REPRESENTATION OF EMPIRICALLY DERIVED CAUSAL RELATIONSHIPS

Robert L. Blum

Department of Computer Science
Stanford University
Stanford, California 94305

ABSTRACT

The objective of this paper is to present a method for the computer representation of empirically derived causal relationships (CR's). This method draws on the theory of multivariate linear models and path analysis. The method is contrasted with the predicate calculus based methods used by most researchers in artificial intelligence.

The representation presented here has been used to store information on medical CR's derived empirically from a large clinical database by a computer program called RX. The principal emphasis in the representation is on capturing the intensities and variances of effects and the variation in the effects across a patient population. Once incorporated into RX's knowledge base, this information is subsequently used by RX in determining the validity of other CR's.

The representation uses a directed graph formalism in which the nodes are frames and the arcs contain seven descriptive features of individual CR's: intensity, distribution, direction, mathematical form, setting, validity, and evidence.

Because natural systems (such as the human body) are inherently probabilistic, linear models are useful in representing causal flow in them. Knowledge of natural systems is fundamentally probabilistic because of 1) Irreducible indeterminism in their component processes, 2) difficulties in accurately measuring all relevant variables, 3) variation among individuals in a population, and 4) inadequate scientific theory.

The principal objective of this paper 1s to present a method for the computer representation of causal relationships relevant to clinical medicine. The representation presented here is used to store information on clinical causal relationships in the medical knowledge base of a large computer program called RX.

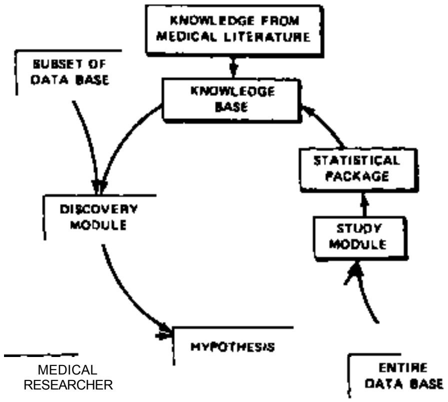
In this brief report I will touch on the following topics: 1) the objectives and methods of the RX Project, 2) the characteristics of the tasks that RX performs that Influence the admissible forms of representation for causal relationships, 3) the method of representation, and 4) a comparison of this method with the work of other AI researchers.

I. THE RX PROJECT: AUTOMATED STUDY AND INCORPORATION OF CAUSAL RELATIONSHIPS

Before presenting our method for representing causal relationships (CR's), it 1s helpful to know the research context in which it was elaborated. Our research project, called the RX Project, was begun in 1978 and 1s a multidisciplinary research effort whose purpose is to develop techniques for deriving various types of medical knowledge from clinical databases. To date we have been exclusively concerned with the detection and study of causal relationships (CR's) in our databases and with their subsequent Incorporation and use by the program.

Specifically, the objects of the project are 1) to increase the validity of CR's derivable from large time-oriented clinical databases, 2) to develop methods for providing intelligent assistance with the task of testing hypothesized CR's against a database, and 3) to study methods for automating the process of discovering CR's. The RX Project is definitively described in [Blum 1982a] and is summarized in [Blum 1982b].

The RX methodology for deriving possible causal relationships from a clinical database employs the following components: a knowledge base (KB), a Discovery Module, a Study Module, a statistical package, and a clinical database. In brief, the system works as follows. The Discovery Module examines relevant subsets of the database to generate an ordered list of causal hypotheses. These hypotheses, of the form "A causes B," are sequentially examined by the Study Module. The Study Module uses the knowledge base to generate a comprehensive epidemiological study design of the hypothesis. This study design is then tested on the statistical package using the entire database. The results are passed back to the Study Module for interpretation. If the results are medically important as well as statistically significant, they are written as a new, machine-readable causal relationship Into the knowledge base. The process of automated study design makes use of previously "learned" causal relationships. The discovery, confirmation, incorporation cycle of RX 1s shown below* The clinical database we have used is a 1700 patient subset of the ARAMIS database [Fries 1979] [McShane 1979] occupying 15,000 pages. RX is written in INTERLISP.



