ELSEVIER

Artificial Intelligence in Medicine 23 (2001) 71–87

Artificial Intelligence in Medicine

www.elsevier.com/locate/artmed

Medical informatics: reasoning methods

William J. Long*

MIT Laboratory for Computer Science, 545 Technology Sq., Rm 420A, Cambridge, MA 02139, USA

Received 5 January 2001; received in revised form 30 March 2001; accepted 24 April 2001

Abstract

The progress of medical informatics has been characterized by the development of a wide range of reasoning methods. These reasoning methods are based on organizing principles that make use of the various relations existing in medical domains: associations, probabilities, causality, functional relationships, temporal relations, locality, similarity, and clinical practice. Some, such as those based on associations and probabilities have been developed to the point where there are off-the-shelf tools available for the researcher to develop new decision support tools. Others such as temporal relations require more effort to use effectively. Even so, we have learned the importance of a separate explicit representation of the domain knowledge and have considerable experience and an impressive armamentarium with which to face the new milieu provided by the Internet.

Keywords: Reasoning methods; Diagnostic reasoning; Knowledge based reasoning; Informatics

1. Introduction

The ubiquitous availability of the Internet has given medical informatics a new direction. The ability to put computer-based systems in the hands of providers, patients, institutions, and researchers and connect them together has completely changed the role of such systems in medicine. Decision support systems can now be thought of as tools fitting into appropriate niches in the practice of medicine and available when it is appropriate to use them. They will be connected in ways that will allow the data to flow where it is needed — the data that has hindered progress in the past. This new milieu brings to the forefront many issues such as interfacing the systems, providing security, and integrating the systems into the practice of medicine. However, before we are too far past this juncture and its necessary change of focus, it is useful to consider what capital the previous era contributes to this new environment.

* Tel.: +1-617-253-3508; fax: +1-617-258-8682.
E-mail address: wjl@mit.edu (W.J. Long).

0933-3657/01/$ – see front matter
PII: S0933-3657(01)00076-8

There are many ways that the research over the past thirty or more years might be characterized, but the one that seems to underpin the rest, indeed to define the influence of medical informatics on domains beyond medicine, is the development of methods for reasoning about the various problems confronted in medicine. Many of the reasoning methods now accepted as the basic tools for computer-based problem solving, including production rules, case based reasoning, neural networks, and Bayesian probability networks, were first developed in a medical context, inspired by medical problems, or greatly influenced in their development by the challenges of medicine. Without attempting to assign credit for the various methods, this paper will take a look at the array of tools the community has amassed to handle the reasoning problems in medicine, the capabilities they provide and their limitations. There are still gaps in this tool set; indeed the most glaring hole is the lack of effective ways of integrating multiple methods to multiply the advantages of each.

1.1. Generic tasks

Medical reasoning can be broken down into three generic tasks based on the time period involved. Diagnostic reasoning is the process of reconstructing the past from available evidence. Planning is reasoning about the effects of actions in the future. Patient management is organizing and dealing with the present. Similarly, medical computing tasks can be generally categorized into one or more of these generic tasks. While the practice of medicine requires all three kinds of reasoning to go on simultaneously, medical computing has not yet advanced to the stage where programs are very successful at addressing more than one of these aspects of reasoning at a time. The primary problem at which much of the early effort was addressed is diagnosis. Much research has also addressed therapy planning and management, but we can look at the particular problems of medical diagnosis as the impetus for much of the development of reasoning methods.

1.2. Requirements for reasoning

The basic problem of diagnosis is to take a set of effects and determine the cause. Actually, the problem is better characterized as looking for an ordered list of possible causes, since some degree of uncertainty is a nearly constant feature of medicine. Because the human body is extremely complex, only partially understood, and not a manufactured artifact that we can easily instrument or modify for ease of maintenance, we need all the leverage possible to solve the clinical problems. Indeed, since there is often no way to be certain of a diagnosis short of autopsy, the result as well as the process of diagnosis involves uncertainty. Thus, a diagnostic program must have a number of characteristics to make it useful, i.e. for users to be willing to use it. Firstly, the program must perform well. Addressing this issue in depth would take us into the issue of evaluation and standards of performance, which is not the primary focus of this paper. Suffice it to say that we have learned that evaluation of decision support systems is difficult and the standards for doing so are only slowly coming together.

Beyond showing that the program performs well on a population of patients, it is necessary to show the user that the program is performing well on the particular case of

concern to the user. To do this, the program must be convincing, either because the results are justified appropriately, explanations are provided, or the reasoning is readily understandable to the user. Likewise, the knowledge of the program needs to be understandable to both the user and the developers, whether they are medical experts or informaticists. These requirements drove the developers of medical reasoning systems away from the traditional programming approach with the algorithms incorporating and implementing the domain knowledge to an approach with a separate knowledge base and inference engine. The separate knowledge base makes several things possible. The tasks of developing the inference engine and knowledge base can be separated and given to experts with different skills. The knowledge base can be examined separately by anyone who understands the paradigm or organizing principles under which it was developed. For the domain expert this is much simpler than attempting to understand a computer program. For example, the domain expert might only need to understand the findings associated with different diseases. The knowledge base can also be used to explain or justify results as well as produce them.

