From: AAAI-83 Proceedings. Copyright ©1983, AAAI (www.aaai.org). All rights reserved.

REASONING ABOUT STATE FROM CAUSATION AND TIME IN A MEDICAL DOMAIN

by
William J. Long[*]

Clinical Decision Making Group, Laboratory for Computer Science
Massachusetts Institute of Technology, Cambridge, Massachusetts

Abstract

The reasoning needed for diagnosis and patient management in a medical domain requires the ability to determine both the aspects of the patient state that are definitely known and those that are possible given what is known about the patient. This paper discusses a mechanism for including the time constraints of causal relationships in the representation and the increased discriminatory power of the reasoning mechanisms when the time relationships are used appropriately. It is further argued that in such a domain where time bounds are weak there is often more information in the relationships between times than in the time values themselves. Thus, it is often necessary to reason from the relationships rather than by comparing time values. A program utilizing the mechanisms outlined is currently under development.

Introduction

The issues of time, causation, and change are central to reasoning in many domains. Since processes take place over time, accounting for the relationships of the changes is essential for the understanding of the processes. The problems of causation and change have been encountered in such expert systems domains as geology, molecular genetics, naive physics, and medicine [Simmons, 1982, Stefik, 1981, Forbus, 1982, Patil *et al.,* 1981]. While considerable research has been focused on causation, the subject has not been exhausted. Other research has been directed specifically toward representing and reasoning about time relationships (see for example [Allen, 1981, Kahn and Gorry, 1977, McDermott 1981]). These efforts have been concerned primarily with answering questions about a history of events, including uncertain time bounds [Kahn and Gorry, 1977] and future events [McDermott 1981]. None of these efforts have looked at the problem of using knowledge about classes of events to answer questions about what events might or must have taken place, especially when there are interactions among processes. The medical domain offers a particular challenge because the patient state has to be inferred from the data, requiring knowledge of the processes involved. Older programs such as Internist [Pople, 1977] and PIP [Pauker *et al.*, 1976] have avoided this issue by treating the possible presentations probabilistically. However, such a methodology neglects time information that may make it possible to discriminate between competing states. This is especially true in programs that reason about patients during the treatment process when the state of the patient is changing. An example from cardiology will illustrate the timing relationships that can arise[**].

If a patient has heart failure caused by a weakened cardiac muscle, the low cardiac output causes water retention which increases the blood volume. The high blood volume can cause edema to develop (abnormal fluid accumulation in the lungs, legs, and possibly other areas). If a diuretic is given to clear the excess fluid, it acts by decreasing the blood volume. As the blood volume decreases, the edematous fluid returns to the circulation. However, it may take time to mobilize the fluid, especially in the legs, and if the diuretic reduces the blood volume more than is appropriate, the low blood volume may cause a further reduction of the cardiac output.

Not all of these processes are instantaneous. In fact there is wide variation in the amount of time taken. If the patient has heart failure or low blood volume, the immediate consequence is low cardiac output. If there is water retention, it will still be a matter of days before the patient has taken in enough fluid for the blood volume to be high. If the blood volume is high — say it happened rapidly from excessive fluid therapy — it would still take at least a few hours for edema to develop. Conversely, once edema has developed it may take days for the fluid to return to the blood stream even though the blood volume is no longer high. The diuretic also takes hours to remove the excess fluid from the blood volume. Because of these time relationships, there are certain combinations of states that simply could not occur. For example, edema would not exist unless heart failure had existed for days or the patient had been given fluid therapy at least hours earlier. However, it is still possible for edema to be present in the patient at the same time blood volume is low if diuretics had been administered hours earlier.

---

[*] This investigation was supported in part by National Library of Medicine Grant No. 5 P01 LM03374-04, in part by the Whitaker Health Sciences Fund, and in part by BRSG S07 RR 07047-17, awarded by the Biomedical Research Support Grant, Division of Research Resources, National Institutes of Health.

[**] The examples used should not be taken to represent medical reality since there are many factors not included and the relationships of the factors included are greatly simplified, but the classes of relationships represented are characteristic of the larger problem and are sufficient to present the issues.

Representation of State, Time, and Causation

The appropriate representation for a problem depends on the kinds of reasoning to be done and the characteristics of the domain. The reasoning we are concerned with is found in both diagnosis and patient management — we need to determine what is *known* about the patient state from the available information, including what is definite and what is uncertain but possible. It is also useful to assign likelihood measures to *possible* conclusions, but that is beyond the scope of this paper — the determination of possible and definite in this paper can be viewed as a controlled framework in which a likelihood mechanism could operate.