II. DESIGN CRITERIA FOR THE REPRESENTATION OF CAUSAL RELATIONSHIPS FOR RX

The form in which we have chosen to represent CR's in RX has been strongly influenced both by the necessity of capturing detailed information on them and by the necessity of using that information for the subsequent study and confirmation of other CR's.

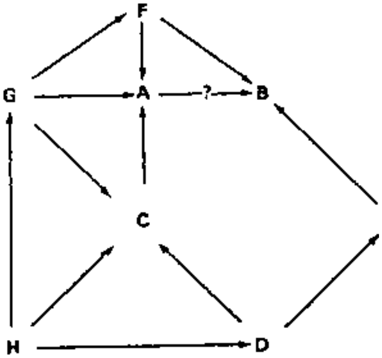
The principal objective of the RX Project is to derive and incorporate detailed knowledge of causal relationships from large time-oriented databases. This knowledge is stored interchangeably with knowledge entered into the medical KB from the medical literature. It is necessary that the representation of CR's be sufficiently rich to enable capturing information on magnitude, frequency, and variability of effect, distribution in a patient population, mathematical form, clinical setting, validity or reliability, and evidential basis.

Although the representation we have designed enables most of these aspects to be encoded in machine-readable form, the most important motivating factor in designing an adequate representation is the tasks for which the encoded information will be used.

The only task we will describe here is the Study Module's use of CR's in creating a study design for a causal hypothesis. This task requires information on the intensity of causal links, their mathematical form, their distribution across patients, and their clinical setting.

A critical step in the design of a study by the-RX Study Module Involves the selection of the set of known clinical events that may confound or bias the results of a study. This set of events 1s known as the set of confounding variables. The control of confounding variables is an essential step in the design of studies using routine health care data. A confounding variable is one that may affect both the causal variable and the effect variable of interest. The objective of control 1s to attempt to isolate the relationship from spurious causal Influence. For example, if A 1s a drug and B is a side effect of Interest, we would like to control for diseases that affect both A and B.

The task of demonstrating that a causal relationship is nonspurious is by far the most difficult task in deriving CR's from large clinical databases. Unfortunately, the confounding variables may exert their influence quite indirectly as shown below. In the figure there are four confounding variables: F, G, H, and D. If we were, however, to examine only the list of variables that directly affect B, we would only find F and E. The node E is not a confounding variable, since it is known not to affect A.



To determine the set of confounding variables for a hypothesis "A causes B," the Study Module uses a function called Confoundinq-Variables to traverse a directed graph whose arcs are CR's. The function determines the set of all nodes that may have medically significant effects (greater than some magnitude) on both A and B for a given clinical setting.

The Study Module actually controls for only a subset of the confounding variables called the causal dominators. This subset is defined as the smallest subset through which all known causal influence on both A and B must flow. In the figure this set = { F D }.

III. THE REPRESENTATION OF CAUSAL RELATIONSHIPS IN THE RX KB

In brief, CR's are represented as labeled arcs 1n a directed graph 1n which the nodes are frames. For the sake of this discussion we will assume that X 1s a causal node and that Y is an effect node. They are connected together by a CR whose components will be described in detail. We specify both the X and Y objects as having real-valued intensities. In other words, in this discussion we will model them as real-valued variables. We assume the following relationship between their intensities:

$$Y(t + tau) = bf[X(t)] + e .$$

That is, Y's value at time t + tau is linearly related to some function of X's value at time t (+ an error term e). We further assume the relationship is causal. That is, a change 1n X of one unit induces a change 1n Y of bf[l] after tau times units. X 1s assumed to cause Y 1n the probabilistic sense that 1t accounts for some of its variance. In the usual path analysis or regression model, we could estimate the parameters of the model by using a database of pairs of measurements of X and Y. The estimate of greatest importance is the unstandardized regression coefficient b. If b is significantly different from zero, we then posit that X causes Y.

The basic information conveyed by the model is that an increase in X by one unit causes a change in Y by b units, where b is the unstandardized regression coefficient of Y on X. Labeled with just these regression coefficients, a simple directed graph can be set up to represent chains of causality.

<intensity distribution direction functional-form
setting validity evidence>

A causal link in the RX knowledge base is composed of seven components shown above. In what follows we have assumed that both the causal variable and the effect variable connected by this causal relationship are real-valued. In our future work we intend to generalize this formalism so that binary and rank-valued variables may also be arbitrarily connected.

