In *Working notes of the AAAI-90 Workshop on Explanation,*
pages 133-141, Boston, MA, July 1990

Using Scenarios to Explain Probabilistic Inference*

Marek J. Druzdzel
Carnegie Mellon University
Department of Engineering and Public Policy
Pittsburgh, PA 15213
md33+@andrew.cmu.edu

Max Henrion
Rockwell International Science Center
Palo Alto Laboratory, 444 High St, Palo Alto, CA 94301
and
Carnegie Mellon University
Department of Engineering and Public Policy
henrion@sumex–aim.stanford.edu

31 May 1990

Abstract

On theoretical grounds Bayesian probability theory is arguably the soundest approach to uncertain reasoning, but it has been criticized as being hard for humans to understand. The acceptability and effectiveness of decision support systems (DSSs) depends on their ability to explain and justify their conclusions to users. We describe a method of explaining probabilistic inference by considering the relative plausibility of alternative scenarios. Scenarios consist of sequences of events, often appearing as coherent causal explanations of observed evidence. Selecting only the most relevant variables and the most probable scenarios allows control over the simplicity and precision of the explanations. Process tracing studies of human uncertain reasoning suggests that this scheme resembles the way humans normally reason.

*This research was supported by the National Science Foundation under grant IRI–8807061 to Carnegie Mellon University and by the Rockwell International Science Center.

Introduction

For a person to come to trust a computerized decision aid, decision support system (DSS) or expert system, requires that the system can communicate the reasoning underlying its conclusions. Ideally, a session with a DSS should resemble a conversation having as goal improving users' insight into a problem.

One possible approach to building a DSS is to try to imitate the reasoning of human experts. This underlies much expert systems work. Presumably, if the system's reasoning imitates human expert's reasoning, it should be relatively easy to produce explanations comprehensible to the user. Unfortunately, there is a danger that this approach will reproduce the deficiencies of human intuitive reasoning [5, 7]. There is a large experimental literature demonstrating that Bayesian reasoning is not a good model of human reasoning [8]. This literature also demonstrates that human reasoning under uncertainty is liable to systematic biases and inconsistencies. Indeed, this is part of the argument for using normatively–based decision aids to help improve on unaided human intuition. This poses a dilemma: Must we choose between the soundness and the explainability of inference methods? The goal of the research described here is to see whether, in fact, it is possible to produce comprehensible and insightful explanations of probabilistic reasoning.

Bayesian Belief Networks

Influence diagrams and Bayesian belief networks (BBN) provide convenient ways to represent a decision problem. These diagrams are directed graphs, in which nodes represent random or deterministic events, decisions, or values. Arcs represent influences, that is probabilistic or informational relations between nodes [12]. The presence or absence of arcs provide a qualitative representation of the conditional dependence or independence between variables. The strength of influence is quantified by a conditional probability distribution for each variable given its predecessors. Beliefs about nodes with no predecessors are characterized by prior probabilities. There is no inherent formal connection between the direction of an influence arc and causality,[1] but it is generally most convenient to encode influences in the causal direction, with effect conditional on cause.

An example of a BBN is given in Figure 1. This BBN describes knowledge about possible causes of sneezing of an individual visiting a strange house. In particular, *Sneezing* can be caused by a *Cold* or an *Allergy* to *Cats*. Both a *Cat* and a *Dog* may leave *Paw Marks*. A priori beliefs about the likely presence of *Cold*, *Cat*, and *Dog* are

[1] The direction of arcs can be reversed by means of the Bayes' theorem.

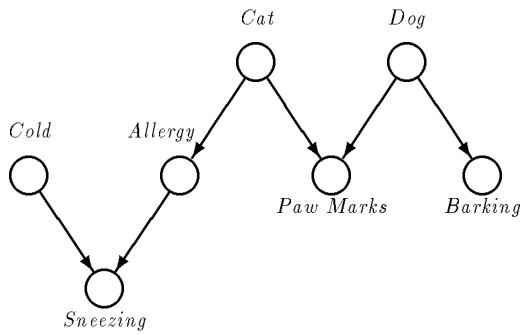

Figure 1: An example of a Bayesian belief network

described by their prior probabilities. The probability of any other event can easily be derived from these priors and conditional probabilities associated with the arcs.