In medicine the meanings of measurements of the patient state tend to vary from patient to patient. While "normal" ranges are identified for many parameters, they are only confidence intervals for the population. The important determination is the relationship between parameters. That is, a parameter is abnormal because it has an effect on something else. Physicians tend to express this by using *qualitative* measures for parameter values (*e.g.*, low cardiac output) even when quantitative measures are available. Thus, we have chosen to represent parameter states qualitatively with the values defined in terms of other parameters. For example, blood volume may be low, normal, or high. A low blood volume is a blood volume that *will result in low* cardiac output, whatever the value might be for an individual patient. Defining the qualitative values in this way means that it is usually not possible to associate exact numbers with the values — there are no universally valid "boiling points" as in physical systems [Forbus, 1982]. This places more burden on the interpretation of patient measurements, but simplifies the reasoning problem.

Similarly, *time relations* in medicine tend to be inexact. For example, there is no exact time when low cardiac output started. Although one can say it has been present for days but not weeks, more precision is impossible. Expressing the templates for causality also requires that the range of possible time delays between cause and effect be represented. The bounds of these ranges are likewise difficult to express in exact terms — it takes from hours to days to mobilize edema. Rather than require artificially exact bounds to be specified, we use *qualitative times* related by a partial ordering. Thus the time bounds of a relation can be viewed as a qualitatively specified confidence interval.

Causation in the medical domain has other characteristics. 1) The factors influencing a particular parameter are limited and each can be suitably represented. Thus, it is appropriate to reason in terms of a closed domain of possible influences. 2) Some of the cause-effect relations are only probabilistic at the present state of medical knowledge. These will be represented as possible, without dealing with the likelihood. 3) The domain is also dominated by stable feedback systems. As a result, it is computationally more reasonable to accept the tendency of the systems to return to stable states as given and only represent the influences on the abnormal states.

Thus, the representation of potential influences on an abnormal parameter value include the possible causes for the value with the time relationships between cause and effect, the possible corrections for the value with those time relations, and the time requirements for the parameter to return to normal in the absence of causes or corrections. The causes and corrections can further be divided into those that make the state possible and those that will definitely result in the state. This yields the following template for representing the causes of a state:

P + : causal conditions that make the state possible
D + : causal conditions that make the state definite
P – : corrective conditions that possibly stop the state
D – : corrective conditions that definitely stop the state
relax : time range for the state to end after the cause ends

The causes for *high blood volume* within the simplified domain are represented as follows:

P + retain water (t1), fluid therapy (0)
D + retain water (t2)
P – no fluid therapy and lose water (t3)
D – no fluid therapy and lose water (t4)
relax (t5 t6) *; minimum and maximum relaxation times*

Retaining water for at least time t1 could cause high blood volume as could fluid therapy for any amount of time. Retaining water for time t2 would definitely cause high blood volume. If the patient stops retaining water and does not receive fluid therapy, the blood volume could return to normal possibly by time t5 and certainly will by t6. Actively losing water will hasten this process, restoring normal blood volume possibly by time t3 and certainly will by t4. This representation of cause extends the notion of *continuous causation* and *thresholds* to include the temporal relationships between cause and effect [Reiger and Grinberg, 1977]. The resulting representation is sufficient to represent the properties needed for answering questions about the patient state.

Many rules about causation are general. For example, all abnormal states have causes; the effect depends on the sum of the influences; and causes must start prior to or simultaneously with the effect. In this domain there are additional properties useful as rules for reasoning: 1) The cause and the effect must overlap, *i.e.*, the cause can not end before the effect starts. 2) The state changes must be from the adjacent states, *e.g.*, for retaining water to cause high blood volume the blood volume must be normal already. Since the causes usually also correct the opposing abnormal state, this is useful to keep the time bounds more precise where possible. 3) Causation once started continues until there is some change either in the cause or the corrective influences. Thus, if high blood volume were causing edema, it would continue to cause edema until the cause were changed, even though high blood volume does not necessarily cause edema (in this model). (In the actual domain it is also necessary to represent *precipitating factors* — factors that can mean the difference between causing and not causing the effect, but are incapable of causing the effect themselves. The extension follows without difficulty.)

