BAYESIAN NETWORKS*

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

Bayesian networks were developed in the late 1970's to model distributed processing in reading comprehension, where both semantical expectations and perceptual evidence must be combined to form a coherent interpretation. The ability to coordinate bi-directional inferences filled a void in expert systems technology of the early 1980's, and Bayesian networks have emerged as a general representation scheme for uncertain knowledge [Pearl, 1988, Heckerman *et al.*, 1995, Jensen, 1996, Castillo *et al.*, 1997].

Bayesian networks are directed acyclic graphs (DAGs) in which the nodes represent variables of interest (e.g., the temperature of a device, the gender of a patient, a feature of an object, the occurrence of an event) and the links represent informational or causal dependencies among the variables. The strength of a dependency is represented by conditional probabilities that are attached to each cluster of parents-child nodes in the network.

Figure 1 illustrates a simple yet typical Bayesian network. It describes the causal relationships among the season of the year (X_1), whether rain falls (X_2) during the season, whether the sprinkler is on (X_3) during that season, whether the pavement would get wet (X_4), and whether the pavement would be slippery (X_5). Here, the absence of a direct link between X_1 and X_5, for example, captures our understanding that the influence of seasonal variations on the slipperiness of the pavement is mediated by other conditions (e.g., wetness).

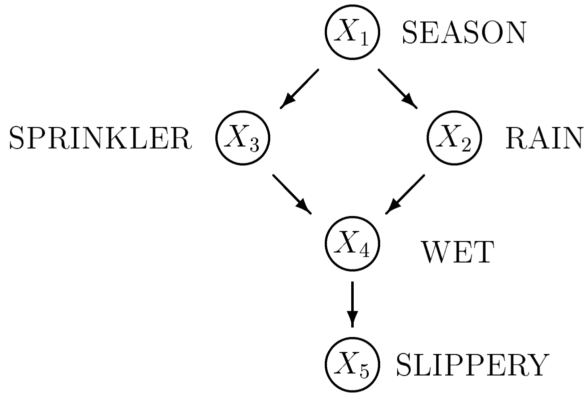

Figure 1: A Bayesian network representing causal influences among five variables.

*To appear in *MIT Encyclopedia of the Cognitive Sciences*

As this example illustrates, a Bayesian network constitutes a model of the environment rather than, as in many other knowledge representation schemes (e.g., rule-based systems and neural networks), a model of the reasoning process. It simulates, in fact, the mechanisms that operate in the environment, and thus facilitates diverse modes of reasoning, including prediction, abduction and control.

Evidential reasoning. Prediction and abduction require an economical representation of a joint distribution over the variables involved. Bayesian networks achieve such economy by specifying, for each variable, X_i, the conditional probabilities $P(x_i|pa_i)$ where pa_i are a set of predecessors (of X_i) which render X_i independent of of all its other predecessors. Variables judged to be the direct causes of X_i satisfy this property, and these are depicted as the parents of X_i in the graph. Given this specification, the joint distribution is given by the product

$$P(x_1, ..., x_n) = \prod_i P(x_i \mid pa_i) \tag{1}$$

from which all probabilistic queries (e.g., find the most likely explanation for the evidence) can be answered coherently using probability calculus.

The first algorithms proposed for probabilistic calculations in Bayesian networks used message-passing architecture and were limited to trees [Pearl, 1982, Kim and Pearl, 1983]. Each variable was assigned a simple processor, permitted to pass messages asynchronously with its neighbors, until equilibrium is achieved. Techniques have since been developed to extend this tree-propagation method to general networks. Among the most popular are Lauritzen and Spiegelhalter's method of join-tree propagation, and the method of loop-cut conditioning (see [Pearl, 1988, Jensen, 1996]).

While inference in general networks is NP-hard, the complexity for each of the methods cited above can be estimated prior to actual processing. When the estimates exceed reasonable bounds, an approximation method such as stochastic simulation [Pearl, 1988] can be used instead. Learning techniques have also been developed for systematic updating of the conditional probabilities $P(x_i|pa_i)$, as well as the structure of the network, so as to match empirical data (see Spiegelhalter and Lauritzen in [Shachter, 1990] and [Cooper and Herskovits, 1990]).

Reasoning about actions. The most distinctive feature of Bayesian networks, stemming largely from their causal organization, is their ability to represent and respond to changing configurations. Any local reconfiguration of the mechanisms in the environment can be translated, with only minor modification, into an isomorphic reconfiguration of the network topology. For example, to represent a disabled sprinkler, we simply delete from the network all links incident to the node "Sprinkler". To represent the policy of turning the sprinkler off if it rains, we simply add a link between "Rain" and "Sprinkler" and revise $P(x_3|x_1, x_2)$. This flexibility is often cited as the ingredient that marks the division between deliberative and reactive agents, and that enables the former to manage novel situations instantaneously, without requiring retraining or adaptation.

