How to
Believe, Desire, Reason, Decide, and Act

or

What has AI been up to Lately?

or

Revenge of the Hungry Monkey

or

Pearl for Swine

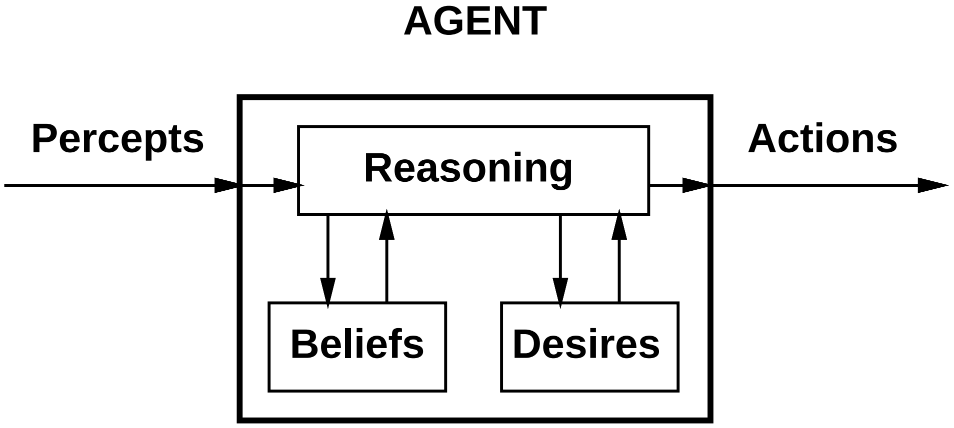
Peter Norvig

Sun Microsystems Labs Inc.

January 24, 1994

Purpose

- It has been suggested we consider certain programs as *agents:* functions from percepts to actions in the context of an environment. This talk describes an emerging consensus of how agents in an uncertain environment should be programmed.
- Advantages of agents over traditional programs: Skeptical, Purposeful, Introspective, Adaptable, Capable of Learning
- Agents are more powerful if they keep internal state representing their *beliefs* about the environment.
- Before it can *act*, an agent should have *desires* that it is attempting to fulfill.
- The agent should *reason* about its beliefs to *decide* what action is best.



Conclusions

1. Uncertain reasoning allows us to usefully summarize ignorance.
2. Probability theory is about the structure of reasoning, not about numbers.
3. Probabilities are uniquely suited to represent degree of belief.
4. Belief networks are an efficient means of representing and computing belief.
5. Before you can decide how to act, you have to know what you desire.
6. If you want to act successfully, then you must obey the rules of Decision Theory.
7. To reason properly is still NP-complete, but there are new representations and algorithms that perform well in special cases.
8. Because of these algorithms, many practitioners are moving away from the lamp post of computational tractability towards the keys of semantic clarity.
9. The semantically clear systems outperform the old ones.

The Purpose of Uncertain Reasoning

Neither $Cavity \Rightarrow Toothache$
nor $Toothache \Rightarrow Cavity$ is categorically true.

- Most environments are too complex. There are always exceptions. We can ignore, enumerate, or summarize the exceptions.
- Uncertain reasoning is the process of *summarizing* ignorance.
- Three kinds: laziness, theoretical ignorance, practical ignorance.
- Uncertain reasoning ascribes a degree of belief to states of the world.
- Assigning a degree of truth is a completely separate question that we will not address.

Approaches can be categorized by

- School: logical, computational, probabilistic
- Model: extensional (syntactic) or intensional (semantic)

Extensional (Syntactic) Reasoning

In extensional systems all we have to know about a proposition is its degree of belief. First order logic is an extensional system, where degree of belief is truth value. Advantages:

- *Locality*: Given $A \Rightarrow B$ we can prove B by proving A, without worrying about any other rules.
- *Monotonicity*: Once we prove a theorem is true, it is always true.
- *Detachment*: We can discard the evidence used to prove a proposition, once we prove it.
- *Substitutability*: If A_1 and A_2 have the same truth value, then $A_1 \vee B$ and $A_2 \vee B$ will also have the same truth value.
- *Modularity*: One can work on and debug rules in isolation (locality + detachment).
- *Incrementality*: We can add new evidence one piece at a time.

Because of these advantages, people tried to extend them to uncertain reasoning (non-monotonic logic, certainty factors, fuzzy logic, Dempster-Shafer theory).

