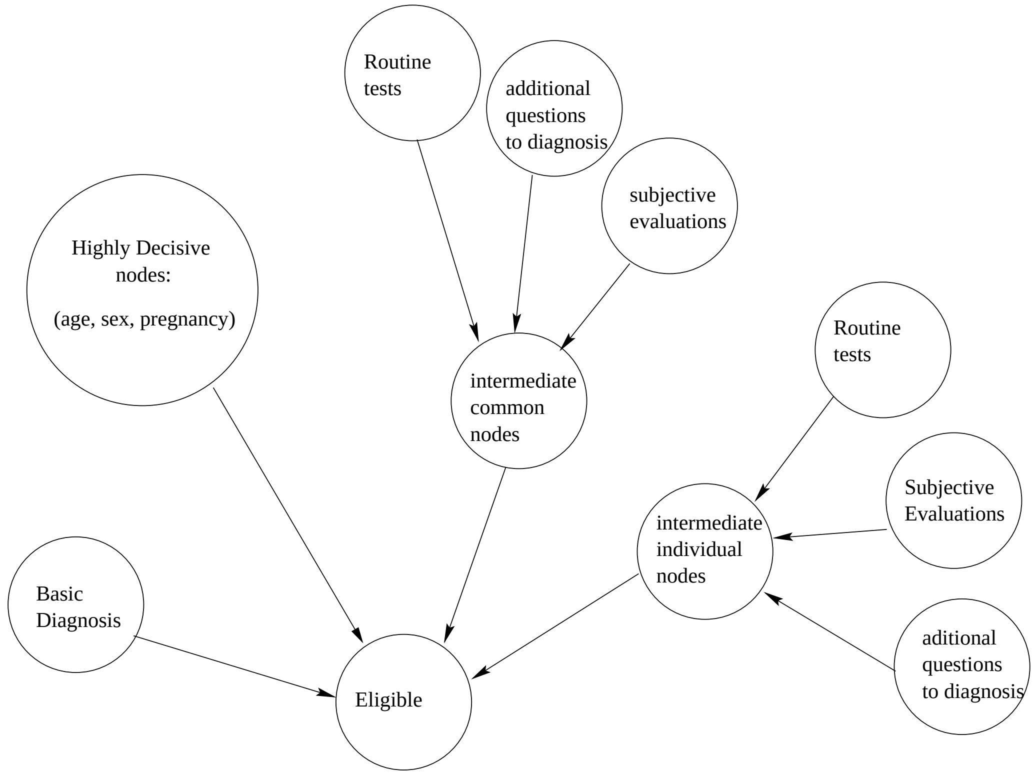

Figure 3: A Model Of A Protocol As A Bayesian Network: The criteria were devoded into fout basic categories as shown in the diagam.

2.3 Web Interface

The user interacts with the system via a World Wide Web (WWW) browser. A typical session is described below. The physician logs in to the system and s/he selects a patient to be evaluated by searching a list of patients. The search can be done by medical record number or by patient name (see Figure 5).

The system then retrieves facts known about the patient (basic eligibilty criteria like diagnosis, sex, age, etc) from external sources (admissions records, previous encounters records, etc). Using this pre-existing knowledge, the system does an initial propagation and outputs a patient eligibility score for each protocol available (see Figure 6).

At this time the user can evaluate the specific protocols. The user can select the protocol the patient is most likely to satisfy (Protocol with largest eligibility score. Eligibility scores are shown next to the protocol name as shown in Figure 6). When a protocol is selected for evaluation, all inclusion and exclusion criteria are presented to the user (see Figure 7). The existing probability assignment is shown in parenthesis next to each inclusion/exclusion criteria question choices. For example, if no evidence is available the probability for a yes/no answer choice could be 0.5 and 0.5, respectively. If evidence exists, these probabilities will be reflected. The user can now change facts and supply answers to questions. The system tries to guide the user by suggesting the next question to be specified. This question is determined by the system as the one that will most change the eligibility score at the time.