Probabilistic Inference Within BBN

Belief Propagation

A BBN in itself is a clear representation for showing the qualitative structure of probabilistic dependence and independence of a model. One way to explain inference within a BBN, that we call “belief propagation”, is to trace the impact of evidence through the network from node to node. Causal inference follows the direction of the arrows, for example observing *cat* would increase expectation of *allergy*, and hence of *sneezing*. Diagnostic inference goes in the reverse direction, for example observing *sneezing* would increase belief in the existence of *allergy* and hence of a *cat*. Intercausal inference (or “explaining away”) is between alternative causes of a common effect. For example, given *paw marks* have been observed, observation of a *dog* provides an explanation of the *paw marks* and hence may reduces the evidence for a *cat*. Such qualitative inferences can be chained along any singly–connected network [6, 14]. Sember and Zukerman [11] propose an algorithm for generating explanations of quantitative probabilistic interactions between neighboring nodes in a BBN, similar to this scheme.

Human Judgment Under Uncertainty

If a computer system does not use similar reasoning schemes to humans, then it should at least be able to translate from its own representation to ones that are compatible to the way people think. As already mentioned, there is a large empirical literature showing that there are serious discrepancies between the way people estimate and process uncertainty and the prescriptions of the axioms of probability theory [8, 10]. These studies have found that people appear to use simple, intuitive strategies or *heuristics* in estimating probabilities and processing uncertain information. These make the tasks cognitively tractable, but often result in systematic biases and inconsistencies. While this literature shows that Bayesian inference is not a good descriptive model of human reasoning under uncertainty, it offers only weak suggestions of what better models might be. Despite common claims that alternative formal representations, such as fuzzy logic or certainty factors, are better models of human reasoning, there is little empirical work to substantiate this.

Existing experimental comparisons of human reasoning and Bayesian inference have concentrated mainly on simple situations, often with one hypothesis variable and one evidence variable. To improve our insight into people's plausible reasoning strategies in more complex tasks, we conducted a series of experiments using verbal "think aloud" protocols of subjects solving various problems under uncertainty [1]. One interesting finding was subjects' propensity to think in terms of deterministic scenarios. A scenario is a sequence of events, that is outcomes of all relevant variables, often forming a coherent, causal story. In one experiment subjects were given a version of the example presented above, told that sneezing and paw marks had been observed, and asked what the effect on the probability of a cold would be of hearing barking. Two plausible scenarios might be: (A) "There is a dog, which caused the barking and paw marks, but no cat; the sneezing was caused by a cold" and (B) "There is a dog, which caused the barking and paw marks; there is no cold; but there is a cat, which caused the allergy resulting in sneezing". Weighting of selected, most plausible scenarios that included a hypothesis, or its complement, formed the basis for judgment of likelihood of that hypothesis. Strategies resembling this scenario–based approach appeared more popular among the protocols than schemes analogous to belief propagation.

Scenario–Based Explanations

Scenario–Based Reasoning and Bayesian Inference

The scenario–based approach can easily be formalized into a normatively correct algorithm. The prior probability of a scenario is equal to the product of the conditional probabilities of each event conditional on its predecessor events (if any). The prior probability of a hypothesis is equal to the sum of the probabilities of all scenarios that are compatible with that hypothesis. Evidence, or observed nodes, will rule out some scenarios as impossible; for example observation of sneezing rules out all scenarios with no sneezing. The posterior probability of an event, e.g. cold, is the ratio of the sum of the probabilities of all scenarios with a cold and compatible with evidence, to the total probability of all scenarios compatible with the evidence. This renormalization of the probabilities after eliminating impossible scenario is just another way to apply Bayes' rule. It provides a way of doing diagnostic inference, reversing the direction of the influence arrow, by performing what seems like only forward or causal inference in assessing the probability of scenarios.

Generally, a model with n uncertain binary random variables generates 2^n possible scenarios. Of course, due to our cognitive limitations, we will not be able to consider more than a few scenarios, generally a tiny fraction of the number of possible scenarios in a complex problem. But fortunately, in many problems the bulk of the probability occurs in just a few scenarios, and so considering only the most probable ones is sufficient for a good approximation.

