CAUSAL EXPLANATION FROM MATHEMATICAL MODELS

MARIA LEE* PAUL COMPTON† BOB JANSEN*

** CSIRO Division of Information Technology*
P.O.Box 1599, North Ryde 2113, AUSTRALIA

† Artificial Intelligence Lab, School of Computer Science and Engineering,
University of New South Wales, AUSTRALIA

A paper presented in AI'92, 16-18 Nov 1992, Tasmania, Australia

ABSTRACT

Decision support systems may require the use of existing complex mathematical models. It is desirable to reduce the equations of such a model to an explanatory causal form to support decision making, for example in making decisions about the greenhouse effect. This paper presents an approach to constructing causal explanation from existing complex mathematical models. The method combines information from the user and the use of heuristics to reduce the equations to a suitable form. It is then able to provide a reasonable causal explanation about the behaviour of the system modelled. The method is generic and has been tested on a wide range of domains. This method has been implemented in Hypercard on a Macintosh to provide an accessible graphically oriented decision support tool.

1. Introduction

Experts are often requested to make judgements and provide advice in the context of moderation of future change. Human-induced greenhouse warming of the Earth is an example of dealing with complex issues in providing expert advice to policy makers. The development of methodologies for climate impact advice requires large amounts of knowledge from diverse sources. This paper concentrates on the acquisition of knowledge from equations in a mathematical model.

Even a simple mathematical model can have extraordinarily dynamic behaviour [11]. However, one of the limits of the mathematical models is that they provide no explicit knowledge of how to perform analysis or to interpret results [10]. When domain experts are asked to explain a result from numerical simulation, they often describe behaviour as a set of numerical values along some successive time points, but are often unable to give a causal explanation of why things behave as they do [1].

One approach is a domain specific expert system to interpret simulation results and present them in an accessible form [12]. Our intention is to use the existing mathematical model itself to assist the system in explaining and justifying its advice for decision making. To understand better the various climate-related processes is to improve our predictive capability and to facilitate decision making. Users are not trying to make new scientific interpretations, rather the intent is to permit decision makers to make decisions based on a good qualitative understanding of complex mathematical systems. The need to produce a meaningful causal explanation for justifying an hypothesis is often stated as a fundamental part of an intelligent decision support system [13].

We propose to use any information available from the user which can identify the causal role of parameters in the model. Secondly we propose to use heuristics to reduce equations to a suitable form to produce a reasonable causal explanation. Inputs and outputs are known in many systems. Users and experts can predetermine inputs and outputs of a system. For instance, an electrical engineer recognises as obvious the inputs and outputs when designing a regulator circuit although this is not explicit in the model. Knowledge of inputs and outputs can support the identification of dependencies among variables. We then use the dependency characteristic of variables to reconstruct equations to be asymmetric causal equations. An asymmetric equation is one where the variables on the right hand side (RHS) and left hand side (LHS) cannot be exchanged. In this form the LHS of an equation is the dependent variable. Causality is then explicitly represented in the asymmetric causal equations.

Researchers have worked on constructing causal explanation from equations, such as Iwasaki and Simon's *causal ordering* [5,6,7,8], and de Kleer and Brown's *mythical causality* [3,4]. However, none of them have applied and used input, output and other causal information, as well as heuristics to produce "normal" structural equations and "reasonable" causality. As well, the heuristics we propose appear to overcome the practical deficiencies of causal ordering in handling feedback [5,13].

2. Causality

Decision makers need support to be able to identify important criteria and the ability to examine "what-if" scenarios. For example, typical questions for a policy maker to ask concerning the Greenhouse effect would be:

> What would be the effect of reducing CO_2 in the atmosphere?
> What is responsible for an increase in cyclones?

A simulation will answer the first question, but the second question must be answered from multiple explanatory simulations on some sort of sensitivity analysis of the equations. Quantitative differences between causes are vital to a full decision support system, but at this stage we are simply trying to identify causality.

Causal explanations describe how the behaviours of individual components contribute to the overall behaviour of a system. This knowledge is important for understanding, designing and diagnosis. For example, in an electronic circuit design, a complete algebraic analysis may not be available, but knowledge of how the individual components support the overall behaviour can certainly improve the efficiency analysis [3,9].

What is the concept of causality? According to Bunge [2], causation can be denoted as a connection between two events - *If X is a cause of Y, then a change in X causes a change in Y.* Such a change can either be a discrete change or it can be a gradual change over an interval.

One of the advantages of constructing causal explanations from mathematical equations is that it is a more abstract description of the domain model. Such an abstraction can capture useful features of a domain and allow for a better understanding of the internal behaviour. Also, precise quantitative information is not required in a causal model. The only concern is the relationships between variables.