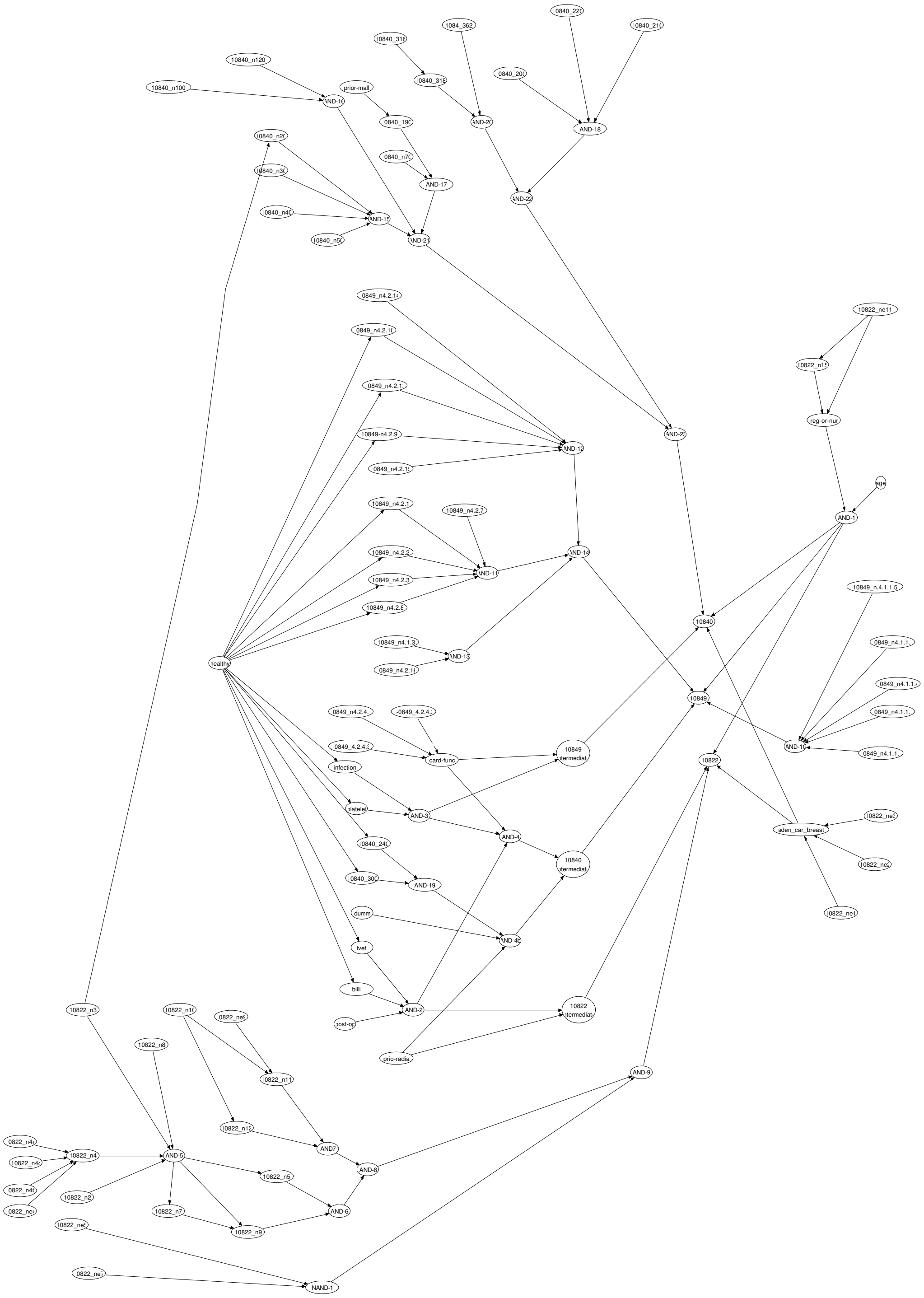

Figure 4: Structure of Bayesian Network Used

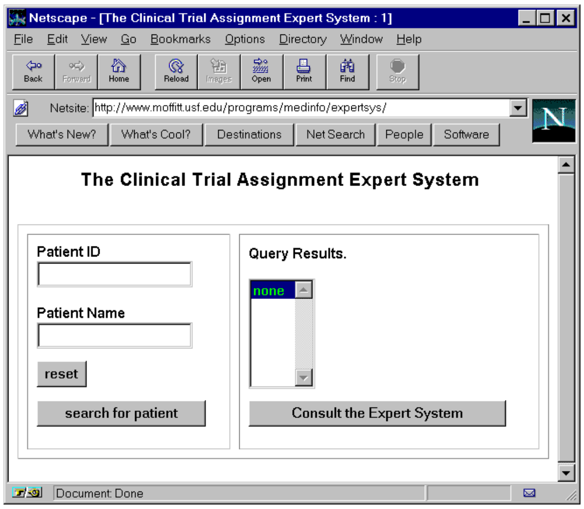

Figure 5: Web Interface for choosing a patient

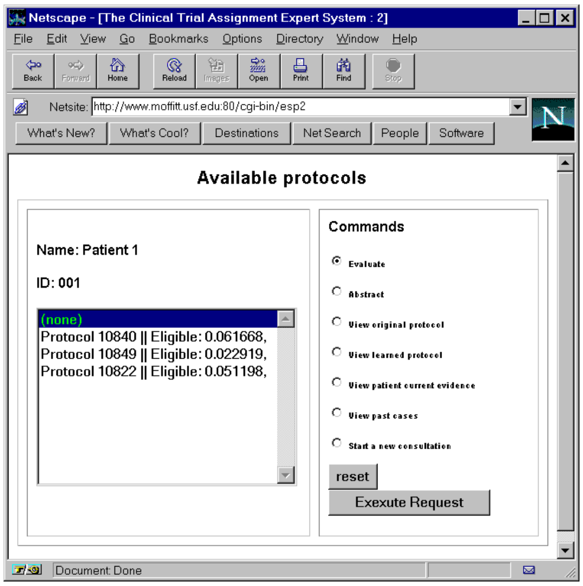

Figure 6: Web Interface showing the eligibility scores for each protocol available for a patient

The idea here is that if the user specifies the most relevant questions first, the system will converge quickly to an eligibility determination. The suggested question is highlighted in green color.

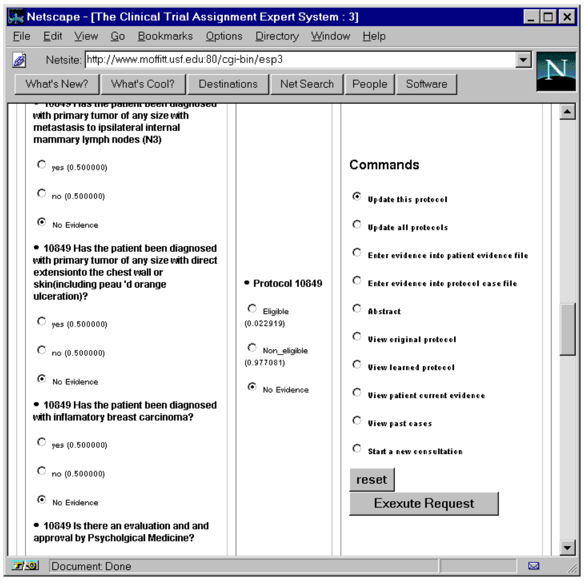

Figure 7: Web Interface showing the eligibility criteria of a particular protocol

The user can make any changes to the evidence about the inclusion/exclusion criteria and re-evaluate (propagate) to generate an updated eligibility score. As additional evidence is collected, the eligibility scores for all protocols are updated. Since protocols are related (share common facts; common nodes of the Bayesian network) changing the facts for one protocol may affect the eligibility score for one or more different protocols. The user may evaluate any protocol that looks promising. The user has also the option to use all of the evidence for a particular evaluation as a learning case for the Bayesian network.

