PROBABILISTIC REASONING: Introduction to Special Issue IEEE PAMI

Guest Editors: H. R. Keshavan, J. A. Barnett,
D. Geiger, & T. Verma

March, 1993

Probability theory offers a well understood language for describing uncertainty and for drawing inferences from partial knowledge. Thus, its methods are applicable to any task requiring inferences from incomplete information. Typical applications include diagnosis, forecasting, and classification. However, computing with probability theory is not without its problems because the size of a joint distribution grows exponentially with the number of variables. Therefore, it is impractical to elicit, store, and compute with an entire distribution unless the domain is small.

One traditional approach to overcome complexity problems is to make independence and uniformity assumptions. Independence assumptions permit a joint distribution to be factored into a product of smaller marginal distributions, while uniformity assumptions permit identical distributions to be substituted for many of the marginals. These assumptions significantly reduce computational requirements and facilitate data acquisition. On the other hand, the resulting models may be over-simplified and can entail questionable conclusions.

Scientists working with probabilistic models are often forced to trade fidelity to save computational resources. In applications such as image processing there may be an enormous number of variables, e.g., pixel features. In such cases, statistical inference is not practical without making simplifying assumptions. A typical assumption is that the distribution of a pixel feature is determined by conditioning on just the features of neighboring pixels. This is tantamount to assuming that features of distant pixels are independent. While this model may lead to less than perfect conclusions, it does have the important property that some conclusions can be drawn within the limitations of the available computational resources. In applications such as medical diagnosis, the trade-off is different: the cost of questionable conclusions can be so severe that fidelity must dominate. A ramification of this priority is that, until recently, probabilistic methods were only applied to limited medical domains.

This special issue addresses the trade-off between fidelity and computational resources in areas where probabilistic reasoning is applicable. With the recent development of network-based approaches such as Bayesian networks, chain graphs, influence diagrams, recursive models, and Markov networks, wide ranges of choices between the two extremes have become available. These approaches are based on the ability to exactly model the dependencies required by the domain, while assuming independence, by default, whenever it is consistent with the resulting model. The ability to use ad hoc dependencies enables the creation of realistic models, while independence defaults reduce time and space complexities.

Bayesian Networks

Bayesian networks are the most popular of the current network-based approaches. They are parameterized by conditional probabilities and offer the ability to explicitly represent dependencies among random variables at arbitrary levels of detail. Further, there are many published algorithms to reason with and about them. The networks possess straightforward interpretations and this fact reduces the effort required to elicit the parameters as well as validate the encoded domain model.

The time and space complexities for the representation of a Bayesian network can range from a linear to an exponential function of network size and topology. The space complexity is dominated by the size of the network parameters, and for many classes of networks, is bound by a polynomial in the number of nodes. For example, if the in-degree of all nodes is bound by a constant, then the space bound is linear. In addition, if the networks are singly connected, then the time complexity is linear as well. However, general probabilistic inference is an NP-hard problem, even for the restricted class of networks where each node has at most two parents.

Formal Definition

A Bayesian network is a directed acyclic graph whose nodes represent random variables and whose edges represent probabilistic dependencies. Under a causal interpretation, the edges represent direct causal influences that are quantified by probability distributions associated with the nodes. For example, in Figure 1, A and B directly influence C. The influence of A and B

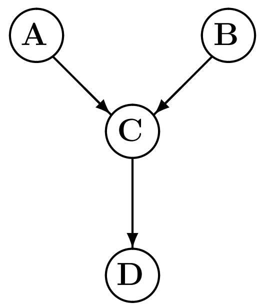

Figure 1: A Bayesian Network.

upon C is quantified by the conditional probability distribution, $P(C|AB)$, associated with C. Similarly, the influence of C upon D is quantified by the conditional probability distribution, $P(D|C)$, associated with D. The variables, A and B, are respectively quantified by the marginal distributions $P(A)$ and $P(B)$. Finally, the principle of assuming independence where possible is a type of closed world assumption used to delineate the semantics of the network. Thus, the joint distribution over all variables,

$$P(ABCD) \quad = \quad P(A)P(B)P(C|AB)P(D|C),$$

is assumed to be the product of the distributions associated with the nodes.

The edges of the Bayesian network in Figure 1 denote direct dependencies. Since A influences C and C influences D it follows that A indirectly influences D. However, there is no edge joining A and D, so A has no direct influence upon D. Thus, the influence of A upon D is completely mediated by C. This means, in terms of probability theory, that A and D are conditionally independent given C. This conditional independence is denoted by $I(A, D|C)$. Similarly, since no variable influences both A and B, they are marginally independent, written $I(A, B|\emptyset)$. The derivation of such I-statements is a manifestation of the default rule that assumes independence whenever it is consistent with the network.