3. Constructing Causal Explanation From Equations

Generally one describes the causal connections between variables by defining variables as independent and dependent variables. A dependent variable is a variable that is directly caused or influenced by any other variables. An independent variable is not influenced by other variables and is constant in any specific simulation run. Dependent variables represent "effects" and independent variables represent "causes". However, when a system has feedback involve, causal analyses becomes very complicated. Some variable can both be causes and effects. Note the distinction between a potential cause - an independent variable - and a numerical constant used in a functional relationship in the model.

3.1 Identification of dependency
The identification of the dependency characteristics of variables is provided by users or from heuristic assumptions. Mathematical models offer a complementary relation among equations and describe the structure and function of system [10]. Often the causal dependency of a variable is defined within equations. For instance, in an equation X = c1, where X is a variable and c1 is a constant. We define X as an exogenous or independent variable, which is determined by variables that are outside the system [5,6,14].

Sometimes it is impossible to identify inputs, especially in feedback systems, so that it is necessary to apply heuristics to determine the causality which will be identified.

3.2 Definition of asymmetric causal equations
An asymmetric causal equation is an equation with independent and dependent variables represented in an ordered form. Implicit in the asymmetry is the notion that the variable on the LHS is dependant on the variables on the RHS, hence, the asymmetry of only one variable on the LHS of each equation. The definition of an asymmetric causal equation is:

(a) If an **output** variable appears on the LHS of the equation, it is the variable whose behaviour is of interest.
(b) An **input** or **independent** variable appears on the RHS of the equation.
(c) A **dependent** variable can be either side of an equation, with each dependent variable occurring once and only once on the LHS in the set of equations representing the model.
(d) The differential equation is a canonical form [8]. The **derivative** is on the LHS of the equation, with only one derivative in each equation.

3.3 Dependency heuristic
There may be a number of asymmetric causal forms for a given set of equations. If so an equation with greatest number of least occurring variables is chosen. The equation is manipulated so that the least occurrence variables are on the RHS. The remaining of the equations are then manipulated to give an asymmetric causal form. The first equation is now considered fixed and the process repeated for the rest of the equations etc.

3.4 A bathtub example
To further explain asymmetric causal equations, consider a bathtub example in [8]. Figure 1 shows a bathtub.

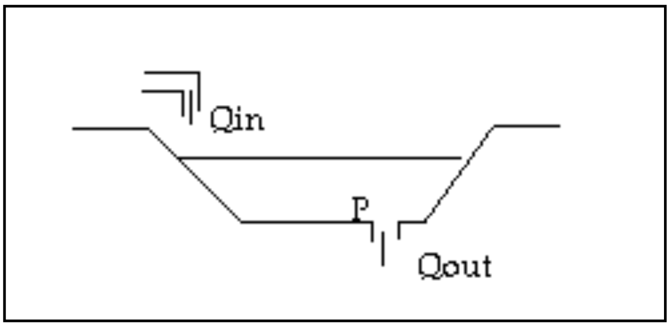

Figure 1. Bathtub (source from [8]).

The following equations describe the bathtub in equilibrium.

$$Q_{out} = K * P \qquad (1)$$
(the output flow rate is proportional to the pressure)
$$A = c1 * P \qquad (2)$$
(the pressure is proportional to the amount of water)
$$Q_{out} = Q_{in} \qquad (3)$$
(when the system is in equilibrium, the input flow rate equals to the output flow rate)

From the description of the bathtub in equilibrium, the user states that the amount of water, *A*, is the output of the system. The dependency heuristic determines the lowest influence variables, these are the input flow rate, *Qin*, and valve opening, *K*, - the independent variables of the system. By excluding the output, input or independent variables, the intermediate dependent variable then is the pressure, P and the output flow rate, Qout.

We then manipulate variables within the equations to produce asymmetric causal equations of the bathtub model. The asymmetric causal equations now become :

$$P = Qout / K \qquad (1)$$
$$A = c1 * P \qquad (2)$$
$$Qout = Qin \qquad (3)$$

There are now three dependent variables in three equations which describe the dynamic behaviour of the system. Once we get this asymmetric causal equation form of the model, we can then determine causal links for these equations. The procedure operates by removing the constant variables first. Then causality in this model can be simply stated as:

P <= Qout, K
A <= P
Qout <= Qin

The arrows indicate that the LHS are variables that depend on the RHS variables. We can represent the causal links in a graphical representation. Figure 2 shows a causal graph of the bathtub in equilibrium. The causal explanation of the bathtub mathematical model is that the output flow rate depends on the input flow rate; the pressure depends on the output flow rate and the amount of water depends on the pressure.

