Design Considerations for Intelligent Data Entry: Development of MedIO

Ralph J. DeFriece, MD, MEE
Division of Medical Informatics, Duke University Medical Center, Durham NC

Capturing clinical data is a multi-faceted problem. This paper discusses clinical data entry problems encountered during the development of an intelligent clinical data entry system. Based on a review of the problems, recommendations are made for an approach to the design of clinical data entry programs. These recommendations include a discussion of key components in the design process as illustrated by the development of MedIO, a C++ computer program for the entry of history and physical exam information.

INTRODUCTION

Several hundred million physician office visits occur in the U.S. each year. Some studies indicate that clinicians spend roughly 38% of their time documenting clinical data [1]. Although there are many proposed benefits of using a computer system for clinical data entry, there are currently few widely accepted, uniform, yet flexible computer interfaces in use in the U.S.. Until the clinical computer interface used for transfer of clinical data approximates the speed and flexibility of the data transfer available in simple dictation and transcription systems, direct electronic capture of clinical data will be difficult.

Interface Design Problems

Many issues must be considered when designing an interface with an efficient mechanism for transferring clinical data into the computer. In examining how clinicians process data and how computers can be used in this process, it is instructive to think of the flow of information between the patient, the clinician and the computer. The clinician collects information from a patient and mentally processes or filters it based on his/her experience and knowledge base. The information deemed worthy of documentation must be transmitted through the interface into the computer. Computer interfaces represent another filter of the information as it passes from the clinician's mind into the computer. The quality of this filter affects how closely the clinician's image of a patient matches the image of the patient as stored in the computer. The general goal of a good interface for capturing clinical data should be to get an accurate clinical image of a patient into the computer in as efficient a manner as possible.

A problem that greatly complicates the design of computer programs to support the transfer of data is the variability in the entire process of clinical data collection and documentation. There are many sources of variability in the clinical data collection process. This variability is complicated by the very nature of mental processing of clinical data by different clinicians. The term "problem space" has been used in artificial intelligence to describe the way a person internally represents the facts and rules they use to deal with a problem. The structure of a clinician's problem space for a particular clinical problem may be a very complicated structure since it involves a loosely organized framework involving a large number of facts and rules.

To illustrate the design issues for an intelligent computer interface, we selected the problem space of upper respiratory infections. Several reasons led to the selection of this clinical problem area. First, upper respiratory infections (URI's) are a fairly common problem. As such, they demonstrate some of the features of common clinical data collection and documentation problems. The data collection and documentation patterns may usually be described as repetitive, with a relatively uncomplicated, shallow space of complexity.

As will be seen later in the design of a computer program to support this process, it will be important to be able to characterize the amount of variability of the data and where the variability occurs in the data collection process. By identifying where the most variability occurs in the process, one can also identify where the computer program to support this process must be the most flexible, and therefore most able to handle this variability.

A review of the prior work pertinent to the variability of data collection and documentation of URI's will help to determine and characterize this variability. Variability in the clinical documentation process can be examined across three axes: clinician, case, and module. Several studies have tended to support the idea that clinical observations by clinicians in respiratory disease may be very unreliable, or highly variable from one clinician to the next [2] [3], [4]. A second facet to examine in the clinical data

0195–4210/95/$5.00 © 1995 AMIA, Inc.

collection process involves variability across cases. In one study, it was found that the largest variability in the data collected and documented was due to the variability in presentation of the various patients for a particular problem [5]. The final facet to study involves variability across four modules, or phases of the clinical management process: History, Physical Examination, Evaluation, and Therapy [6]. This study concentrated on the reproducibility of clinical data and decisions in the management of URI's. The most notable results of their findings were that each group was most consistent in documenting historical parameters. This finding makes sense since documentation of the history involves more of a recording of medical facts found during the conversation pertaining to the patient's history. The other modules all involve a variety of more complex assessments and judgments based on the historical information found. There was a corresponding decrease in agreement between providers in these modules.

METHODS

A preliminary study was done to subjectively examine the nature and amount of variability of the data collected by various clinicians. This study was intended to be a more comprehensive look at the entire process of data collection and documentation, with an eye to narrowing the focus and refining the design of a computer program to support this process. The study was also helpful for studying specific aspects of data collection amenable to generic methods used in the design of a proposed data entry computer program. The preliminary study was also intended to give a more subjective view of the ways clinicians collect and document data.