Organizing one's knowledge around stable mechanism provides a basis for planning under uncertainty [Pearl, 1996]. Once we know the identity of the mechanism altered by the intervention and the nature of the alteration, the overall effect of an intervention can be predicted by modifying the corresponding factors in Eq. (1) and using the modified product to compute a new probability function. For example, to represent the action "turning the sprinkler ON" in the network of Figure 1, we delete the link $X_1 \rightarrow X_3$ and fix the value of

X_3 to ON. The resulting joint distribution on the remaining variables will be

$$P(x_1, x_2, x_4, x_5) = P(x_1)\ P(x_2|x_1)\ P(x_4|x_2, X_3 = \text{ON})\ P(x_5|x_4) \tag{2}$$

Note the difference between the action $do(X_3 = \text{ON})$ and the observation $X_3 = \text{ON}$. The latter is encoded by ordinary Bayesian conditioning, while the former by conditioning a mutilated graph, with the link $X_1 \rightarrow X_3$ removed. This mirrors indeed the difference between seeing and doing: after observing that the sprinkler is ON, we wish to infer that the season is dry, that it probably did not rain, and so on; no such inferences should be drawn in evaluating the effects the contemplated action "turning the sprinkler ON".

Causal Discovery. One of the most exciting prospects in recent years has been the possibility of using Bayesian networks to discover causal structures in raw statistical data. [Pearl and Verma, 1991, Spirtes *et al.*, 1993]. Although any inference from association to causation is bound to be less reliable than inference based on controlled experiment, one can still guarantee an aspect of reliability called "stability": any alternative structure compatible with the data must be less stable than the one(s) inferred; namely, slight fluctuations in conditions will render that structure no longer compatible with the data. With this form of guarantee, the theory provides criteria for identifying genuine and spurious causes, with or without temporal information, and yields algorithms for recovering causal structures with hidden variables from empirical data.

Plain Beliefs. In mundane decision making, beliefs are revised not by adjusting numerical probabilities but by tentatively accepting some sentences as "true for all practical purposes". Such sentences, called *plain beliefs*, exhibit both logical and probabilistic character. As in classical logic, they are propositional and deductively closed; as in probability, they are subject to retraction and to varying degrees of entrenchment. Bayesian networks can be adopted to model the dynamics of plain beliefs by replacing ordinary probabilities with non-standard probabilities, that is, probabilities that are infinitesimally close to either zero or one [Goldszmidt and Pearl, 1996].

Models of cognition. Although Bayesian networks can model a wide spectrum of cognitive activity, their greatest strength is in causal reasoning, which, in turn, facilitates reasoning about actions, explanations, counterfactuals, and preferences. Such capabilities are not easily implemented in neural networks, whose strengths lie in quick adaptation of simple motor-visual functions.

Some questions arise: Does an architecture resembling that of Bayesian networks exist anywhere in the human brain? If not, how does the brain perform those cognitive functions in which Bayesian networks excel? One plausible answer is that fragmented structures of causal organizations are constantly being assembled on the fly, as needed, from a stock of functional building blocks. For example, the network of Figure 1 may be assembled from several neural networks, one specializing in the experience surrounding seasons and rains, another in the properties of wet pavements, and so forth. Such specialized networks are probably stored permanently in some mental library, from which they are drawn and assembled into the structure shown in Figure 1 only when a specific problem presents itself, for example, to determine whether an operating sprinkler could explain why a certain person slipped and broke a leg in the middle of a dry season.

Thus, Bayesian networks are particularly useful in studying higher cognitive functions, where the problem of organizing and supervising large assemblies of specialized neural networks becomes important.

References

[Castillo *et al.*, 1997] E. Castillo, J.M. Gutierrez, and A.S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer-Verlag, New York, 1997.

[Cooper and Herskovits, 1990] G.F. Cooper and E. Herskovits. A Bayesian method for constructing Bayesian belief networks from databases. *Proceedings of the Conference on Uncertainty in AI*, pages 86–94, 1990.

[Goldszmidt and Pearl, 1996] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1–2):57–112, July 1996.

[Heckerman *et al.*, 1995] Guest Editors: David Heckerman, Abe Mamdani, and Michael P. Wellman. Real-world applications of bayesian networks. *Communications of the ACM*, 38(3):24–68, March 1995.

[Jensen, 1996] Finn V. Jensen. *An Introduction to Bayesian Networks*. Springer, New York, 1996.

[Kim and Pearl, 1983] J.H. Kim and J. Pearl. A computational model for combined causal and diagnostic reasoning in inference systems. In *Proceedings IJCAI-83*, pages 190–193, Karlsruhe, Germany, 1983.

[Pearl and Verma, 1991] J. Pearl and T. Verma. A theory of inferred causation. In J.A. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452. Morgan Kaufmann, San Mateo, CA, 1991.

[Pearl, 1982] J. Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. In *Proceedings AAAI National Conference on AI*, pages 133–136, Pittsburgh, PA, 1982.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligence Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

[Pearl, 1996] J. Pearl. Causation, action, and counterfactuals. In Y. Shoham, editor, *Theoretical Aspects of Rationality and Knowledge, Proceedings of the Sixth Conference*, pages 51–73. Morgan Kaufmann, San Francisco, CA, 1996.

[Shachter, 1990] R.D. Shachter. Special issue on influence diagrams. *Networks: An International Journal*, 20(5), August 1990.

[Spirtes *et al.*, 1993] P. Spirtes, C. Glymour, and R. Schienes. *Causation, Prediction, and Search*. Springer-Verlag, New York, 1993.