The separation of the knowledge base and inference engine imply that there needs to be a clear semantics of the knowledge base that the inference engine can reliably use to do the reasoning. That is, the inference engine must do reasoning consistent with the intended meaning of the represented knowledge. The knowledge base must represent well-understood relationships among the clinical concepts that can be consistently manipulated by the domain experts separate from the inference engine and can be faithfully reasoned with by the inference engine. There are a number of relationships available in the medical domain that make this possible. These include associations, probabilities, causality, functional relationships, temporal constraints, locality, similarity, and clinical practice. These relationships provide organizing principles around which a mechanism can combine knowledge to do the necessary reasoning to solve the clinical problem. The organizing principle in turn enables a view of the problem that recommends ways of solving it.

There is also a tension in designing an organizing principle for a system. To make the knowledge base easy to understand and the inference engine reliable and relatively transparent, it would be best to pick a single relationship and use it to develop a simple uniform mechanism. However, medicine is more complex than that and such approaches will be limited. For example, a diagnosis program based on associations may present diagnoses that are clearly wrong when temporal relations are considered. Still, the first generation of reasoning methods relied on uniform mechanisms based on single relationships and we will consider the strengths and weaknesses of some of the more influential methods. The development of methods presents somewhat of a logical progression, although it is more of a tree than a line. The following paragraphs will explore the progression.

2. Organizing principles

2.1. Associations

The association between diseases and findings was one of the first organizing principles used to tackle the problem of medical diagnosis. The key observation is that the same

disease in different patients tends to have similar findings. There are often differences depending on severity, complications, chronicity, and other factors, but there is usually enough overlap in the findings to be recognized as being caused by the same disease. That is, for each disease there is a list of findings that might be present when the disease is present. An instance of the disease may not manifest all of the findings, but there are usually enough to distinguish the disease from other diseases. This view of the medical knowledge suggests a simple approach to diagnosis — simply invert the association to get all of the diseases associated with each observed finding and search this list to find the disease or diseases that account for the largest number of findings. The use of simple associations goes back to schemes to use punch cards with diseases on one axis and findings on the other to ''shake out'' the diseases associated with the findings. This simple notion is still the basis for Weed's Knowledge Couplers [48], which attempt to provide a comprehensive list of findings for each disease in order to assist the physician in considering as complete as possible a list of diseases that could account for the findings.

Associations by themselves convey only a small part of the knowledge we have about how diseases relate to findings. The next step was to improve upon associations by observing that there are strengths of association. When a disease is present, some findings are always present, some rarely, and others with intermediate frequency. On the other hand, some findings, even some that rarely occur are only caused by one or two diseases, and thus are very useful for diagnosis. These ''enhanced associations'' are the organizing principles for such programs as QMR [15,36] and Dxplain [5], which provide diagnostic coverage for large parts of medicine. QMR, for example, uses three properties besides the associations themselves for diagnostic reasoning: the frequency of the finding with the disease, the frequency of the disease with the finding, and the importance of the finding. This last measure helps the diagnostic algorithm ensure that its solutions cover the important findings and if any findings are left unaccounted, they can be reasonably ignored. It would seem that the evoking strengths should be derivable from the frequencies if the prior probabilities of the diseases are known. However, the implicit independence assumption for diseases and findings inherent in this view of the problem is rarely true and providing explicit evoking strengths is one way to compensate for this. The frequencies do implicitly constrain the evoking strengths, so care must be taken to ensure consistency among the numbers.

The advantages of associations stem from their simplicity. Very little information is needed and it is easily understood. Thus, it is easy for someone with medical knowledge but little computational expertise to add knowledge to such a system. The semantics of the representation is also very simple, allowing sophisticated kinds of analysis and multiple uses of the data. For example, it is easy to use the knowledge base of associations to determine all of the common or uncommon findings caused by a disease, the diseases that might cause a finding, or the findings that might distinguish between a couple of diseases. Considerable theoretical work has been done using this model of diagnosis [44] and powerful algorithms have emerged.

2.2. *Criteria tables*

One of the disadvantages of these simple associations is the difficulty of distinguishing between diseases with similar findings. This is partly because there are intermediate states

between the diseases and findings that increase the co-occurrence of findings of the same intermediate states. That is, treating all of the findings as providing independent evidence leads to unreliable differentiation of diseases with similar findings. A heuristic method for making decisions about such diseases led to the use of criteria tables. Criteria tables associate diseases with clusters of findings. That is, they utilize the lack of independence between findings for a disease. Criteria tables were first developed for glaucoma diagnosis in CASNET [29] and then used in the context of rheumatology, where the concept of developing lists of major and minor criteria for recognizing a disease was already part of medical training. Because criteria tables are part of the medical context, systems such as AI-Rheum [31,24] are easy for the physician to understand and it is easy for the medical expert to contribute to the knowledge base. A weakness of criteria tables is that, since the important differentials are encapsulated in the criteria for diagnosis of the different diseases the implications of new knowledge are not easy to see. For example, if such a program were used in a context with different priors for the diseases, the criteria tables would need to change to reflect the appropriate sensitivity and specificity to the diseases.