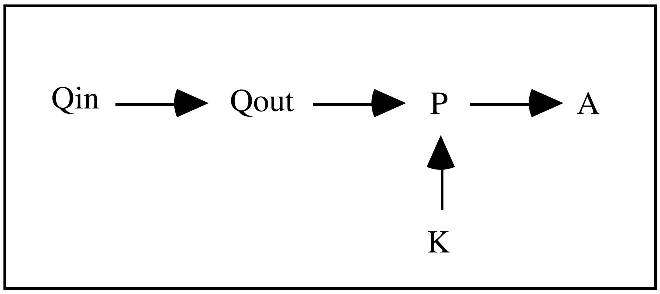

Figure 2. The causal graph of Bathtub in equilibrium structure.

It should be noted that the decision support tool (an intrinsic part of the application) also provides a model revision and consistency checking facility. The tool allows the expert to edit the above causal graphs, then the system backtracks from the edited graph back to a new set of asymmetric causal equations. These can be checked for consistency between equations and the user identified causality. If the dependency of variables is indeterminate, multiple causal models are produced.

4. Aggregation

When we study a mathematical model that involves many variables and relations, it becomes very complex to understand. We can aggregate the causal links as required to a number of small subgroups and show the relatively important parts. We have applied the method to a Global Energy Model with 41 equations and 73 parameters, and generated 107 causal links. According to the expert's requirements, we then aggregated these causal links into 8 subgroups to show the relatively more important causal corrections.

5. Conclusion

We have described a method for determining causal explanation of the behaviour of mathematical models. This method is based on applying heuristics to analyse the equations as well as using information from the user where possible.

We have tested the method on more than 20 mathematical models including all of those in the relevant literature and in all cases the method discovers the "correct" causality. However, equations in mathematical models do not model any original causality. We need to further investigate what class of equations (if any) have implicit causality and what classes of equations do not. However, our heuristic does manage to recapture the original causality where relevant, or at least a reasonable causality where the equation was not based on an initial causal model.

The next stage of this research is to attempt to understand what classes of models this heuristic works for and all in what way causality is implicit in these models. Also, it is essential to determine for what models the system gives causality that conflicts with the original causality the expert had in mind when devising the models.

The method not only generates causal pathways for explanation but also permits an expert to edit the causal paths to grasp new insights into a domain. The proposed method also produces a maximum causal resolution and automatically determines exogeneity, which makes knowledge acquisition easier. The construction of an asymmetric equation can constrain and derive the causal relationships among variables. Thus, causal dependencies are determined by other causal assignments within equations.

Also, we need to incorporate the concept of time in differential equations and assign the cause-effect signs (+/-) between causal links.

Acknowledgments

We would like to thank Ashesh Mahidadia, Tim Menzies, Phil Preston and Claude Sammut at the University of New South Wales for their crucial criticisms; Paul Guignard at the CSIRO Division of Information Technology for discussing of the physical phenomena; Ann Henderson-Sellers and Xiaohua Yuan at Macquarie University for giving climate impact advice.

References

1. A. Barr, P. Cohen and Feigenbaum (ed) *The Handbook of Artificial Intelligence Volume IV*, Addison-Wesley Publishing Company, 1989.

2. M. Bunge, *Causality and Modern Science*, New York, 1979.

3. J. de Kleer and J. Brown, *A Qualitative Physics Based on Confluences,* Artificial Intelligence 24, 1984, pp7-83.

4. J. de Kleer and J. Brown, *Theories of Causal Ordering*, Artificial Intelligence 29, 1986, pp33-61.

5. Y. Iwasaki and H. Simon, *Causality in Device Behaviour*, Artificial Intelligence 29, 1986, pp1-33.

6. Y. Iwasaki and H. Simon, *Theories of Causal Ordering: Reply to de Kleer and Brown*, Artificial Intelligence 29, 1986, pp 73-117.

7. Y. Iwasaki, *Model Based Reasoning of Device Behaviour with Causal Ordering*, a PhD thesis from Carnegie Mellon University, 1988.

8. Y. Iwasaki, *Causal Ordering in a Mixed Structure* , in the Proceeding of AAAI-88, 1988, pp 313-318.

9. B. Kuipers, *Commensence Reasoning about Causality: Deriving Behaviour from Structure,* Artificial Intelligence 24, 1985, pp169-203.

10. J. Kunz, M. Stelzner and M. Williams, *From Classic Expert Systems to Models: Introduction to a Methodology for Building Model-Based Systems*, in Topics in Expert System Design, Elsevier Science Publishers, 1989.

11. R. May, *Simple Mathematical Models with very Complicated Dynamics* in Nature Vol 261, 1976.

12. T. Menzies, J. Black, M. Dean, and J. Fleming, *Combining Heuristics and Simulation Models: An Expert System for the Optimal Management of Pigs*, in Lecture Notes in Artificial Intelligence vol 406, Springer-Verlag, 1990.

13. R. Patil, *Causal Representation of Patient Illness for Electrolyte and Acidbase Diagnosis*, PhD Thesis, Massachusetts Institute of Technology, 1981.

14. J. Top and H. Akkermans, *Computational and Physical Causality*, IJCAI-91.