Three simulated cases of upper respiratory infection were presented to three family practice physicians in a hospital in rural North Carolina. The cases consisted of paper simulations, with some pictures or visual aids used for parts of the physical exam. A predetermined patient scenario was presented by the interviewer, with possible positives and negatives in the history prerecorded on paper. To be as realistic as possible, it was stressed that clinicians were free to elicit and record whatever information that was felt to be necessary, in whatever order required. After the interview was performed and the physical exam data presented, further work-up proceeded as usual, with options for ordering labs, imaging tests, and medications. Laboratory, imaging, and medication or therapy planning were also discussed during the review of the case scenario. The clinicians were given flexibility in how much and what type of data they would enter into the medical record. This removed the restriction of constraining a clinician to a particular mode of data entry. The interview was taped on audiocassette. During natural breaks in the data collection process, or whenever possible, the clinician was asked to think aloud about what data items he/she would record in the medical record and why they were being recorded. Data for further analysis was obtained by reviewing the cassette tape of the patient interview and exam.

The entire documentation process was split into four areas for transcription and study: (1) The Interview, (2) Dictation, (3) Action Questions, and (4) Data Element Questions. During the interview section, the case simulation was presented and data elicited during the patient case simulation were documented. Following this, the clinician was asked to dictate the case that was presented. This dictation was transcribed for study.

To provide data relevant to some proposed mechanisms for a clinical data entry computer program, the "action questions" and "data element questions" sections were formulated. The "action questions" section focused on why a clinician chose to do a specific action. The "data element questions" focused on trying to elicit some of the common, or critical data elements a clinician might use for the given clinical problems.

The fundamental units of analysis for the simulation were determined from the data collected during the taped interviews. There were two categories of units of analysis: data elements, and actions. Data elements included all information pertaining to physical findings and historical facts available during the interview. Actions included all tests ordered, therapies, instructions, and plans for the patient's care. Trends across the three axes of: case, clinician, and module were examined. The modules consisted of: History, Past Medical History, Physical Exam, Lab Studies, and Plan of Therapies. Issues across the entire documentation process were examined to narrow the focus of the proposed computer program.

RESULTS AND DISCUSSION

The previously cited variability studies tended to support the idea that the data collected during the history and physical modules were somewhat more uniform than the plans of action that clinicians take. That is, it was somewhat more difficult to predict what a particular clinician's plan of action would be for a particular clinical problem. The preliminary study also seemed to suggest that clinicians were most consistent versus themselves in the number and type of data elements collected across cases, in the history and physical modules.

The subjective findings of the study included gaining more insight into the relative difficulty of trying to design a computer program based on specifications elicited from a set of doctors. The clinicians were comfortable with providing generalities with regard to how and why they collected data. They were not as able to describe specific instructions or specifications with regard to the sets of data elements they would like to collect and with the actions they might want to take at various points in the data collection process. So, a computer system designer may be unable to get clear specifications from the clinician user population. In addition, there is much variability regarding how clinicians go through the data collection process and what different clinicians need at various points in the data collection process. It would therefore seem to be a difficult task to design a computer system that does exactly what a clinician wants since it may be difficult for a clinician to clearly specify what he or she wants in a variety of clinical situations.

What does all of this mean to someone designing a system to support the data collection process? One of the main conclusions is that it might not be possible to efficiently support a wide range of providers using a fixed set of computer tools for all phases of data collection. Flowsheets, or fixed data entry screens have a limited number of options on them and may only capture some percentage of what a clinician might want at various points in the data collection and documentation process.

Another way to look at the studies of variability involves using a broad interpretation of the variability data. That is, modules that have the most variability will also tend to be the modules that are the most difficult to design. These modules will need to be programmed to be more flexible, or plastic. Modules with less variability between providers might be programmed to have the most fixed components, or most fixed data elements and actions.

What these studies of variability might also suggest is that it might be easier to try to focus on a particular axis of variability in the design of a computer system. The three axes described above involve patients, providers, and modules of data collection. Both clinicians and patients introduce a large amount of variability into the process. Some mechanism for tracking this variability, or remembering the variability that occurs from patient to patient and from clinician to clinician, would also be very helpful in designing an efficient medical interface.

PROPOSED DESIGN MECHANISMS

Current approaches to designing clinical data entry programs involve trying to encode a large, complicated set of rules and data elements for each of thousands of different clinical problems and situations. Construction of computer programs in this manner is analogous to trying to extract and encode the knowledge embodied in the problem spaces of different clinicians to develop a generic clinical problem space for each of thousands of clinical problems.