2.3. Bayesian networks

The reliance on heuristic methods to deal with intermediate physiologic states and syndromes as well as the heuristic algorithms for combining evidence led a number of researchers to look for a sounder mathematical footing for the diagnostic programs. Fortunately, Kim and Pearl published a model for sound and efficient probabilistic reasoning in the form of Bayesian networks [23,40] and Lauritzen and Spiegelhalter [30] published an effective method for handling networks with rejoining branches. With this machinery, it was possible to develop probabilistic networks representing diseases, their mechanisms, intermediate and ultimate effects. The basic assumption of a Bayesian network is that the probability of a node is completely determined by the nodes linked as immediate inputs to it. If one of those nodes is true (or has a certain value), the state of its ancestors and the state of its other effects is irrelevant. This assumption provides the separation between nodes that enables efficient computation of the probabilities in the network, and hence, provides the power of this method. The basic weakness of Bayesian networks is the assumption that everything important about the relationship between two linked nodes (e.g. a disease and physiologic state) is captured by a probability. No distinction is made among the various ways that a cause can produce an effect. The strength, of course, is the solid foundation in probability theory. Indeed, some researchers view Bayesian networks as the standard against which any other method should be measured.

One of the first programs to use these mechanisms was MUNIN [37], a program for diagnosis of neurologic disorders. There are too many other medical programs based on Bayesian networks to mention here. However, once in a while there are problems that are made to order for a particular methodology. Genetic counseling, that is, determining the probability of genetic disease as determined from the pedigree of the family fits perfectly the reasoning view of Bayesian networks. *Geninfer* is a program that uses the family tree as a Bayesian network to provide genetic counseling [47]. Bayesian networks have proven useful enough that several shells have appeared for building and using them. The first of

these was HUGIN [3], which came out of the work on MUNIN. More recent Bayesian network shells have found their way from beginnings in medical applications all the way to the Microsoft Answer Wizard [20].

2.4. Causality

Another relationship that has been used as an organizing principle in a number of diagnostic systems is causality. One of the first to explicitly use causality was CASNET, although its notion of causality is more of historical ordering or phases of the diseases. Exactly what constitutes causal reasoning varies from system to system. Most systems purporting to do diagnostic reasoning based on causality have a fairly complex knowledge representation system that includes relations between concepts describing the aspects of causality deemed important in that domain. On top of this is an inference engine for reasoning with the knowledge structure by using a number of properties of causality to infer the necessary diagnostic conclusions. Perhaps because the particular choices have been strongly influenced by the domain of the program or because the complexities of representation and reasoning have not resulted in a compelling view, none of these systems has been successfully generalized to other domains.

Since it is easy to think of causality in probabilistic terms, a number of efforts have used Bayesian networks to represent the causal structure of medical domains [13]. However, it should be noted that the direction of links in a Bayesian network do not necessarily correspond to any notion of causality since a probability on a link only means that the second node occurs with that probability when the first one is true. Causality is a difficult concept to capture satisfactorily. While there are certainly rules that causal relations obey, the assignment of causal structure in systems characterized by feedback can become an arbitrary linearization of a process better characterized as mutual constraints. Still, causality is important because users tend to think in terms of causality and the closer a system's reasoning corresponds to that of the user, the easier it is for the user to understand and accept the conclusions reached. An indication of the difficulty of using causality as an organizing principle is that no one has developed an effective tool for reasoning with causal relations. That may be changing with considerable theoretical work being published on causality [17,41].

In the Heart Disease Program [32,33] we chose to use a causal network representation with probabilities on the links, but forgo the formal semantics of a Bayesian network in order to preserve the sense of causality. In particular, in the pathophysiologic model of the cardiovascular and related systems (about 200 nodes) there are a number of forward loops in the network, which make it impossible to assign a consistent set of probabilities to the nodes. A simple example is that renal insufficiency can cause hypertension and hypertension can cause renal insufficiency. Both definitely occur in life. In a particular case, the program chooses one or the other depending on any other relevant findings to maximize the overall probability. Thus, in a particular hypothesis (that is, a subset of the pathophysiologic model instantiated for the patient) there are no forward loops and there is a consistent interpretation of the overall probability. This representation of the domain knowledge as a probability network has proven very useful. It is easy to think about the relationships and get an expert to assign nodes, links and probabilities to the links. Care must be taken to

include all intermediate nodes that are measurable or represent common pathways to findings to preserve the power of the reasoning because there is a natural tendency to eliminate the intermediate nodes from our descriptions to one another. There are, however, limitations to the representation. Firstly, as diseases or pathological states get more severe, they have more effects and the effects are more certain to happen. The result is that the probabilities on the links are dependent on the severity of the node at their causal end, not just its presence or absence. We have dealt with this by allowing many of the nodes in the network to have multiple states representing different severities. This complicates the reasoning considerably but by judicious use of such states we are able to represent the severity differences that are clinically important. A second limitation is that the timing of diseases and findings interacts with the semantics of a probability network, but that will be discussed in a few paragraphs.

2.5. Functional relationships

Functional relationships, while not as common as associations or probabilities, are very effective for reasoning when they are available. The ABEL program for acid–base and electrolyte diagnosis [39] is an example of a program that makes use of a known functional relationship for diagnosis. In this case, the Henderson–Hasselbalch equation accurately reflects relationships so that the blood gases and electrolytes can be used to filter the causes down to one or two conditions or combinations. Since this diagnosis is at the level of acute or chronic metabolic or respiratory acidosis or alkalosis, it is still necessary to determine the cause of the condition. At that level there are no easily measurable parameters and other kinds of reasoning must take over.

One use of functional relationships that waxes and wanes is simulation of physiologic systems. Guyton et al. [18] have developed some impressive simulations of the cardiovascular system, including one with hundreds of parameters. These have been used in teaching [12] but not for problems involving individual patients. The difficulty in applying such models to patients is that many of the relationships have significant variation from person to person so individual reaction will vary from the average reaction determined by physiologic studies. Furthermore, to customize the model to an individual it would be necessary to measure many parameters, including some whose measurement is not compatible with the health of the patient. Because of the nature of simulation, the model must be completely specified in order to reason with it. Thus, quantitative simulation has not proven to be an effective tool for the clinical setting although it is a powerful teaching tool.