3 Results

Experiments were designed to test for accuracy and usability of the system. The testing revealed the following:[16]

- For cases with no missing evidence, one of the following would occur:

    1. If there was strong evidence that the patient was eligible for the specified protocol, the system derived a high eligibility score (above 0.9).
    2. If there was strong evidence that the patient was not eligible for the specified protocol, the system derived a low eligibility score (below 0.5).
    3. If *class 1* criteria were satisfied (criteria that are unlikely to change) and some *class 2* criteria (criteria that could change, routine laboratory tests, symptoms,

etc) were not satisfied, the system would give eligibility scores between 0.5 and 0.7. This score implies that the protocol is still promising since the patient can become eligible if some *class 2* criteria are satisfied at some later time.

- For cases where information was missing, one of the following would occur:
    1. The system produced an eligibility score above 0.5 to indicate that the protocol looked promising.
    2. The system produced an eligibility score below 0.5 to indicate that protocol eligibility did not look promising.

When the user evaluates a protocol, the system attempts to determine which question, if answered next, would have the greatest effect on the eligibility score and suggests that question as the next question to be answered. Our testing showed that if the user follows this recommended sequence, then the eligibility score would converge faster than if other question sequences were used.

When eligibility score convergence is reached, additional answers do not substantially change the eligibility score. The system is currently unable to recommend the next question to be answered. When convergence is reached, the user can stop providing further evidence and still have a good indication of how promising the protocol eligibility is. Figure 8 illustrates eligibility score convergence. The top line represents the system suggested sequence of questioning. From the figure, we see that it converges more rapidly than other user defined sequences. A potential benefit of this observation is that a user can quickly evaluate many protocols to determine a set of protocols that are promising.

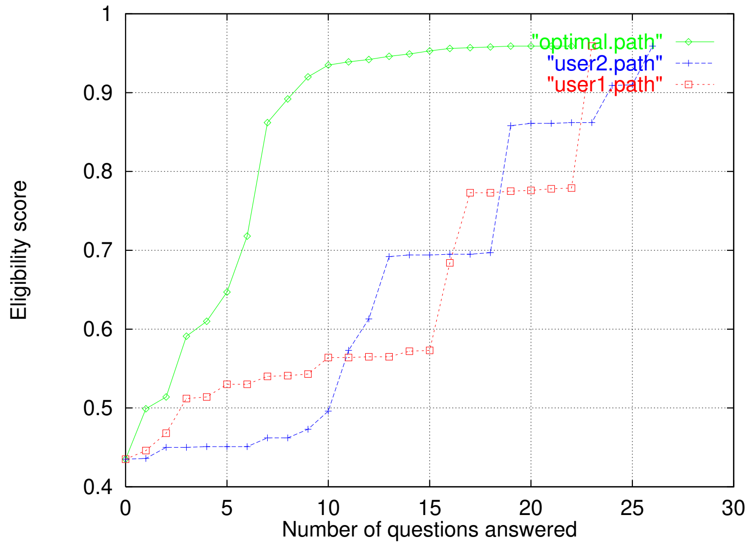

Figure 8: The convergence graph for protocol "10849"

The fact that the system converges early suggests that Bayesian networks can point out redundant questions and help refine the protocol questionnaire.

A test was also designed to investigate the learning capability of the system. Two Bayesian networks were used for this experiment. The first Bayesian network was defined with prior and conditional probabilities by the expert. The second network had the same structure (nodes and links) but most prior and conditional probabilities were set to 0.5 (only AND and OR nodes were left with their proper probabilities). Fifty training cases were used to train the second network. We observed that as more cases were presented for learning, the prior and conditional probabilities and the behavior of the trained network started to approximate the behavior of the first network.[16]

4 Discussion

The propagation algorithm used is slow when large cliques of variables are formulated. We are using an exact algorithm for propagation. All exact algorithms based on clique size have an exponential runtime.[17] The implementation could become faster by modifying the propagation algorithm to use an approximate solution.[18]

The system executes faster when the number of parents of each node is small. It is better to break down the parent configuration by adding intermediate nodes. This accelerates the propagation of evidence and reduces the number of conditional probabilities to be specified. The propagation becomes even faster when we avoid arcs from nodes that point to their children but also to their grandchildren or even worse, to nodes further down in the network that have originated from their children. Further investigation is needed to determine how the system performs when large number of protocols are implemented.