Problems with Extensional Reasoning

Using extensional reasoning with uncertainty has problems:

- Lack of bidirectional inference. If both $Cavity \Rightarrow_{.9} Toothache$ and $Toothache \Rightarrow_{.7} Cavity$ are allowed we get unwanted cycles.
- Lack of mixed (causal/diagnostic) reasoning. Given: $Rain$, $Rain \Rightarrow_{.9} Wet$, $Wet \Rightarrow_{.3} Sprinkler$ we want belief in $Sprinkler$ to go down, not up.
- Lack of "explaining away" Given: Wet, $Rain \Rightarrow Wet$, $Sprinkler \Rightarrow Wet$, evidence for rain is evidence *against* sprinkler. We can only fix this by destroying modularity.
- Lack of detachment: $Enquirer \Rightarrow_{\epsilon} Elvis$ Limits us to tree-shaped rule bases.
- Lack of substitutability: H_x = coin x is heads:

E	$P(E)$	$P(E \wedge H_1)$
$\neg H_1$	.5	.00
H_2	.5	.25
H_1	.5	.50

- In conclusion, if we want semantic coherence, then *none* of the nice properties of certain reasoning hold.
- This rules out certainty factors, some versions of Dempster-Shafer, and some versions of fuzzy logic.

Intensional (Semantic) Reasoning

In intensional systems, uncertainty is attached to possible worlds. Substitutability holds only for states that represent the same set of possible worlds, not just the same *proportion* of possible worlds.

Intensional systems use probability as a measure of belief:

- Supports all modes of reasoning
- Mathematically well-founded
  (Centuries of Probability and Utility theory)
- Easy to use (with experience and understanding)
- Good computational methods now exist (except in the worst case)

Conclusion (2): Probability theory is about the structure of reasoning, not about numbers.

Objections to Probabilities

Many reasons have been cited for ignoring centuries of work on probability and inventing something new:

- They are "epistemologically inadequate"
  – McCarthy and Hayes (1969)
- Priors assume too much *(no harm no foul)*
- Belief should be a range *(ditto; range of P)*
- Requires massive data *(only frequentist)*
- Requires the enumeration of all possibilities *(ditto)*
- Don't know all numbers *(use what you have)*
- People are bad at estimating probabilities
  (so they need help)
- Requires unwarranted independence assumptions *(only idiot Bayes)*
- Not cognitively real *(Pearl argues the opposite)*
- Can't do learning *(au contraire)*
- Can't generate good explanations
  (harder domain, but in principle better results)

Arguments mostly accepted within AI for 25 years.

Statisticians unimpressed with "currently fashionable ad-hoc quantitative mumbo-jumbo."

Objections to Probability (continued)

> "Translated to the notation of conditional probability, this rule $(s_1, s_2, s_3 \Rightarrow h_1)$ seems to say $P(h_1|s_1, s_2, s_3) = 0.7$ ... Questioning of the expert gradually reveals ... he becomes uneasy when he attempts to follow the logical conclusion that $P(\neg h_1|s_1, s_2, s_3) = 0.3$. He claims that the three observations are evidence in favor of the conclusion ... and should not be construed as evidence against ..." – *Buchanan and Shortliffe*

The expert was presented with a situation with priors $P(h_1) = P(\neg h_1) = 0$, and becomes uneasy when $P(\neg h_1)$ goes up. No wonder!

If we don't ignore priors, then, say, $P(h_1) = .01$ and $P(\neg h_1) = .99$, and the rule makes $P(\neg h_1)$ go down, as it should.

Argument for probability

Cox's argument *American Journal of Physics*, 1946:

- Six intuitively appealing axioms of belief
  (Not completely uncontroversial)
- Only probabilities satisfy the axioms

de Finetti's argument:

- One person announces odds (3 to 1 the NFC wins Superbowl; 2 to 1 its a blow-out)
- The other person places bets in any amount
- If the announced odds disobey axioms of probability then bettor can guarantee a win

Conclusion (3): Probabilities are uniquely suited to represent belief

Probabilities can combine subjective and objective information in a way that no other method can.