One program that has successfully combined a constrained but effective functional model with a Bayesian network is VentPlan, a program for managing a ventilator [45]. It uses discrete data about the patient to set what is known in the Bayesian network. Some of the nodes in the network represent parameters in a mathematical respitatory model. The probabilities on these nodes are interpreted as probability distributions of the parameters of the mathematical model. Then using these and other measured data an empirical Bayes estimator estimates the model parameters for simulation. This is a very effective and appropriate combination of methodologies, but unfortunately has not spawned any similar efforts.

2.6. Qualitative simulation

Qualitative simulation was invented to deal with some of the problems of quantitative simulation [28]. Qualitative simulation simplifies the functional relationships to ones of direction of influence. This removes problems of parameter values and the need to know exact relationships while retaining the ability to determine some aspects of the nature of responses to changes. However, with the complex feedback systems in the human body and the need to know more than just directions of response, qualitative simulation has not been very useful for clinical problems.

2.7. Temporal relationships

Temporal relationships are obviously of great importance in medicine and many programs have attempted to make use of them in one form or another. However, temporal relationships are generally not sufficient by themselves for diagnosis. Therefore, they have usually been added to some other relationship and have been implemented within a system using some other organizing principle. An interesting example of this is a program for managing diabetes by modeling the day-by-day progression of the disease and symptoms as repeating parts of a Bayesian network [38]. The nodes from the previous day determine the prior probabilities for each day. This approach has the advantage that the semantics and reasoning machinery are well understood. It has the drawback that only a limited number of days can be modeled before the number of nodes in the network makes it computationally expensive. This approach works for a domain like diabetes where the time period of concern is hours and longer or shorter periods can be ignored, at least for daily management. Actually, temporal extensions to Bayesian networks has been somewhat of a growth industry in the last few years [2], but no particular consensus or set of rules has developed about how best to integrate Bayesian networks and temporal reasoning. The challenge for any approach to temporal reasoning, particularly for diagnosis, is to avoid a computational explosion when there is little data to constrain the temporal relationships among the entities of concern. Any approach that attempts to represent all of the possibilities is likely to fail in a domain of any complexity.

Often the first step in reasoning about temporal data is to determine intervals in which the data behaves in recognized ways (constant, increasing, in a range, etc.). These intervals may be determined strictly from the data but usually any events that affect the data have an impact on the appropriate bounds for the intervals. A number of systems have taken this approach, providing a mechanism to generate the intervals from which other reasoning could take place. A mechanism that takes a more model-based or expectation driven approach to temporal relationships is trend templates [19]. These were inspired by the clinical approach to growth disorders in children and have proven useful in providing computerized diagnosis of these diseases. They have also been applied in the interpretation of the much more frequent data acquired in the intensive care unit. The basic approach is to provide templates of expected behavior with time intervals over which changes normally take place. These are compared to the actual data and deviations from the expectations become the evidence for the conditions of interest. The strength of this approach is that it nicely encapsulates the temporal information about the domain in a form that can be used

for analyzing temporal data and translating it into a succinct conclusion that can then be used as a finding for other diagnostic reasoning. Thus, it is a good way of integrating temporal reasoning into another methodology.

We took a different approach to temporal reasoning in the current version of the Heart Disease Program [35]. It was necessary to include temporal reasoning in the program because the assumption of causal separation that gives power to the probability network representation ignores important temporal information. For example, a myocardial infarction (heart attack) can cause pulmonary edema and pulmonary edema can cause pleural effusion, among other things. However, pleural effusion requires weeks of pulmonary edema and so is not accounted for by a myocardial infarction a few hours ago. The assumption that the probability of pulmonary edema causing pleural effusion is independent of the causes of the pulmonary edema is violated. To handle such reasoning, we added temporal constraints to the nodes and links and their instantiation for the particular case and propagate the constraints through the network as necessary. If there is more than one temporal extent consistent with the findings in the case (e.g. acute and chronic pulmonary edema) the instantiated node is duplicated to allow the alternatives and combinations to be explored. The multiplication of nodes is tightly controlled to prevent the network from growing too large. This approach to temporal reasoning has proven very effective for representing and reasoning with the kinds of cases encountered in the cardiovascular domain. The approach of replicating the whole probability network as was done for diabetes would not work because there are many different time frames pertinent to cardiovascular problems: arrhythmias can change in minutes, a myocardial infarction in hours, peripheral edema in days, cardiac dilatation in weeks, and ventricular hypertrophy in months or years.

2.8. Location and similarity

Still other relations are available in the medical domain and are useful for some diagnostic problems. *Location* is useful in at least a couple of contexts: trauma and tumor diagnosis. For trauma diagnosis it is necessary to know what organs might be affected by trauma in a known location as well as inferring the location of trauma from findings. These principles have been used in the development of TraumAID [16] for reasoning about various kinds of wounds. The same kinds of anatomical information have been used by Banks et al. for the diagnosis of neurologic disorders in Caduceus [4] and more recently for analyzing CT and MRI images using a neuroanatomic knowledge base [11]. To accomplish this they developed an Octree representation for the data that allows the accessing of the neurological information by location and refinement to the appropriate structures to account for the findings. Such tools provide a way of using locality to produce findings that can be used with other findings for reasoning with another methodology. In our own work, we have noticed that a more general notion of ''closeness'' is often needed for reasoning. In the cardiovascular domain, there are a number of findings that can be confused, particularly by less experienced users. Using a similarity measure, which could be dependent on factors such as background noise level for auscultation, is very useful to account for findings that could otherwise lead a diagnosis program astray.

