Knowledge Acquisition for Clinical-Trial Selection

Savvas Nikiforou  
*University of South Florida*

Eugene Fink  
*University of South Florida*

Lawrence O. Hall  
*University of South Florida*

Dimitry B. Goldgof  
*University of South Florida*

Jeffrey P. Krischer  
*University of South Florida*

Follow this and additional works at: [http://repository.cmu.edu/compsci](http://repository.cmu.edu/compsci)
Knowledge Acquisition for Clinical-Trial Selection

Savvas Nikiforou, Eugene Fink, Lawrence O. Hall, Dmitry B. Goldgof, and Jeffrey P. Krischer

Computer Science and Engineering, University of South Florida, Tampa, Florida 33620

Abstract—When medical researchers test a new treatment procedure, they recruit patients with appropriate medical histories. An experiment with a new procedure is called a clinical trial. The selection of patients for clinical trials has traditionally been a labor-intensive task, which involves the matching of medical records with a list of eligibility criteria, and studies have shown that clinicians can miss up to 60% of the eligible patients. A recent project at the University of South Florida has been aimed at the automation of this task. We have developed an intelligent agent that selects trials for eligible patients. We report the work on the representation and entry of the related knowledge about clinical trials. We describe the structure of the agent’s knowledge base and the interface for adding new trials.

Keywords—Knowledge representation, medical expert systems, user interfaces.

I. INTRODUCTION

Cancer causes 550,000 deaths in the United States every year, and the treatment of cancer is an active research area. Medical experts explore new treatment methods, such as drugs, surgery techniques, and radiation therapies. An experiment with a new treatment procedure is called a clinical trial. When researchers conduct a trial, they recruit patients with an appropriate cancer type and medical history. The selection of patients has traditionally been a manual procedure, and studies have shown that clinicians can miss up to 60% of the eligible patients [12, 22, 30].

A recent project at the University of South Florida has been aimed at automatic selection of patients for clinical trials. We have developed an intelligent agent that prompts a clinician to enter the results of medical tests, and studies have shown that clinicians can miss up to 60% of the eligible patients [12, 22, 30].

We report the work on a web-based interface that enables a clinician to enter new trials without the help of a programmer. We have used the interface to build a knowledge base for clinical trials at the Moffitt Cancer Center, located at the University of South Florida. We review the previous work on medical expert systems (Section II), explain the knowledge representation in the developed agent (Section III), and describe the interface for adding new knowledge (Section IV).

II. PREVIOUS WORK

Researchers began to work on medical applications of artificial intelligence in the early seventies. Shortliffe and his colleagues developed the MYCIN system, which diagnosed bacterial diseases [5, 25, 26]. Experiments showed the effectiveness of MYCIN, which led to the development of other medical systems [5, 14], such as NEOMYCIN, PUFF, CENTAUR, and VM.

Musen et al. built a rule-based system, called EON, that selected AIDS patents for clinical trials [17]. Ohno-Machado et al. developed the AIDS system, which also assigned AIDS patients to clinical trials [19]. Bouaud et al. created a cancer expert system, called ONCODOC, that suggested alternative trials for each patient and allowed a physician to choose among them [3, 4]. Séroussi used ONCODOC to select participants for clinical trials at two hospitals, which helped to increase the number of selected patients by a factor of three [23, 24].

Early expert systems did not have knowledge-acquisition tools, and programmers hand-coded the related rules. To simplify knowledge entry, researchers implemented specialized tools for some systems [13, 15].

Eriksson pointed out the need for tools that would allow efficient knowledge acquisition, and described a system for building such tools [6]. Tallis et al. developed a library of scripts for modifying knowledge bases, which helped to enforce the consistency of the modified knowledge [7, 27, 28, 29]. Kim and Gil considered the use of scripts for building new knowledge-acquisition tools, and created a system for evaluating these tools [9, 10]. Blythe et al. designed a general knowledge-acquisition interface based on previous techniques [2].