While the interpretation of Bayesian networks was suggested by causal considerations, the formal semantics make no reference to causality. The nodes correspond to random variables and the topology of the graph represents a decomposition of the joint distribution into a product of several smaller conditional distributions. Thus, Bayesian networks are applicable to domains even when little or no causal information is available.

Papers in this Special Issue

Empirical problem-solving begins with an assessment of the type and quality of the domain knowledge available. The papers in this special issue describe methodology and theory that is useful when the available knowledge is probabilistic in nature. Four of the papers discuss methods to draw inferences when the available quantitative knowledge is insufficient to build a traditional Bayesian network. Three papers present meta-technology that automatically tailors Bayesian networks to particular problem instances. The remaining papers are concerned with the issues of representation, verification, and inference.

Bayesian networks require the specification of exact probability values. The possible sources of this data are physical models, statistical samples, and expert judgements. However, these sources are not always capable of providing sufficient quantitative information. The papers by Fertig & Breese, Goldszmidt, Morris & Pearl, Wellman & Henrion, and Bhatnagar & Kanal discuss methods to approximate probabilistic reasoning when the required distributions are not available. For example, Fertig & Breese develop inference techniques for Bayesian networks that are quantified by bounds rather than exact probability values.

Goldszmidt, Morris & Pearl discuss reasoning in a framework where the knowledge consists of qualitative default rules admitting exceptions. These rules are modeled as statements whose probabilities are infinitesimally removed from 1. Inferences are made using maximum entropy to minimize the number of extraneous dependency assumptions. While this paper maintains the spirit of assuming independence whenever consistent with a network of dependencies, it marks a departure from the standard Bayesian network formalism. The remaining papers discuss methods that are specifically related to Bayesian networks.

The parameters of a Bayesian network can encode many qualitative relationships that are not reflected in the graphical structure. For example, Wellman & Henrion identify monotonic influences and annotate links of the network with signs indicating the polarity of their influences. Bhatnagar & Kanal use directed hyper-edges as a sharper causal language to cluster the parents of a node into subsets that represent individual qualitative causal relationships. Both papers identify sources of hidden qualitative information and describe methods that enable inferences based soly upon this information. Thus, one contribution of these papers is that they provide methods for performing inference when only some qualitative and no quantitative information is available.

Generally, domain models encode the knowledge necessary to solve several related problems. When a model is large and direct computations with it are not feasible, techniques are required to specialize the model. Provan & Clarke accomplish this task using a data base of prior cases, while Goldman & Charniak use domain-specific meta-rules to construct an appropriate network for a given problem. Sarkar & Boyer tackle vision problems using network-templates that represent geometric relations. Each of these papers describes a method for using general domain knowledge to construct a Bayesian network that is tailored to a particular problem instance. The use of a specific network not only makes the inference process faster, it provides a basis for generating better explanations as well.

The remaining papers in this issue concern problems of representation, verification, and inference that occur when currently available algorithms are applied to the network. Olesen extends current practice with technology that permits both discrete and continuous variables to be used in the same Bayesian network. Cowell, Dawid & Spiegelhalter present a method to monitor inference processes within a Bayesian network to help a scientist assess the fidelity of the encoded model. The papers by Dagum & Chavez and Heckerman, Horvitz & Middleton both address methods that speed up the inference process. The former use stochastic simulation to approximate posterior probabilities to a given precision for a given certainty factor. The latter use a cost-benefit analysis to assess the utility of making observations and recomputing the posteriors.

Additional References

The following references summarize the foundations for the research reported in this special issue. They recast prior work in probabilistic reasoning in the framework of a network-based approach. Though they differ in exact notation, all focus on the issue of how to use knowledge of probabilistic independence to advantage.

R. A. Howard & J. E. Matheson, (1984). Influence Diagrams, in *Principles and applications of decision analysis*, pp. 720–762. Strategic Decisions Group: Menlo Park, California.

S. L. Lauritzen & D. J. Spiegelhalter, (1988). Local computations with probabilities on graphical structures and their applications to expert systems, in *J. Royal Statist. Soc.*, ser. B, vol. 50, No. 2, pp. 157–224.

R. E. Neapolitan, (1990). *Probabilistic Reasoning in Expert Systems*. John Wiley and Sons.

J. Pearl, (1988). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann: San Mateo, California.

Acknowledgement

We are gratified by the large number of papers submitted for this special issue and would like to thank the authors and referees who have invested the time and effort required to make this a successful endeavor. In addition, we would like to thank PAMI for the opportunity to introduce this topic to a readership that is larger than otherwise possible.