2.9. Case-based reasoning

It has been observed that physicians often relate the present cases to those seen in the past. From this observation in medicine and other fields grew the paradigm of case-based reasoning. The essential idea is that if a suitable measure of similarity exists, the new case can be related to one or more similar past cases in an appropriately indexed database. From that case or cases the appropriate diagnosis or other question can be answered by analogy or simply by copying the answer if the match is close enough. This approach was first tried by Kolodner in psychiatry [25] and a number of other domains. In our laboratory we compared case-based reasoning to model-based reasoning for the diagnosis of cardiovascular disease, using the Heart Disease Program as a model. Having the physiologic model made it possible to identify similarities in findings by looking for common immediate causes. The initial experiments were very promising [26,27], but the cases were restricted to a very similar set of diseases and combinations. Once we extended the case set to the mix of cardiovascular cases encountered in a tertiary case hospital, the performance fell dramatically [1]. The basic problem is determining what cases and what aspects of the cases are similar. To the physicians who reviewed the 240 cases, many cases appeared essentially the same because the important aspects of the cases were similar. However, to the program each case appeared different because each had findings not shared with other cases, even using the physiologic model for matching. Often the differences were among chronic stable previously existing diseases, but without more knowledge the program was unable to discount such differences. We have since looked at ways of dividing known cases and combining parts of them to produce pseudo-cases from which the case-based program could apply its reasoning [22]. Still, the weakness of case-based reasoning for medical diagnosis remains the difficulty of producing an appropriate metric for deciding that two cases are similar enough in the aspects that matter for solving the problem at hand. One way of overcoming this problem is to narrow the domain while keeping the database as large as possible so there are always similar if not identical cases to use for reasoning. Thus, all of the case-based programs in the literature are in restricted domains.

2.10. Flowcharts

Probably the first abstraction from programs was the use of flowcharts. Indeed, flowcharts have been used as a tool of programmers for designing code, since the early days of assembly language coding. Flowcharts have been used in a clinical setting to capture various procedures, such as the appropriate care of diseases. They also have been used to capture the diagnostic process and whole books of diagnosis flowcharts have been published. Flowcharts are available on the Web for patients to use for symptom triage. However, when attempts were made to use computers to do diagnosis by following the flowcharts it became clear that diagnosis is rarely an ordered process while flowcharts require an inflexible order. Unless all of the data is immediately available (such as when asking patients for their symptoms), it is necessary to duplicate parts of the flowchart to handle data being available in the wrong order. A more difficult problem is handling missing data when data can be costly or unavailable. Even so, a few successful programs

were developed in this way, such as Bliech's acid–base diagnosis program [8], which is still in use.

2.11. Production rules

Because of the need for flexibility in determining what data was available and in what order, production rules were developed. The primary exemplar of production rules is the Mycin system for diagnosis of bacteremia [10]. The organizing principle of a production rule system is that knowledge can be stated as an unordered set of rules with a set of antecedents and a set (usually just one) of consequents. That is, if all of the antecedents are true, the consequent is true, no matter when the rule is applied. One of the innovations of the Mycin system was to include uncertainty in the rules. Each consequent has a certainty factor between 0 and 1 representing the confidence in the deduction with a scheme for combining certainty factors if multiple rules have consequents relating to a concept. A production system does diagnosis by backward chaining. That is, trying to establish rules to meet a goal — initially, ''what's wrong with the patient?'' To do that it uses each unknown antecedent as a goal until it gets to antecedents that can be answered by asking the user. The advantage of a production rule system besides the flexibility of application is the fine-grained modularity of the representation. That is, each rule is complete unto itself. In addition, production systems have extremely wide applicability, not depending on anything unique to medicine. They can also be used for inference using forward chaining. That is, any time the antecedents of a rule are true, its consequents can be asserted. This method of reasoning has been used to great effect for planning systems. There are also drawbacks to production rules, basically the inverse of the advantages. The wide applicability means that rules can encode medical knowledge or control knowledge and no distinction is made. Thus, different kinds of knowledge are mixed and the clean separation between medical knowledge and process is lost. The flexible ordering means that each rule must explicitly state all of its assumptions or it may be applied in an unanticipated context with unforeseen consequences. Still, production rule systems have a compelling simplicity and generality that has almost made them synonymous with expert systems.

3. Acquiring the knowledge

A constant issue with all expert systems is where to get the knowledge. Relying on expert judgment has drawbacks. It is extremely costly, very time consuming, and experts do not always agree. These issues have led to efforts to learn the knowledge from data. Such efforts in machine learning have taken several forms. For example, considerable work has been done in automatically learning Bayesian networks from data. Systems have been developed that learn the probabilities in a network of known topology [6] and others that learn the structure of the network as well [14,21]. Learning the structure is very difficult for several reasons: firstly, there is noise in medical data and that along with the randomness of case selection can easily lead to spurious dependencies and independences, and hence, to extra or missing connections. Secondly, in most medical problems there are nodes that make sense physiologically for which there is no direct measurement available. Finally, the

best arrangement of probabilistic dependencies as inferred from the data does not always correspond to the medical understanding of causality, making such networks less understandable to the user even if they have adequate performance.