The meaning of each of these seven components is summarized below:

Intensity: the expected change in the effect given a change in the cause, expressed as an unstandardized regression coefficient.

Distribution: the distribution of the Intensity of the effect across patients.

Direction: either increases or decreases.

Setting: the circumstances under which the causal relationship was derived, encoded as a Boolean expression with time-dependent predicates.

Functional Form: the complete mathematical model relating Y to X, encoded in an algebraic language.

Validity: the state of proof of the causal relationship on a 1 to 10 scale: 1 means highly tentative, 10 means beyond reasonable doubt.

Evidence: a summary of the evidence on which the relationship is based: either literature citations or a summary of the study performed by the Study Module.

In the RX Study Module the intensity and the direction components are derived from the fitted regression model that is stored in machine-readable form as the functional-form component.

The distribution component records the density function of the estimated regression coefficients across patients. In otherwords, this component enables us to record the varying intensities with which a population of patients exhibits the effect of interest. This capability for encoding unexplained variation in an effect 1s an Important aspect of our representation scheme. This density function 1s encoded by storing the mass under ten contiguous regions of the curve. The choice of the nine cut points 1s based on prior medical knowledge of the effect variable.

The setting component allows the explicit inclusion of the setting 1n which the causal relationship is believed to be true. For relationships that have been empirically derived by the Study Module, the setting component encodes the inclusion and exclusion criteria that were used to select time intervals from patient records for study. In English a typical setting might read "between two months and six months after myocardial infraction — but not during an episode of congestive heart failure." This is stored as a logical expression with time-dependent functions: for example, (Concurrent (After Myocardial-Infraction 2 months 6 months) (Not (During Congestive-Heart-Failure))).

The validity of a causal link depends on how extensively and under what circumstances it has been tested. Validity, as defined here, pertains to the state of proof of a causal relationship, and not to the relationship itself. Causal relationships are widely regarded as valid if they have been repeatedly confirmed, particularly 1n prospective, randomized studies. At the opposite extreme are relationships based on a single retrospective study of a small number of patients.

IV. LINEAR MODELS VERSUS PREDICATE CALCULUS

The representation for CR's presented here was strongly influenced by the methods of linear models, multivariate analysis, and path analysis. This body of theory has largely been developed and applied by psychologists,economists, and biologists. Excellent reviews appear in [Helse 19761, [Kenny 1979], and [Bentler 1980]. In contrast, the predicate calculus representations developed by AI workers (for example, [Rieger 1977}, [Rieger 1978], and [de Kleer 1981]) have largely been applied to the simulation and understanding of mechanical and electrical devices.

Why have multivariate linear models been used for certain applications and predicate calculus models for others? The answer is profound and important: linear models capture crucial features of natural systems, predicate calculus captures crucial features of artifacts.

Natural systems are Inherently probabilistic. Medical phenomena, at least at the clinical level, are typically quite Indeterminate. This probabilistic character arises from at least four sources: 1) the Inherently probabilistic nature of the component phenomena (at all levels of detail) that comprise the working human body, 2) our Inability as observers to accurately measure these phenomena in a given patient, 3) the variability of effects across patients, and 4) the Inadequacy of current biological theory as a basis for explanation. The role of probability In models of causality 1s lucidly discussed in [Suppes 1970].

Capturing this variability of clinical phenomena 1n a sufficiently detailed manner to allow Its subsequent scientific analysis dictates that detailed quantitative information on the intensities of effects and their variation be captured in the representation. This is largely what has motivated our adoption of multivariate linear models and extensions to them. In RX we start with detailed quantitative data in our database. We have tried to preserve as much of that detail as possible in the statistical summaries that comprise the data in the CP's.

In contrast, the predicate calculus representations developed by AI workers have largely been applied to the simulation of mechanical devices or to the modeling of human understanding of the CR's that describe these devices. Rieger et al. and de Kleer et al . share an interest in qualitative models, which they (and I) believe are used by human beings in understanding machines, as opposed to detailed mathematical models.

Central to the CSA Project of Rieger and Grinberg is a collection of types of causal links, which they feel provides a useful and comprehensive set for modeling mechanical devices. There are ten types of causal links in their set, and they have used them to model a host of devices. Their declarative representation may be transformed into a procedural representation for device simulation. The emphasis in [Rieger 1977] is on providing a qualitative functional description that emulates human understanding of a device.