Musen developed the PROTEGÉ environment for creating knowledge-acquisition tools [14, 16], which proved effective for the development of knowledge systems, including the AIDS expert systems [20], asthma treatment selection [8], and elevator-design rules [21].

III. KNOWLEDGE BASE

Physicians at the Moffitt Cancer Center have about 150 clinical trials available for cancer patients. They have identified criteria that determine a patient’s eligibility for each trial, and they use these criteria to select trials for eligible patients. Traditionally, physicians have selected trials by a manual analysis of patients’ data. The review of resulting selections has shown that they usually do not check all clinical trials and occasionally miss an appropriate trial.

To address this problem, we have built an intelligent agent that helps to select trials for each patient. It prompts a clinician to enter the results of medical tests, and uses them to identify appropriate trials.

In Figure 1(a), we give a simplified example of eligibility criteria for a clinical trial. This trial is for young and
(a) Eligibility criteria

1. The patient is female.
2. She is at most forty-five years old.
3. Her cancer stage is II or III.
4. Her cancer is not invasive.
5. At most three lymph nodes have tumor cells.
6. Either
   - the patient has no cardiac arrhythmias, or
   - all tumors are smaller than 2.5 centimeters.

(b) Tests and questions

General information
What is the patient’s sex?
What is the patient’s age?

Mammogram, Cost is $150
What is the cancer stage?
Does the patient have invasive cancer?

Biopsy, Cost is $300
What is the cancer stage?
How many lymph nodes have tumor cells?
What is the greatest tumor diameter?

Electrocardiogram, Cost is $200
Does the patient have cardiac arrhythmias?

(c) Eligibility expression.

\[
\text{sex} = \text{FEMALE} \quad \text{and} \\
\text{age} \leq 45 \quad \text{and} \\
\text{cancer-stage} \in \{\text{II}, \text{III}\} \quad \text{and} \\
\text{invasive-cancer} = \text{NO} \quad \text{and} \\
\text{lymph-nodes} \leq 3 \quad \text{and} \\
(\text{arrhythmias} = \text{NO} \quad \text{or} \\
\text{tumor-diameter} \leq 2.5)
\]

Fig. 1. Example of eligibility criteria, tests, and questions.

We encode the eligibility for a clinical trial by a logical expression, which may include variables that represent the available medical data, as well as equalities, inequalities, “set-element” relations, conjunctions, and disjunctions. For example, we encode the criteria in Figure 1(a) by the expression in Figure 1(c).

The agent collects data until it can determine whether the eligibility expression is true or false. For instance, if a patient’s sex is male, then the expression in Figure 1(c) is false, and the agent immediately rejects this trial. If the sex is female, the agent has to ask more questions. If the knowledge base includes many clinical trials, the agent checks a patient’s eligibility for each of them. It first asks for the tests related to multiple trials, and then requests additional tests for specific trials.

IV. Entering Eligibility Criteria

We have designed a web-based interface for adding new clinical trials [18], which consists of two main parts; the first part is for adding information about medical tests (Figure 2), and the second is for eligibility criteria (Figure 3). The interface includes ten screens; two of them are “start screens,” which can be reached from any other screen. We give an example of entering eligibility criteria, describe the two parts of the interface, and present experiments on its effectiveness.

Example: Suppose that a user needs to enter the criteria shown in Figure 1. First, she utilizes the “Adding tests” screen to enter the three tests (Figure 4). Then, she adds the related questions; to enter questions for a specific test, she selects the test and clicks “Modify” (Figure 4), and the agent displays the “Modifying a test” screen (Figure 5). To add a question, she clicks the appropriate button at the bottom (Figure 5) and then types the question (Figure 6).

After adding the questions for all tests, the user goes to the “Adding clinical trials” screen and initializes a new trial (Figure 7). She gets the “Selecting tests” screen and chooses the tests related to the current trial (Figure 8). Then, she marks relevant questions and the answers that make a patient eligible (Figure 9). If the eligibility criteria include disjunctions, she has to use the screen for composing logical expressions (Figure 10).