One of the problems with such machine learning approaches is that a large amount of data is required to learn even a small network. In the Heart Disease Program, we have only used data to substantiate probabilities in very focused situations, such as the changes in the probabilities of various cardiovascular conditions with diabetes, and only then when there was independent literature to support the probabilities because of the potential biases of our own data.

3.1. Classification

Because of the difficulty in developing large models from data most machine learning in the medical context has focused on tightly constrained diagnostic problems, more properly called ''classification problems''. For example, a problem that has received considerable attention is whether or not a patient in the emergency room experiencing chest pain or shortness of breath is having a myocardial infarction (heart attack). Three approaches have emerged to deal with such problems: classification trees, logistic regression, and neural networks. They all take the same view of the problem: there is a set of cases, each consisting of a set of values to a fixed set of parameters and each having an appropriate classification. Usually the classification is binary, although only logistic regression requires this. The three approaches take different views of the statistical properties of the data and produce different kinds of solutions.

3.2. Classification trees

Classification trees [43,9] recursively pick a parameter on which to partition the data set into high and low concentrations of the desired classification. As it recurses, each subproblem only considers the data that made it to that branch. Thus, there is a tendency for the branches near the leaves to be less generalizable than the top branches since they are based entirely on small subsets of the data. Classification trees have the advantage that the criteria they use to make decisions is explicit — it is a series of questions answered with yes or no. A number of efforts have improved the performance of classification trees by such means as generating a number of trees and using a voting scheme to do classification. While improving the performance of the trees, this also makes them much more difficult to understand.

3.3. Logistic regression

Logistic regression [42] is a method from the statistical community used to solve the classification problem. This method takes the view that some subset of the parameters makes independent contributions in different degrees to the classification. These become the parameters of the logistic equation computing the probability of the classification. The advantage of this approach is that the minimal set of parameters contributes to the classification and their relative contributions are explicit. For example, in the classification

of myocardial ischemia [21], each case had 59 parameters but only seven made it into the final equation. Thus, it is easy to understand what goes into making the decision. The disadvantage is that no benefit is derived from the rest of the parameters. This might be important for a case in which one or more of the desired parameters are unknown and some excluded parameter or combination would provide that information. The only way to deal with missing parameters is to generate multiple logistic regression equations using different subsets of the available parameters.

3.4. Neural networks

The neural network approach [7] takes all of the parameters as nodes, optimizes the weights of those as inputs to a layer or two of hidden nodes and the weights of those nodes combine to determine the classification. The result is a non-linear combination of the inputs contributing to the classification, optimized for the available data. The non-linear relationship that is produced makes it possible to represent complex parameter interactions that could only be handled with logistic regression by adding explicit functions of multiple variables (products, etc.). The disadvantage of neural networks is the difficulty in understanding what they do. That is, they essentially put all of the inputs through a black box that produces a classification (actually a number that can be interpreted as a probability). Many people have made some progress toward glimpses inside the black box, such as average measures of the contributions of each input, but nothing has proven satisfactory yet for convincing the user that the data was appropriately considered.

As for the relative performance of these methods, our own studies on a large data set for classifying myocardial ischemia in the emergency room have shown neural networks to be the most accurate, logistic regression second, and classification trees (generating a single tree) the least [34,46]. However, the ROC curves for all three methods were close enough that it would be hard to show a significant difference in a prospective study. All three are viable methods for doing classification and the appropriate selection of a method needs to consider how the user will understand the result as well as the performance of the method.

4. Representation of knowledge

One of the lessons learned in developing reasoning methods for medical problems is the importance of an explicit representation for knowledge in a program. The development of a medical decision support system is the melding of two different disciplines and ways of thinking. Providing a clean separation between the medical knowledge and the reasoning method allows each to be considered in its own appropriate context by the appropriate experts. Exactly where this separation should take place is dependent on the medical problem. For example, process knowledge, decisions about what to do next and what data to gather may be considered part of the medicine or part of the reasoning. In some contexts, the program needs to make use of whatever data is available, identifying the uncertainty as appropriate and the gathering of the data is an external issue. In other contexts, the process of gathering the data is part of the medical knowledge that needs to be in the knowledge

base — the kinds of medical situations where a protocol captures the current best medical practice.

Certainly there is a whole body of research on knowledge representation, but most of the issues are orthogonal to the issues of the kinds of reasoning we are discussing. The essential features of a suitable knowledge representation are that it has clear semantics, is easy for the domain expert to understand and manipulate, and that it provides efficient indexing of the kinds needed by the reasoning engine.

4.1. Explanation

Some of the advantages of an explicit representation became clear in the development of explanation systems for programs [10]. Unless the explanation is generated from the same knowledge that the program was using to do the reasoning, there is no way to guarantee that the explanation reflects what the program knows or is doing. It might be argued that it is faster and easier just to write an explanation facility once the program is complete (being very careful that it is accurate!). However, programs, especially ones based on complex evolving knowledge, are never complete. In practice, the problem is to keep the explanations consistent with the changing knowledge base and the only way to do this is for both to operate from the same knowledge base. Furthermore, it was discovered that an explanation facility is useful to the developers in determining whether the knowledge base is consistent and whether the program is correctly using the knowledge. The single representation is also useful in generating suitable interfaces for the program. For example, the Heart Disease Program uses the knowledge base to generate the input forms used to enter a case. There is a module that goes through the findings used in the knowledge base, combines these with formatting information and generates the file that drives an HTML form generator. This may sound complicated, but it has saved considerable work that would have been required to update the forms as the knowledge base evolved.

