In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, pp. 190-193, August 8-12, 1983.

TECHNICAL REPORT
R-25
May 1983

A COMPUTATIONAL MODEL FOR CAUSAL AND DIAGNOSTIC REASONING IN INFERENCE SYSTEMS[1]

Jin H. Kim[2] and Judea Pearl

Cognitive Systems Laboratory
University of California, Los Angeles

ABSTRACT

This paper introduces a representation of evidential relationships which permits updating of belief in two simultaneous modes: causal (i.e. top-down) and diagnostic (i.e. bottom-up). It extends the hierarchical tree representation by allowing multiple causes to a given manifestation. We develop an updating scheme that obeys the axioms of probability, is computationally efficient, and is compatible with experts reasoning. The belief parameters of each variable are defined and updated by those of its neighbors in such a way that the impact of each new evidence propagates and settles through the network in a single pass.

I INTRODUCTION

The integration of new pieces of information to existing body of knowledge constitutes a fundamental problem in a class of decision-making tasks such as situation assessment, diagnosis, pattern recognition and speech understanding. Knowledge-based expert systems and decision support systems must handle this problem to achieve expert's level performance and to derive valid recommendations. This paper addresses the issues of efficiently propagating the impact of new evidence and beliefs through a hierarchically organized inference network. The inference procedure described here simultaneously models both causal and diagnostic modes of reasoning. The causal mode of reasoning refers to the inference process of updating the likelihood of an event due to modified belief in its causal factors while the diagnostic mode of reasoning refers to that of updating the likelihood of an event as a result of an update in some of its manifestations (Tversky and Kahneman 1979).

The inference procedure described here is a generalization of the Bayesian methods previously applied to trees (DDI 1973, Pearl 1983) toward a class of hierarchical networks suitable to model multiple causes. The tree representation insists that only one variable be considered a cause of any other variable. This restriction simplifies computations and avoids the problem of maintaining consistency among interrelated variables. However, its representational power is so restricted that many real problems cannot be modeled naturally. In order to comply with the requirements imposed by the tree structure, we must group together all the causal factors as the set of states of one single variable. By contrast, when people associate a given observation with multiple potential causes, they weigh one causal factor against another as independent variables, each pointing to a specialized area of knowledge. As an illustration, consider the following situation:

> Mr. Holmes received a telephone call from his neighbor notifying him that she heard a burglar alarm sound from the direction of his home. As he was preparing to rush home, Mr. Holmes recalled that last time the alarm had been triggered by an earthquake. On his way driving home, he heard a radio newscast reporting an earthquake 200 miles away.

Mr. Holmes perceives two episodes which may be potential causes for the alarm sound, an attempted burglary and an earthquake. Even though these two events are a priori independent and so, not mutually exclusive, still the radio anouncement reduces the likelihood of a burglary, as it "explains away" the alarm sound. Moreover, the two causal events are perceived as individual variables each pointing to a separate frame of knowledge. The computational scheme described here uses Bayes calculus to model that kind of interaction among causes in addition to the ususal interaction among diagnostic indicators.

This paper is organized as follows. After presenting the basic concepts and definitions, we introduce two kinds of independencies which typically characterize the interactions among the various causes of a common manifestation and among the various manifestations of a common cause. Exploiting these independencies, belief parameters are identified and an efficient belief propagation scheme is developed which updates the beliefs of all variables in a single pass through the network, avoiding infinite relaxations.

II HIERARCHICAL CAUSAL NETWORK

The basic definitions and concepts used here are

---

1. Supported in part by the National Science Foundation grant IST 8119045.

2. Currently with Hughes Research Laboratories, Malibu, California.

borrowed from Pearl (Pearl 1983). A node in a causal netwrok represents a variable. Let a variable be labeled by a capital letter, e.g., A, B, ..., X, Y, and its various states subscripted by numbers, e.g., X_1, Y_2. A causal network is a directed graph where each link X --> Y represents the relationship 'X causes Y', and is quantified by a conditional probability matrix $\underline{M}(Y|X)$ with entries:

$$(1) \quad \underline{M}(Y|X)_{i,j} = \text{Prob}(Y_i|X_j).$$

We restrict the arrows to follow the direction of causality insisting that variables be only related by conditional probabilities where the cause, not the effect, is the conditioned variable. The reason is that usually the probability P(manifestation|cause) is psychologically more available (Tversky and Kahneman, 1979), and therefore, can be elicited with greater ease and validity than its counterpart, P(cause|manifestation) (Burns and Pearl, 1981).