Generating Scenario–Based Explanations

We have employed a scenario–based scheme as a way to generate explanations of probabilistic inference in belief nets, as an alternative to belief propagation. This scheme has several advantages: Its basis in coherent, causal scenarios seems to be compatible with the way people normally reason; it avoids the diagnostic direction of reasoning that is generally more troublesome for people [3, 13]. It combines probabilistic and deterministic reasoning into a kind of possible–world logic. It also allows ways to tradeoff between the simplicity and accuracy of the explanation in terms of the number of scenarios used and the fraction of the total probability accounted for.

We have implemented this scheme in a user interface to a simple general purpose BBN–based DSS shell (Allegro Common LISP, MacIntosh). This scheme will be one among several explanation and communication methods comprising an interface to a quantitative probabilistic reasoner, under a common name QIQ (Qualitative Interface to the Quantitative). We intend to avoid committing ourselves to one single scheme, and treat scenario–based explanations as one of the possible ways of explaining probabilistic inference, leaving the choice to the user.

Generating an explanation in each case involves identifying those nodes (i.e. random variables) in the belief network that are directly or indirectly relevant to the hypotheses of interest. Often, many nodes can be rules out using Pearl's d–separation criterion [4]. For example, if the presence of paw marks is unknown, then the dog or barking are irrelevant to belief in the cat and other events, given the independence assumptions expressed by the diagram. If the cat is observed directly, then the presence of paw marks, dog or barking, are all made irrelevant. This can allow substantial simplification of the inference and explanation problems.

The remaining set of variables is then used for generating scenarios, which are sequences of outcomes of each of the random variables in the set. Two of the 2^7 scenarios in the complete example network are: *cold sneezing no–allergy no–cat paw–marks dog barking* and *no–cold sneezing allergy cat paw–marks dog barking.* For sufficiently small sets, exhaustive generation of scenarios is feasible. The probability of each scenario is computed by multiplying the probabilities of its component events, conditional on its predecessor events. These probabilities are easy to retrieve from the network.

Scenarios are then divided into two groups: those compatible and those incompatible with the hypothesis. A scenario is compatible with the hypothesis only if it includes the outcome in question. Explanations are based on selected scenarios from these two groups. It selects for use the most probable scenarios that collectively account for more than x% of the probability of the hypothesis. A typical value of this threshold is 90%, but it is adjustable according to the tradeoff between simplicity and precision desired. An example explanation generated by the system for the example BBN is given in Figure 2. It explains the probability of a cold given sneezing, paw marks, and barking have been observed.

```
? (why 'cold)
Given:
  Sneezing must have been caused by cold or allergy.
  Paw Marks could have been caused by cat or dog or
  another unknown cause.
  Barking must have been caused by dog.

Scenario(s) compatible with cold:
A. No cat, therefore no allergy, cold, and
   therefore sneezing.                            0.38
B. Cat, therefore allergy, cold, and
   therefore sneezing.                            0.05
   Other less probable scenario(s)                0.01
   Total probability of cold                      0.44

Scenario(s) incompatible with cold:
C. No cold, cat, and therefore allergy, and
   therefore sneezing.                            0.56

Therefore cold is almost as likely as not (p = 0.44).
```

Figure 2: Example of a scenario–based explanation

The explanation starts by listing as "Given" the observed evidence and its relevance. It then gives two lists of scenarios: those compatible and those incompatible with the hypothesis. Although the underlying system's reasoning is quantitative, it can provide numbers and/or verbal phrases to express probabilities according to the preferences of the user. Using numbers, it provides a summary of the posterior probabilities in a kind of ledger sheet comparing the contributions of the scenarios. The total probability of the less probable scenarios (under threshold) is also provided as a check on the completeness of the explanation.

These mappings or conversion tables between numerical and verbal probabilities have been derived from empirical studies in the literature [9]. The literature shows these mappings to be somewhat variable and context dependent. To allow for this, different mappings can be selected for the different networks, or different ones can be selected for each variable. Giving the number to supplement or even replace the verbal probability phrase may be more appealing to some users to avoid these contextual complexities.