The irony of explanations is that users rarely use them in the interactive way that was first expected. Even so, explanation facilities have proven to be of great benefit by providing justification for the results and by making the functioning of the program accessible to the developers.

5. Discussion

This is a necessarily partial and very sketchy overview of the progress in reasoning mechanisms. There are many other systems that could be mentioned, some of which probably had more influence than some that are listed. Thus, apologies if your favorite system was not mentioned. However, the intent was to review the degree of coverage we have achieved relative to the relationships in medicine suitable for automated reasoning.

It is clear that we have acquired a significant armamentarium to face the challenges of medicine. The set of tools is far from perfect. In some areas, such as production rules or Bayesian networks, there are off-the-shelf tools for implementing a system. In other areas, such as temporal reasoning, even though much research has been done, there are no tools

that one would feel comfortable recommending to someone new to the field. More research must be done before the right set of tools emerges. In other areas, such as reasoning with location information, there are good ideas and the theoretical basis for good tools, but few compelling domains to push the development forward.

5.1. Picking a method

So how does one pick a method when facing the need for a decision support system in a particular domain? The typical answer is to pick the method with which you have most experience. A better answer can be determined from the answers to three questions: (1) What kind of result is needed? (2) What is the nature of the medical knowledge? (3) Where will the knowledge come from? The result may range from a simple yes or no or probability to a complete explanation of how a set of findings is accounted for by diseases, physiologic mechanisms, and complications. The medical knowledge may range from associations to exact functional relations to rules for determining similarity. The knowledge may come from experts or need to be learned from data. Usually, the knowledge will be of several kinds and since there are a few tools that effectively handle more than one kind of relation, the challenge is to pick the primary organizing principle through which the problem will be viewed and then invent ways of handling knowledge that does not fit the paradigm. When considering the knowledge, one must always be asking the question: ''How will this make a difference clinically?'' That is, what aspects of the knowledge need to be accessible to the reasoning engine to produce the conclusions that are important to the user. Keeping this in mind allows the designer to make the tradeoffs in organizing the system that will allow the system to capture the essence of the medicine and still run effectively. Given the incomplete state of reasoning methods, it may be necessary to invent a new methodology to serve the requirements of the domain. However, there are now enough tools using enough of the relationships available in medicine that it is generally advisable to first implement a prototype using the tools most closely matching the characteristics of the domain and then assess its shortcomings in the situations that really arise rather than ones that are theoretically possible.

We have an assortment of reasoning methods and a good deal of experience building systems that do reasoning from diagnosis to therapy planning to patient management tasks in a variety of medical domains. What now? The Internet has provided us with connectedness; speech understanding technology is easing the input burden; ubiquitous inexpensive computers provide the platforms. It only remains for us to infuse decision support into the normal practice of medicine. That is, answer the questions the various care providers need answered in a way that requires no effort on their part beyond that necessary to carry out the normal functions of medicine.

Acknowledgements

The Heart Disease Program and the derivative research on case-based reasoning was supported by National Institutes of Health Grant No. R01-HL33041 from the National Heart, Lung, and Blood Institute.

References

[1] Aghassi DS. Evaluating case-based reasoning for heart failure diagnosis. MIT/LCS/TR-478, 1990.
[2] Aliferis CF, Cooper GF. A structurally and temporally extended Bayesian belief network model: definitions, properties, and modeling techniques. In: Proceedings of the Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, 1996, pp. 28–39.
[3] Andersen SK, Olesen KG, Jensen FV, Jensen F. HUGIN — a shell for building Bayesian belief universes for expert systems. IJCAI 1989:1080–5.
[4] Banks G, Weimer B. Symbolic coordinate anatomy for neurology (SCAN). J Med Syst 1984;8:157–62.
[5] Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: an evolving diagnostic decision-support system. J Am Med Assoc 1987;285:67–74.
[6] Bauer E, Koller D, Singer Y. Update rules for parameter estimation in Bayesian networks. In: Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI), Providence, Rhode Island, August 1997.
[7] Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. Ann Intern Med 1991;115:843–8.
[8] Bleich HL. Computerbased consultation: electrolyte and acid–base disorders. Am J Med 1972;53:285–91.
[9] Breiman L, Friedman JH, et al. Classification and regression trees. Belmont (CA): Wadsworth, 1984.
[10] Buchanan BG, Shortliffe EH. Rule Based Expert systems: the mycin experiments of the stanford heuristic programming project. Reading (MA): Addison-Wesley, 1984.
[11] Citro G, Banks G, Cooper G. INKBLOT: a neurological diagnostic decision support system integrating causal and anatomical knowledge. Artif Intell Med 1997;10:257–67.
[12] Coleman TG, Randall JE. HUMAN: a comprehensive physiological model. Physiology 1983;26:15–21.
[13] Cooper GF. NESTOR: a computer-based medical diagnostic aid that integrates causal and probabilistic knowledge. Ph.D. Thesis, Stanford, 1984.
[14] Cooper G, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. Mach Learn 1992;9:309–47.
[15] First MB, Soffer LJ, Miller RA. Quick (quick index to caduceus knowledge): using the internist-1/caduceus knowledge base as an electronic textbook of medicine. Comput Biomed Res 1985;18:137–65.
[16] Gertner AS, Webber BL, Clarke JR, Hayward CZ, Santora TA, Wagner DK. On-line quality assurance in the initial definitive management of multiple trauma: evaluating system potential. Artif Intell Med 1997;9:261–82.
[17] Glymour C, Scheines R, Spirtes P, Kelly K. Discovering causal structure. New York: Academic Press, 1987.
[18] Guyton AC, Coleman TG, Granger HJ. Circulation: overall regulation. Ann Rev Physiol 1972;34:13–46.
[19] Haimowitz IJ, Le PP, Kohane IS. Clinical monitoring using regression-based trend templates. Artif Intell Med 1995;7:473–96.
[20] Heckerman D. Probability similarity networks. New York: MIT Press, 1990.
[21] Heckerman D, Geiger D, Chickering D. Learning Bayesian networks: the combination of knowledge and statistical data. Mach Learn 1995;20:197–243.
[22] Jang Y. HYDI: a hybrid system with feedback for diagnosing multiple disorders. Ph.D. Thesis. New York: MIT Press, 1993.
[23] Kim JH, Pearl J. A computational model for causal and diagnostic reasoning in inference systems. IJCAI-83 1983:190–3.
[24] Kingsland III LC. The Evaluation of medical expert systems: experience with the ai/rheum knowledge-based consultant system in rheumatology. In: Proceedings of the Ninth Symposium on Computer Applications in Medical Care, Bethesda, MD, 1985, pp. 292–5.
[25] Kolodner JL, Kolodner RM. Using experience in clinical problem solving: introduction and framework. GIT-ICS-85/21, Georgia Institute of Technology, 1983.
[26] Koton PA. A medical reasoning program that improves with experience. In: Proceedings of the Symposium on Computer Applications in Medical Care Conference, Washington, DC, 1988.
[27] Koton PA, Using experience in learning and problem solving. MIT/LCS TR-441, 1989.
[28] Kuipers B. Qualitative simulation. Artif Intell 1986;29:289–338.
[29] Kulikowski CA, Weiss SM. Representation of expert knowledge for consultation: the Casnet and Expert