We will restrict our attention to a special kind of graph, called Generalized Chow Tree(GCT) where a node may have several parents but at most one underlying path exists between any pair of nodes[3]. Since no cycle exist, a link B --> A partitions the graph into two parts: an upper subgraph, G^+_{BA}, and a lower subgraph, G^-_{BA}. These two graphs constitute hierarchical representations for the set of data which we shall call D^+_{BA} and D^-_{BA}, respectively. These data are defined as the observations and prior beliefs obtained only at the boundaries of network. Likewise, every node A partitions the graph into two parts: above A, G^+_A, and below A, G^-_A, representing the data set D^+_A and D^-_A respectively. Figure 1 shows the causal network representing Mr. Holmes' belief structure.

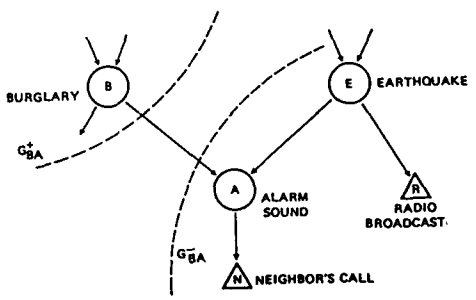

Figure 1 : Mr. Holmes' Belief Structure

III STRUCTURAL ASSUMPTIONS OF INDEPENDENCE

The likelihood of the various states of a variable X would, in general, depend on the entire data observed so far. However, the existence of only one path from G^+_{YX} to X implies that the data, D^+_{YX}, influences X only through the states of Y:

$$(2) \quad P(X_i|Y_j, D^+_{YX}) = P(X_i|Y_j)$$

which leads to

$$(3) \quad P(X_i|D^+_{YX}) = \sum_j P(X_i|Y_j)\, P(Y_j|D^+_{YX}).$$

In other words, the influence of one node on another is compleltely summarized by the intermediate nodes between them.

A special case of this assumption is what we traditionally call "Conditional Independence," which is usually valid among several manifestations of a common cause. If X and Z are successors of Y, we then write

$$(4) \quad P(X_i,Z_j|Y_k) = P(X_i|Y_k)\, P(Z_j|Y_k)$$

which means that X and Z are not independent a priori, but become independent once we know with certainty which state of Y prevails. We will call this relationship inter-symptom independence. (See Figure 2-a.)

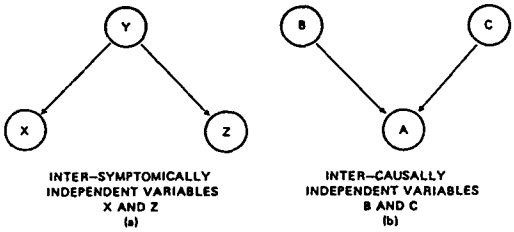

Figure 2: Independence Relationships

The inter-causes relation is typically perceived to work in the opposite direction, i.e., causes are viewed to be a priori independent and once their common symptom is observed they become coupled. In Mr. Holmes' example, home burglaries can safely be assumed independent of earthquakes. However, given the alarm sound, the likelihood of a burglary becomes dependent upon the occurrence of an earthquake. We call this relationship inter-causes independence (see Figure 2-b), and formulate it via

$$(5) \quad P(B_i,C_j|D^+_B,D^+_C) = P(B_i|D^+_B)\, P(C_j|D^+_C).$$

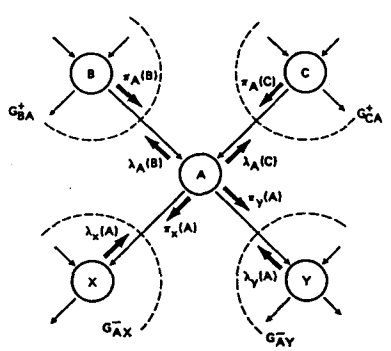

Figure 3: A Fragment of a Causal Network

---

3. A general causal network with cycles may be transformed into a GCT by a systematic treatment such as neglecting the least informative link (Chow 1968, Kim 1983).