All of the various link types proposed in [Rieger 1977], including one-shot enablement, threshold, antagonism, and rate confluence, may be simply represented using linear models. For binary-valued dependent variables logistic regression models may be used that allow for time-dependent independent variables.

To conclude, the CSA representation of Rieger and Grinberg and the tabular representations of de Kleer and Brown were designed to model mechanical devices with discrete states. The objective of these researchers has been the qualitative simulation or "envislonment" of the operation of physical devices. The RX representation, employing linear models and the methods of multivariate analysis, has been designed for real-valued, multivariate, probabilistic domains. The objective has been the detailed analysis and quantification of individual causal links.

V. CONCLUSION

I presented a simple representation for CR's appropriate for modeling clinical medicine. The principal emphasis in the model 1s on capturing intensities of effects across a patient population. The representation arose from the methods of multivariate linear models and path analysis.

Given that detailed quantitative Information is occasionally needed in medical AI programs, how does the user or the program avoid being bogged down in needless complexity? The solution is to maintain detailed Information in the KB, but to translate it, as needed, to appropriately simplified levels. Linear models, in particular, may be automatically simplified into predicate calculus forms (but not vice versa). A method for performing this transformation will be included in a forthcoming paper.

ACKNOWLEDGEMENTS

Funding for this research was provided by the National Center for Health Services Research (NCHSR) through grant HS-04389 and by the National Library of Medicine through grant LM-03370. Computation facilities were provided by SUMEX-AIM through NIH grant RR-00785 from the Biotechnology Resources Program. Clinical data were obtained from the American Rheumatism Association Medical Information System. ARAMIS is sponsored by NIH grants AM-21393 and HS-03802.

REFERENCES

[Bentler 1980] Bentler, P.M., Multivariate Analysis with Latent Variables: Causal Modeling; Annual Reviews of Psychology 31:419-456, 1980.

[Blum 1982a] Blum, R.L., Discovery and Representation of Causal Relationships from a Large Time-Oriented Database: The RX Project; A monograph in the Medical Informatics Series, edited by D. Lindberg and P.Reichertz, Springer-Verlag, 1982.

[Blum 1982b] Blum, R.L., Discovery, Confirmation, and Incorporation of Causal Relationships from a Large Time-Oriented Database: The RX Project; Computers and Biomedical Research 15:164-187, 1982.

[de Kleer 1981] de Kleer,J. and Brown,J.S., Mental Models of Physical Mechanisms and Their Acquisition; in John Anderson (ed.), Cognitive Skills and Their Acquisition, Erlbaum, 1981.

[Fries 1979] Fries,J.F. and McShane,D.J., ARAMIS: A National Chronic Disease Data Bank System; in Proc. of the 3rd Annual Symp. on Computer Applications in Medical Care, pp. 798-801, IEEE, Washington,D.C., Oct., 1979.

[Heise 1975] Heise.D., Path Analysis; John Wiley & Sons, 1975.

[Kenny 1979] Kenny,D., Correlation and Causality; John Wiley & Co., 1979.

[McShane 1979] McShane,D.J., Harlow,A., Kraines, R.G., and Fries.J.F., TOD: A Software System for the ARAMIS Data Bank; Computer 12:34-40, Nov. 1979.

[Rieger 1977] Rieger,C. and Grinberg,M., The Declarative Representation and Procedural Simulation of Causality in Physical Mechanisms; in Proc. of the 5th International Joint Conference on Artificial Intelligence, pp.250-255, IJCAI 1977.

[Rieger 1978] Rieger,C. and Grinberg,M., A System of Cause-Effect Representation and Simulation for Computer-Aided Design; in Latombe (ed.), Artificial Intelligence and Pattern Recognition in Computer-A1ded Design, pp.299-326, North-Holland, 1978.

[Suppes 1970] Suppes.P., A probabilistic Theory of Causality, North-Holland Publishing Co., Amsterdam, 1970.

[Szolovlts 1982] Szolovits.P. (ed.), Artificial Intelligence 1n Medicine; Westview Press, Boulder, Colorado, 1982.