Tests and questions: The interface for adding tests and questions includes six screens (Figure 2). The start screen is for viewing the available tests and defining new ones, whereas the other screens are for modifying tests and adding questions.

We show the start screen in Figure 4; its left-hand side allows viewing questions and going to a modification screen. If the user selects a test and clicks “View,” the agent shows the questions related to this test. If the user clicks “Modify,” it displays the “Modifying a test” screen (Figure 5). The right-hand side of the start screen allows adding a new test by specifying its name and cost.

The “Modifying a test” screen shows the information about a specific test, which includes the test name, cost, and related questions. The user can change the test name and cost; the four bottom buttons allow moving to the screens for adding and deleting questions.
Fig. 2. Entering tests and questions. We show the screens by rectangles and the transitions between them by arrows. The bold rectangle is the start screen.

Fig. 3. Entering eligibility criteria.

Fig. 4. Adding a new test.

Fig. 5. Modifying a test; the bottom buttons are for moving to question-entry screens.

Fig. 6. Adding new questions; the user enters a question and answer options.

Fig. 7. Adding a new clinical trial.
We show the screens for adding yes/no and multiple-choice questions in Figure 6; the screen for numeric questions is similar. The user can enter a new question for the current test, along with a set of allowed answers. If the question is also related to other tests, the user has to mark them in the lower box. The “Deleting questions” screen is for removing old questions.

Eligibility conditions: The mechanism for entering eligibility criteria consists of four screens (Figure 3). The start screen allows the user to initialize a new clinical trial and view the criteria for old trials. If the user needs to modify a clinical trial, the agent first displays the test-selection screen (Figure 8). The user then chooses related tests and question types, and clicks “Continue” to get the question list.

The next screen (Figure 9) allows the user to select specific questions and mark the answers that make a patient eligible. For a multiple-choice question, the user may specify several eligibility options; for example, a patient may be eligible if her cancer stage is II or III. For a numeric question, the user has to specify a range of values; for instance, a patient may be eligible if her age is between 0 and 45 years. If the user clicks “Simple questions,” the agent generates a conjunction of the
selected criteria. If the eligibility conditions involve a more complex expression, the user has to click “Combined question” and then use the screen for composing logical expressions (Figure 10).

**Entry time:** We have run experiments with sixteen novice users, who had no prior experience with the interface. First, every user has entered four sets of medical tests; each set has included three tests and ten questions. Then, each user has added eligibility expressions for ten clinical trials used at the Moffitt Cancer Center; the number of questions in an eligibility expression has varied from ten to thirty-five.

We have measured the entry time for each test set and each eligibility expression. In Figure 11, we show the mean time for every test set and the time per question for the same sets. All users have entered the test sets in the same order, from 1 to 4; since they had no prior experience, their performance has improved during the experiment. In Figure 12, we give similar graphs for the entry of eligibility expressions.

The experiments have shown that novices can efficiently use the interface; they quickly learn its full functionality, and their learning curve flattens after about an hour. The average time per question is 31 seconds for the entry of medical tests and 37 seconds for eligibility criteria, which means that a user can enter all 150 cancer trials used at Moffitt in about two weeks.

**V. Concluding Remarks**

We have developed knowledge-acquisition tools for an agent that automatically assigns cancer patients to clinical trials. We have described the representation of eligibility criteria and a web-based interface for adding new trials. The experiments have shown that a user can enter a new trial in fifteen to thirty minutes. Novices can use the interface without prior instructions, and they reach their full speed after about an hour. Although cancer research at Moffitt has provided the motivation for this work, the agent is not limited to cancer, and we can use it for trials related to other diseases.

**Acknowledgments:** This work has been partially supported by the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command under contract DAMD17-00-1-0244, and by the Moffitt Cancer Center.

**REFERENCES**