IV BELIEF PARAMETERS

Consider the network of Figure 3. The strength of belief $BEL(A_i)$ of A_i should, at any given time, reflect the entire data observed so far, i.e., data from subgraphs G^+_{BA}, G^+_{CA}, G^-_{AX} and G^-_{AY}. Hence, we write

$$(6) \quad BEL(A_i) = P(A_i|D^+_{BA},D^+_{CA},D^-_{AX},D^-_{AY}).$$

According to Bayes rule and the structural assumption (2),

$$(7) \quad BEL(A_i) = \alpha\, P(A_i|D^+_{BA},D^+_{CA})\, P(D^-_{AX},D^-_{AY}|A_i)$$

where α is a normalization constant[4]. Further applying Eq (3) and (5), we get

$$(8) \quad BEL(A_i) = \alpha[\sum_{jk} P(A_i|B_jC_k)\, P(B_j|D^+_{BA})\, P(C_k|D^+_{CA})]\, P(D^-_{AX}|A_i)P(D^-_{AY}|A_i).$$

Eq (8) suggests that the probability distribution of each variable A in the network could be computed if three parameters are made available: 1) the current strength of the causal evidence, π, contributed by each incoming link to A;

$$(9) \quad \pi_A(B_j) = P(B_j|D^+_{BA})$$

2) the current strength of the diagnostic evidence, λ, contributed by each outgoing link from A;

$$(10) \quad \lambda_X(A_i) = P(D^-_{AX}|A_i)$$

and 3) the fixed conditional probability matrix, P(A|B,C), which relates the variable A to its immediate causes. Accordingly, in the propagation scheme which we have devised, we let each link carry two dynamic parameters, π and λ, and let each node store the information contained in P(A|B,C).

V APPROXIMATION OF P(A|B,C)

In principle, the specification of P(A|B,C) requires a table with one entry for each state combination of the variables A, B and C. Needless to say, such a table is rather troublesome to obtain from experts due to its size. For this reason, it is necessary to approximate high-order conditional probabilities P(A|B,C) from pairwise relations P(A|B) and P(A|C).

A description of a state at a given level of detail is an aggregation of states of the next more detailed level (Patil 1981). The state of an aggregated variable is determined by a relationship among its component states. Consider the Mr. Holmes example again. The state 'alarm' is a summarization of its more detailed level states, 'alarm caused by burglar' and 'alarm caused by earthquake'. Moreover, either a burglar or an earthquake may cause the alarm sound separately, while the state 'alarm sound' is false when both 'alarm sound cause by a burglar' and 'alarm sound caused by an earthquake' are false. We say that the state 'alarm sound' dominates its complement state. The dominance relationship is a characteristic of a variable itself, not of the causal relations with its neighbors.

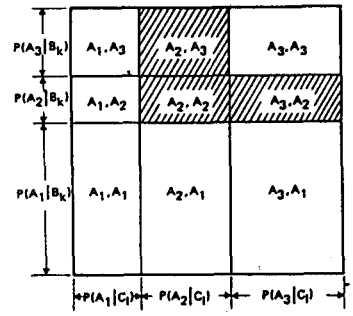

Figure 4: Computation of P(A|BC)

The strength of belief of an aggregated state is computed by the sum of beliefs committed to its component states. This computation is illustrated in Figure 4, in which beliefs supported by two causal states B_k and C_l are combined. The vertical axis represents the belief distribution of A supported by B_k and the horizontal axis represents that of A supported by C_l. If we assume that A_i dominates A_j for $i < j$, then the combining formula is

$$(11) \quad P(A_i|B_k,C_l) = \alpha \sum_{p>i,q>p} \sum P(A_p|B_k)P(A_q|C_l)$$

where α is a normalization constant. Eq (11) means that the regions of conflicting labels are resolved by the dominance relation.

The dominance relation may not hold for some variables. For those, the regions of conflicting labels are ignored and the ratio of the diagonal regions serves to produce belief distribution. For this case, the combining formula is

$$(12) \quad P(A_i|B_k,C_l) = \alpha\, P(A_i|B_k)P(A_i|C_l).$$

Note that this formula resembles the Dempster's rule of combination known as "orthogonal sum" devised for the treatment of ignorance (Shafer 1976).

VI PROPAGATION OF INFORMATION

Assuming that the vectors $\underline{\pi}$ and $\underline{\lambda}$ are stored with each link, our task is now to prescribe how the influence of new information spreads through the network.

A. Updating λ

---

4. We assume that α is chosen to make $\sum BEL(A_i) = 1$. However, one may relax this constraint to represent the degree of ignorance such as Dempster-Shafer system (Shafer 1976).