An Example

The causal relationships for the example are sketched in figure 1 (without the corrections or time relations). Given the corresponding relationships represented by the formalism, the program is able to determine what is definitely known about the patient by propagating the minimum conclusions from the causal relationships. If the patient presently has edema and has not had

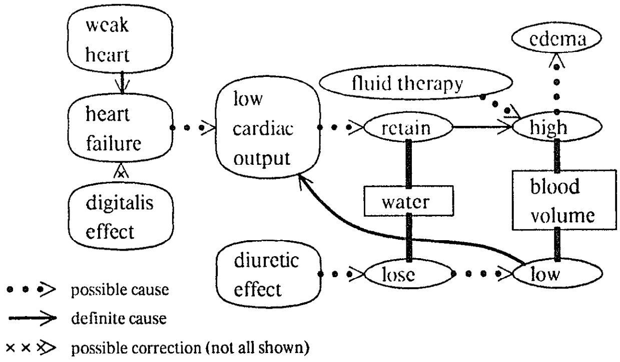

Figure 1. Example Causal Relations

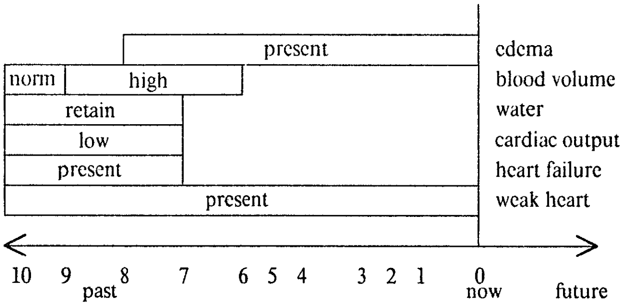

Figure 2. Edema without Fluid Therapy

fluid therapy, the conclusions (somewhat overconstrained by the diagram) are illustrated in figure 2 and deduced as follows:

edema ⇒ high blood volume

High blood volume is the only available cause. It must exist or have existed to account for the edema present. The time from 0 to 6 (t0-6) in the figure is the maximum relaxation time for the edema. That is, the high blood volume had to be true at least that recently or the edema would definitely have disappeared by now. The difference between times 8 and 9 (t8-9) is the minimum time for which high blood volume must persist to cause edema. It is not specified when the edema started, but the high blood volume must have started at least that long before the start of the edema.

high blood volume ∧ fluid therapy: not given ⇒ water retention

The water retention must not have ended more than t6-7 prior to the end of the high blood volume and must have started at least t9-10 prior to the beginning of the high blood volume. Since to cause high blood volume the blood volume must be normal at least for a time, normal blood volume must exist at least t9-10 prior to the high blood volume. (Normal blood volume and retaining water do not have to start at the same time. They just have the same minimal time relationship to high blood volume.)

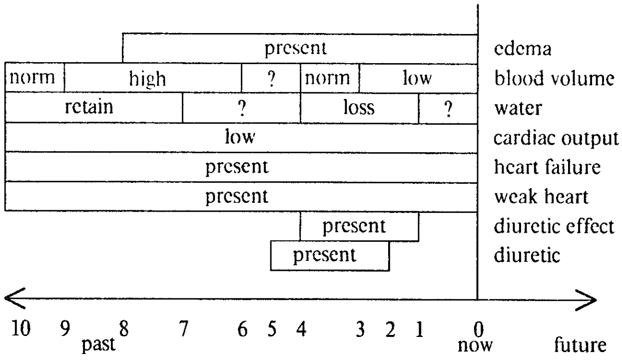

Figure 3. Edema and Low Blood Volume

water retention ⇒ low cardiac output

The other causes have no delay and are therefore simultaneous with the water retention.

low cardiac output ⇒ heart failure V low blood volume

However, low blood volume is not consistent with the normal and high blood volume that must overlap the required time period.

heart failure ⇒ weak heart

The relaxation time for a weak heart is infinite. Thus while the weak heart is deduced true from t7 to t10 to account for the heart failure, once it exists it must remain.

If we assume that nothing is known about therapy in the patient other than the lack of fluid therapy, then nothing more can be concluded. The result of this deduction is a representation of the set of facts that must hold given the edema.