Experimental Verification

We tested belief propagation–based and scenario–based explanations in a pilot experiment involving a simulated session with a medical DSS generating likelihood of breast cancer given patient's risk factors and results of screening tests [2]. Three groups of physicians: *control*, *scenario*, and *belief propagation*, analyzed three patient cases with system's assistance. The groups differed only in the explanations that were given along with the system's answers (no explanations, scenario–based, and belief propagation–based explanations respectively). We tested the degree of learning in the experiment (a measure of understanding the explanations) after the experiment in a test involving similar simple probabilistic problems. We measured the mean difference between the subjects' estimates and the normative answer. With four subjects in each group, we obtained a difference in performance between the treatment and the control groups ($p<0.10$), but no significant difference between the two treatment groups. This pilot experiment gives support for the hypothesis that our explanations improved subjects' insight into system's reasoning, but we did not find significant difference between explanation types.

Conclusion

In an attempt to convey probabilistic inference in a form more compatible with human reasoning, we developed explanations based on generation and subsequent weighting of deterministic scenarios involving the outcome of interest. This method lends itself to automatic generation of explanations by implementing it in an interface to a DSS based on Bayesian probability theory. It also provides several ways of controlling the tradeoffs between the simplicity and completeness of an explanation. Our studies involving judgment under uncertainty suggest that this method may be compatible with the way people usually think. Initial results have been encouraging, and we are now engaged in a number of extensions: To provide explanations in more complex networks situations, to incorporate consideration of utility as well as probability as a criterion (e.g. low probability scenarios resulting in death of a patient may be more important than higher probability scenarios with benign outcomes), to explore other ways to simplify explanations, and to conduct more conclusive empirical comparisons of their effectiveness.

References

[1] Marek J. Druzdzel. Towards process models of judgment under uncertainty. Technical Report CMU–EPP–1990–03–03, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, October 1989.

[2] Marek J. Druzdzel. Scenario based explanations for Bayesian decision support systems. Technical Report CMU–EPP–1990–03–04, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, January 1990.

[3] Hillel J. Einhorn and Robin M. Hogarth. Judging probable cause. *Psychological Bulletin*, 99(1):3–19, January 1986.

[4] Dan Geiger, Thomas Verma, and Judea Pearl. d–Separation: From theorems to algorithms. In *Proceedings of the 5th Workshop on Uncertainty in Artificial Intelligence*, pages 118–125, Windsor, Ontario, August 1989.

[5] Max Henrion. Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In Jeryl Mumpower, Ortwin Renn, Lawrence D. Phillips, and V.R.R. Uppuluri, editors, *Expert Judgment and Expert Systems, Proceedings of the NATO Advanced Research Workshop on Expert Judgment and Expert Systems*, pages 105–129, Porto, Portugal, August 1986. Berlin, Germany: Springer Verlag.

[6] Max Henrion and Marek J. Druzdzel. Qualitative propagation and scenario–based approaches to explanation of probabilistic reasoning. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pages 10–20, Cambridge, MA, July 1990.

[7] Eric J. Horvitz, John S. Breese, and Max Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2(3):247–302, July 1988.

[8] Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge, 1982.

[9] Sarah Lichtenstein and J. Robert Newman. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, 9(10):563–564, December 5 1967.

[10] M. Granger Morgan and Max Henrion. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, 1990.

[11] Peter Sember and Ingrid Zukerman. Strategies for generating micro explanations for Bayesian belief networks. In *Proceedings of the 5th Workshop on Uncertainty in Artificial Intelligence*, pages 295–302, Windsor, Ontario, August 1989.

[12] Ros D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, November–December 1986.

[13] Amos Tversky and Daniel Kahneman. Causal schemata in judgments under uncertainty. In M. Fishbein, editor, *Progress in Social Psychology.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[14] Michael P. Wellman. *Formulation of Tradeoffs in Planning Under Uncertainty.* PhD thesis, MIT, Laboratorium for Computer Science, MIT/LCS, TR-427, Cambridge, MA, August 1988.