Assume that B and C form a super variable which admits all combinations of the states of B and C, then

$$(13)\quad BEL(B_i) = \sum_j BEL(B_i C_j)$$

and at the same time,

$$(14)\quad BEL(B_i) = \alpha\, \pi_A(B_i)\lambda_A(B_i).$$

Equating Eq (13) and (14), we get

$$(15)\quad \lambda_A(B_i) = \alpha\sum_j[\pi_A(C_j)\sum_k \lambda_X(A_k)\lambda_Y(A_k)P(A_k|B_iC_j)].$$

Eq (15) shows that only three parameters (in addition to the conditional probabilities $P(A|B,C)$) need to be involved in updating the diagnostic parameter vector $\lambda_A(B)$ from A to B: $\pi_A(C)$, $\lambda_X(A)$ and $\lambda_Y(A)$. This is expected since $\lambda_A(B)$ stands for $P(B|\bar{D}_{BA})$ and $\bar{D}_{BA}$ is completely summarized by the above three parameters. (See Figure 3.)

B. Updating π

The rule for updating the causal parameter $\pi_X(A)$ can be obtained from formula:

$$(16)\quad \pi_X(A_i) = \alpha\lambda_Y(A_i)[\sum_{jk} P(A_i|B_jC_k)\ \pi_A(B_j)\pi_A(C_k)]$$

Thus, similar to $\lambda_A(B)$, $\pi_X(A)$ is also determined by three neighboring parameters: $\lambda_Y(A)$, $\pi_A(B)$ and $\pi_A(C)$.

Equation (15) and (16) also demonstrate that a perturbation of the causal parameter, π, will not effect the diagnostic parameter, λ, on the same link, and vice versa. Therefore, any perturbation of beliefs due to new evidence propagates through the network and is absorbed at the boundary without reflection. A new equilibrium state will be reached after a finite number of updates which, in the worst case, is equal to the diameter of the network.

Eq (15) reveals that if no data is observed below A, i.e., all λ's to A are an unit vector, then all λ's from A are also an unit vector. This means that evidence gathered at a node does not influence its spouses until their common son gathers diagnostic evidence. In Mr. Holmes' case, for example, seismic data pertaining to earthquakes would not have influenced the likelihood of burglary prior to receiving the neighbor's telephone call. It is a pleasing characteristic. Otherwise, a node may gather support through purely mental constucts void of diagnostic support.

A node which has no predecessor needs a special parameter unless it is a data node. Since no causal influence is available from its predecessors, it requires an external parameter summarizing the background, a priori[5] knowledge pertaining to that node, thus serving the classical role of subjective prior probablility.

Generalization of Eq (15) and (16) for more than two causal factors and more than two sets of manifestations is straight forward (Kim 1983).

VII CONCLUSIONS

We have introduced a formalization for the interaction among multiple causes which reflects the way people often view causal relationships. Based on this formulation, we have extended the tree representation to a class of hierarchical networks capable of modeling multiple causes while still maintaining the computational efficiency provided by the tree representation: belief parameters are updated by local (nearest neighbors) computations, they reach equillibrium after a single pass through the network and remain consistent with the tenets of probability calculus. Additionally, the causal network representation lends itself naturally to object-oriented formulation; each node is an object of the same generic type and the belief parameters are the messages by which neighboring objects communicate.

REFERENCES

1. Tversky, A. and Kahneman, D. "Causal Schemata in judgments Under Uncertainty," in Progress in Social Psychology, M. Fishbein(Ed.), Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

2. Decision and Design Inc. "Handbook of Decision Analysis," McLean, Virginia, 1973.

3. Pearl, J. "Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach," Proceedings of the second annual conference on Artificial Intelligence, Pittsburgh, Pennsylvania, August 1982.

4. Burns, M. and Pearl, J. "Causal and Diagnostic Inferences: A Comparison of Validity," Organizational Behavior and Human Performance 28, 1981.

5. Chow, C. and Liu, C. "Approximating Discrete Probability Distributions with Dependence Trees," IEEE Transactions on Information Theory, May 1968.

6. Patil, R., Szolovits, P. and Schwartz W. "Causal Understanding of Patient Illness in Medical Diagnosis," Proc. of IJCAI-81, Vancouver, Canada, 1981.

7. Shafer, G. A Mathematical Theory of Evidence, Princeton University Press, Princeton, New Jersey 1976.

8. Kim, J. "CONVINCE: A CONversational INference Consolidation Engine," Ph.D. Dissertation, University of California, Los Angeles, 1983.

---

5. According to Webster's New World Dictionary, "a priori" means "from cause to effect".