When specifying Bayesian networks, it is best to relax certain functions so that conflict of evidence does not occur when evidence is propagated in the network. We avoided specifying conditional probabilities using 1 and 0 (we used 0.99, 0.001).

Acquiring the knowledge from the expert is not an easy task. Even after the cause and effect relationships between criteria were established, the network always seemed to require some improvement every time a new case was tested.

We also found that there are no established guidelines to define and word clinical research protocols. Even protocols from the same agency word the same criteria differently. The criteria described are sometimes ambiguous and imprecise. There is a need to develop a formal guideline and lexicon for writing clinical research protocols.

This expert system approach appears to be a useful tool to assist physicians in evaluating patients for entry into clinical protocols. It will always give a correct eligibility score when all evidence is available, but it will also try to predict eligibility at an early stage, and thereby direct the investigator to the protocols that seem most promising. In addition, the system can learn its prior and conditional probabilities by simply looking into a data-base of cases. Finally, the system has the potential to make the screening for eligibility process more efficient and less time consuming.

References

1. Nealon E. *What Are Clinical Trials All About: A Booklet for Patients with Cancer.* National Cancer Institute, 1989.

2. Giarratano J. and Riley G. *Expert Systems: Princibles and Programming.* PWS Publishing Company, 1994.

3. S. W. Tu, C. A. Kemper, N. M. Lane, R. W. Carlson, and M. A. Musen. A methodology for determining patient's eligibility for clinical trials. *Methods of Information in Medicine*, 1993:32;317-25.

4. Jensen F.V. *An introduction to Bayesian Networks.* New York, 1996.

5. Heckerman D. A tutorial on learning bayesian networks. *Technical report MSR-TR-95-06 Microsoft*, Redmond WA'.

6. Ohno-Machado L., Parra E., Tu S.W., and Musen M.A. Aids2: A decision-support tool for decreasing physicians 'uncertainty regarding patient eligibility for hiv treatment protocols. *Proc Annu Symp Comput Appl Med Care*, (1993) p429-33.

7. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman, San Mateo, California, 1988.

8. Howard R. Matheson J. Influence diagrams in howard r. and matheson j. editors. *Readings on the Principles and Aplications of Decision Analysis.*, Volume II, pages 721-762. Strategic Decisions Group, Menlo Park, CA.

9. Olmsted S. representing and solving decision problems. *Ph.D thesis, Department of Engineering-Economic Systems*, Stanford University.

10. Shachter R. Probabilistic inference and influence diagrams. *Operartions Research,*, 36:589-604.

11. Lauritzen S. and Spiegelhalter D. Local computations with probabilities on graphical structures and their application to expert systems. *Royal Statistical Society B*, 50: 157-224. 1988.

12. B. D'Ambrosio. Local expression languages for probabilistic dependence. *In Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence*, Los Angeles CA, pages 95-102.

13. Neapolitan R.E. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms.* New York: John wiley and Sons, 1990.

14. *Netica API Programers Library Reference MAnual.* Version 1.01, November 10, 1995.

15. Heckerman D. and Breese J.S. Causal independence for probability assessment and inference using bayesian networks.

16. Theocharous G. *An Expert System for Assigning Patients into Clinical Trials Based on Bayesian Networks*. 1996.

17. Shimony S.E. Finding maps for belief networks is np-hard. *Artificial Intelligence*, 68.

18. Jensen F.V. and Andersen S.K. Approximations in bayesian belief universes for knowledge-based systems. *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pages 162-169, July 1990.