An alternate approach to clinical data entry is illustrated in the design of MedIO, a data entry program. Rather than try to encode a complicated list of rules that may only apply to a limited number of clinicians, a more practical approach is to support a clinician's data collection habits by automatically organizing the data elements and rules he or she uses in a way that tries to mimic how the clinician's problem space is constructed.

Two general methods for doing this involve pattern accumulation and pattern recognition. Pattern accumulation will be important in addressing the variability that occurs in the data collection process. Pattern recognition will be important in addressing the large number of possible clinical situations that may occur.

One of the major problems is identifying and targeting where the variability in the clinical data collection process occurs. Some reduction in variability might be accomplished by accumulating patterns of data elements that are collected for different clinical problems and situations. Some conclusions from the discussion of variability are helpful in deciding how to organize the accumulation of patterns. Since the most variability in data for a given clinical problem occurs from patient to patient, the proposed mechanism for data collection will sum up this variability by accumulating patterns across a collection of patients. This pattern accumulation will allow flexible, automatic grouping of commonly used data elements on a problem specific and clinician specific basis.

Pattern recognition will be important in determining what a clinician does on the basis of a given set of data elements. Since it is difficult, if not impossible to generate a set of rules that a group of clinicians might apply in a variety of situations, it seems natural to select a method of pattern recognition to try to identify patterns of action of different clinicians. By employing a method of pattern recognition, it should be possible to automatically track how a particular clinician might act given a particular set of data elements. This should produce an interface that anticipates how a particular clinician might act in a given clinical situation, based on his/her patterns of action in the past.

These methods should eliminate or reduce many problems. Problems of consensus about which data elements to include in an interface should be eliminated since the clinicians automatically generate their own sets of commonly used data elements. Since data collection would be based on the prior data collection habits of a clinician, or group of clinicians, the effects of regional variation, differing vocabularies, and differences in training of clinicians should also be lessened.

DESCRIPTION OF DATA ENTRY PROGRAM

MedIO is a C++ computer program under development that uses pattern recognition and pattern accumulation for clinical data entry of history and physical examination information. The program maintains a global pool of data elements specific to a particular class of problems (i.e., Upper Respiratory Infection for this prototype) . These data elements are maintained in four lists: History Phrases, History Symptoms, Physical Exam Anatomical Site, and Physical Exam Descriptors.

A simple method of pattern accumulation is used to generate and store lists of data elements in a compressed, problem specific format. Pattern accumulation addresses the variability that occurs in the data collection process. Over several cases of a particular clinical problem, the computer program will automatically accumulate and filter a set of data elements most frequently used for that problem. The problem-specific accumulated pattern of data elements will be called the composite image because the view it gives of a clinical problem is a composite view over a collection of several patients. Upon selection of a clinical problem from a menu, the composite image for that problem will automatically be loaded into the four list boxes.

The following steps are taken during the process of generating a composite image. The number of occurrences of a particular data element are accumulated. The set of frequently used data elements is organized into a collection of vectors of variable length. The length of the vector will be limited by a filtering process in which less frequently used data elements are excluded from the composite image. The goal of the composite image is to generate a collection of the clinician's most frequently used data elements for a given clinical problem. The collection of less frequently used elements will be relegated to a second easily available list.

Another problem to address in the data entry process involves trying to track how or why clinicians perform actions. In designing computer programs to allow for flexible data entry, clinicians should not be forced to navigate through several different computer screens or forms to search for an action. Yet it is difficult to design a flexible system that can handle a large number of clinical situations given the millions of medical facts and large number of possible combinations of data elements and possible actions that may occur during a variety of clinical encounters. Since different clinicians may tend to collect different data elements and then may act very differently depending on the particular set of data elements, this method will try to identify how a particular clinician acts in different situations. Then, rather than "hard coding" in a particular set of rules to go with a particular set of data elements, the program will be designed to make the given actions available on the computer screen when the clinician completes the collection of data elements.

Unlike data element collection and documentation, action patterns may not be organized on a problem specific basis. Given the large number of clinical situations a clinician might encounter, it is difficult to try to anticipate what actions a clinician might take in a variety of situations. Pattern recognition addresses this by identifying the response patterns of clinicians to different clinical situations and maintaining these responses, or actions, in a fifth list box. These response patterns will be organized into sets called "action-data element pairs". An "action-data element pair" is a collection of critical data elements usually associated with a typical clinical action. After the repetitive presentation of sets of clinical data elements to the pattern recognition portion of the program, the goal of the program will be to identify likely plans of actions based on these data elements.