projects. In: Szolovits P ,editor. Artificial intelligence in medicine. Boulder (CO): Westview Press, 1982, pp. 21–55.
[30] Lauritzen SL, Spiegelhalter DJ. Local computations with probabilities on graphical structures and their application to expert systems. J R Statist Soc B 1988;50:157–224.
[31] Lindberg D, Sharp G, Kingsland L, Weiss S, Hayes S, Ueno H, Hazelwood S, Computer-based rheumatology consultant. In: Proceedings of the Third World Conference on Medical Informatics, Tokyo, 1980, pp. 1311–5.
[32] Long W. Medical diagnosis using a probabilistic causal network. Appl Artif Intell 1989;3:367–83.
[33] Long W, Naimi S, Criscitiello MG. Development of a knowledge base for diagnostic reasoning in cardiology. Comput Biomed Res 1992;25:292–311.
[34] Long W, Griffith JL, Selker HP, D'Agostino RB. A comparison of logistic regression to decision tree induction in a medical domain. Comput Biomed Res 1993;26:74–97.
[35] Long W. Temporal reasoning for diagnosis in a causal probabilistic knowledge base. Artif Intell Med 1996;8:193–215.
[36] Miller RA, Pople HE, Myers JD. Internist-I, an experimental computer-based diagnostic consultant for general internal medicine. New Engl J Med 1982;307:468–76.
[37] Olesen KG, Kjaerulff U, Jensen F, Jensen FV, Falck B, Andreassen S, Andersen SK., A MUNIN network for the median nerve — a case study on loops. In: Horn W, editor. Causal AI models hemisphere, 1990, pp. 301–9.
[38] Oppel UG, Hierle A, Janke L, Moser W. Transformation of compartmental models into sequences of causal probabilistic networks. In: Andreassen S, et al., editors. Artificial intelligence in medicine. Amsterdam: IOS Press, 1993, pp. 87–98.
[39] Patil RS. Causal representation of patient illness for electrolyte and acid–base diagnosis. MIT/LCS/TR-267, October 1981.
[40] Pearl J. Fusion, propagation, and structuring in Bayesian networks. Artif Intell 1986;29:241–88.
[41] Pearl J. Causality: models, inference and reasoning. Cambridge: Cambridge University Press, 2000.
[42] Pozen MW, D'Agostino RB, et al. A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease, New Engl J Med 1984;310:1273–8.
[43] Quinlan JR. Induction of decision trees. Mach Learn 1986;1:81–106.
[44] Reggia JA, Nau JS, Wang Y. Diagnostic expert systems based on a set covering model. Int J Man-Mach Stud 1983;19:437–60.
[45] Rutledge G, Thomsen G, Farr B, Tovar M, Polaschek J, Beinlich I, Sheiner L, Fagan L. The design and implementation of a ventilator-management advisor. Artif Intell Med 1993;5:67–82.
[46] Selker HP, Griffith JL, Patil S, Long W, D'Agostino RB. A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. J Invest Med 1995;43:468–76.
[47] Szolovits P, Pauker SP. Pedigree analysis for genetic counseling. In: Lun KC, et al., editors. Proceedings of the Seventh Conference on Medical Informatics, MEDINFO 92, Amsterdam: Elsevier, North Holland, 1992, pp. 679–3.
[48] Weed LL. New connections between medical knowledge and patient care. Br Med J 1997;315:231–5.