The example has been extended in figure 3 to show that it is possible for the patient to have low blood volume even though edema is present. If we assume the patient has not received digitalis (which could correct the heart failure), the heart failure and low cardiac output must have continued until time *now* because the cause has not changed.

low blood volume ⇒ water loss

The patient must have been losing water at least until time t1 ago, having started losing water and having had a normal blood volume at least t3-4 prior.

water loss ⇒ diuretic effect
diuretic effect ⇒ diuretic

The program can determine a specific time interval during which diuretic therapy would have to have been given for there to be low blood volume now. The more common deduction of the possible effects from known times of therapy could also be made.

For the low blood volume to be consistent with the edema, the end of the high blood volume must be before the start of normal blood volume (and likewise for water retention and loss). Therefore since the maximum time for the edema to be removed is longer than the minimum time that water loss takes to cause low blood volume, it is possible for both to exist simultaneously.

Discussion

There are three observations to be made about the deductions facilitated by the mechanisms presented. First, the mechanisms are capable of discrimination that would be impossible without the use of the time relationships. The ability to filter out the impossible from among the possible contingencies allows a program to have more focused reasoning about the legitimate possibilities. Second, the mechanisms permit reasoning both about the definite and the possible. Thus, the same mechanism is useful for determining what is known about the patient state and deducing the possible effects of therapy.

Finally, let us consider a little more deeply the reasoning that eliminated low blood volume as a cause of the low cardiac output. All that is known about the time of the high blood volume is that it must have begun at least t8-9 (say hours) ago but could have begun and ended t0-6 (say days) ago. Similarly, the cause of the low cardiac output began at least t8-9 + t9-10 (hours) ago but could have begun and ended t0-6 + t6-7 (days) ago. It is impossible to conclude that the low blood volume is incompatible with the high blood volume by comparing these ranges. The important information is the temporal relationship between cause and effect. The cause for the low cardiac output must continue until the low cardiac output ends, which is when the retaining of water ends, which is necessarily after the blood volume is high. Thus, any representation of the times of these states must preserve both the constraints on the times and the relationships between the states to make all of the reasoning possible.

Acknowledgements

The work reported here is part of the development of a program to assist the physician in the diagnosis and management of heart failure [Long *et al.,* 1982]. The ideas are presented as separate from that work, but the results will become a part of that work as it progresses. For an understanding of some of the problems and issues confronting the physician, I thank our collaborators on that project, Drs. Criscitiello, Naimi, and Pauker of New England Medical Center Hospital.

References

[1] Allen, J. F., "Maintaining Knowledge about Temporal Intervals," *University of Rochester Department of Computer Science TR 86*, January 1981.

[2] Forbus, K. D., "Qualitative Process Theory," *MIT AI-Memo 664*, February 1982.

[3] Kahn, K. and Gorry, G. A., "Mechanizing Temporal Knowledge," *Artificial Intelligence 9*, (1977) 87-108.

[4] Long, W., Naimi, S., and Criscitiello, M. G., "A Knowledge Representation for Reasoning about the Management of Heart Failure," *Computers in Cardiology Conference*, October 1982.

[5] McDermott, D., "A Temporal Logic for Reasoning About Processes and Plans," *Yale University Department of Computer Science Research Report 196*, March 1981.

[6] Patil, R. S., Szolovits, P., and Schwartz, W. B., "Causal Understanding of Patient Illness in Medical Diagnosis," *Proceedings of IJCAI-7*, August 1981, pp. 893-899.

[7] Pauker, S. G., Gorry, G. A., Kassirer, J. P., and Schwartz, W. B., "Toward the Simulation of Clinical Cognition: Taking a Present Illness by Computer," *The American Journal of Medicine 60*, (1976) 981-995.

[8] Pople, H. E., Jr., "The Formation of Composite Hypotheses in Diagnostic Problem Solving: an Exercise in Synthetic Reasoning," *Proceedings of IJCAI-5*, August 1977, pp. 1030-1037.

[9] Reiger, C. and Grinberg, M., "The Declarative Representation and Procedural Simulation of Causality in Physical Mechanisms," *Proceedings of IJCAI-5*, August 1977, pp. 250-256.

[10] Simmons, R. G., "Spatial and Temporal Reasoning in Geologic Map Interpretation," *Proceedings of AAAI-82*, August 1982, pp. 152-154.

[11] Stefik, M., "Planning with Constraints (MOLGEN: Part I)," *Artificial Intelligence 16*, (1981) 111-140.