An associative network is a type of neural network particularly suited to this task [7]. Their function is to associate one set of vectors with another set of vectors. Autoassociative networks can be used to regenerate noise-free, complete patterns from one that is noisy or incomplete. In this program, there will be three input vectors: the history data element vector, the physical exam data element vector, and the action vector. Data elements will be limited to

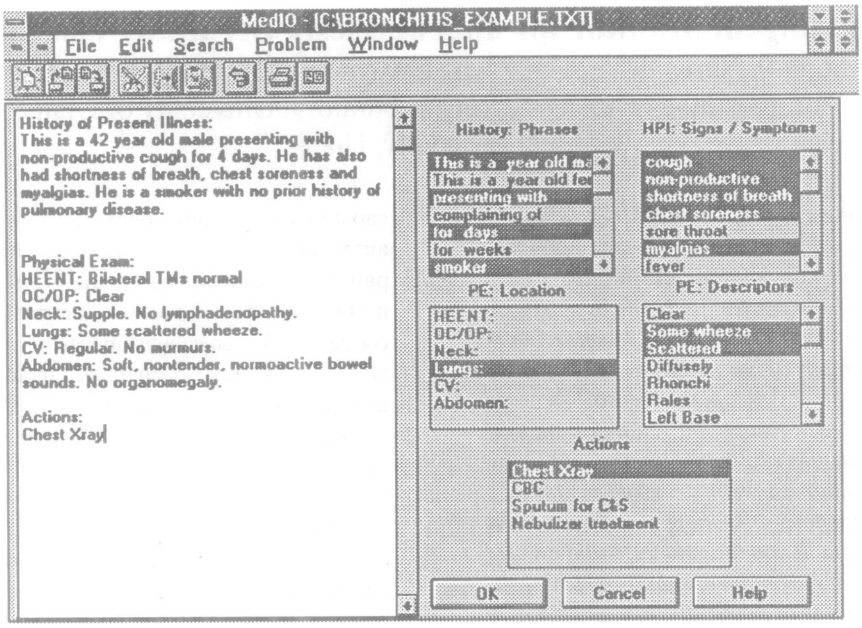

Figure 1: Main Screen for MedIO

those that are in the composite image for a particular problem. During the training phase, some number of clinical cases will be input to establish the weights of the nodes of the network. In the prediction phase, a particular case will be entered, with an empty action vector. The network will then generate an output pattern, completing the sets of actions to be taken based on its prior set of learned associations of data elements and action patterns.

The data entry screen uses a simple, Windows multiple document interface [Fig. 1]. Upon selection of a particular clinical problem from a menu, a template document is pulled up, with an editing window and the 5 data entry list boxes mentioned previously. Data may be entered by typing in the edit window and/or by selecting data elements from the various lists of frequently used data elements and actions.

The goal of the program is to remember the most frequently used data elements and actions for a particular problem and make them quickly accessible. This design approach attempts to provide a uniform, yet flexible data collection program while also reducing the complexity and repetition inherent in the design of clinical data collection programs.

The efficiency and content of clinical data entry using MedIO running on an IBM PC will be evaluated in a clinical setting. Results from this evaluation should help define the data entry time trade-offs between customizing an interface to the individual and using a more uniform interface for a group of clinicians.

References

1. Mamlin JJ, Baker DH. Combined time-motion and work sampling study in general medicine clinic. Medical Care. 1973; 11(5): 449-456.
2. Schilling RSF, Hughes JPW, Dingwall-Fordyce I. Disagreement between observers in an epidemiological study of respiratory disease. British Medical Journal . 1955; 65-68.
3. Smyllie HC, Blendis LM, Armitage P. Observer disagreement in physical signs of the respiratory system. The Lancet . 1965; 412-413.
4. Garland LH. Studies on the accuracy of diagnostic procedures. American Journal of Roentgenology. 1959; 82:25-38.
5. Revicki DA. 1984. The dependability of medical encounter diagnostic information. Medical Care. 1984; 22: 661-669.
6. Wood RW, Diehr P, Wolcott BW, Slay L, Tompkins RK. Reproducibility of clinical data and decisions in the management of upper respiratory illnesses. A comparison of physician and non-physician providers. Medical Care. 1979; 17(7): 767-779.
7. Hecht-Nielsen R. 1990. Neurocomputing. Addison-Wesley Publishing Company, New York. 1991. pp. 354-357.
8. Anderson JA. Cognitive capabilities of a parallel system. NATO ASI Series F20. 1986: 209-226.

This work supported in part by National Library of Medicine Training Grant #LM07071-3