Computing with Probabilities

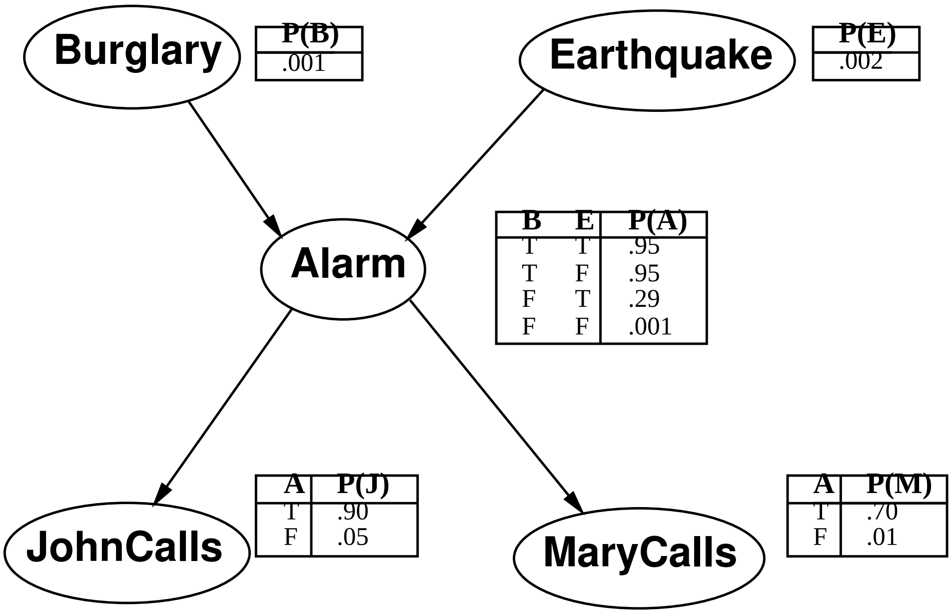
One way to use probabilities is to compute the entire *joint probability distribution*: the probability of every possible combination of states.

Problem: This is of exponential size.

Solution: use *belief networks*, which:

- Consist of random variables, links, and conditional probability tables
- Explicitly represent independence
- Represent no more than is needed, and thus require no more numbers than are needed but are always coherent.
- Represent the entire joint
- Have tractable algorithms for simple topologies
- Are (necessarily) exponential in the general case

Belief Networks



Or with rules where $A =$ `Alarm`, $B =$ `Burglary`, ...

$$True \Rightarrow_{.001} B \qquad True \Rightarrow_{.002} E$$

$$B \Rightarrow_{.95} A \qquad \neg B \wedge E \Rightarrow_{.29} A$$

$$\neg B \wedge \neg E \Rightarrow_{.001} A$$

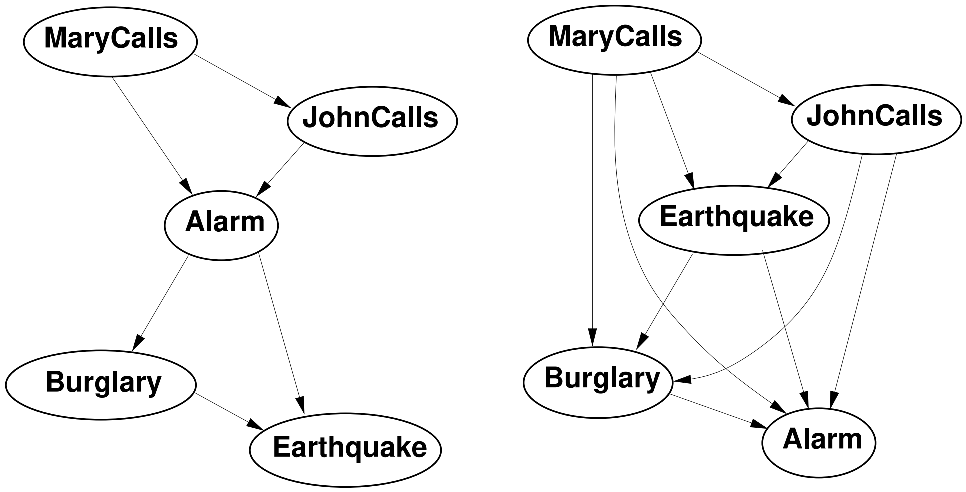
$$A \Rightarrow_{.90} J \qquad A \Rightarrow_{.70} M$$

$$\neg A \Rightarrow_{.05} J \qquad \neg A \Rightarrow_{.01} M$$

Belief Networks or Causal Networks?

Links represent non-independence, but not necessarily causation.

Causal networks tend to be both more compact and transparent.

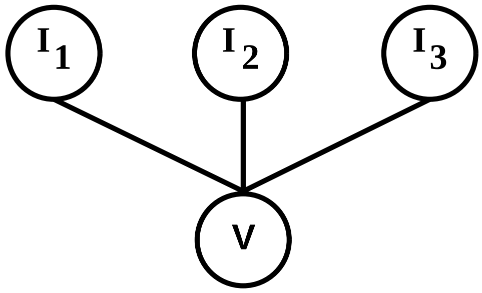


$1 + 2 + 4 + 2 + 4 = 13$ $\quad$ $1 + 2 + 4 + 8 + 16 = 31$

Representing Conditional Tables

A conditional probability table can be represented by:

- A complete table of $(|V|-1)\prod|I|$ entries (2^n for booleans)
- Deterministic logical relation
- Deterministic numeric relation
- Noisy logical relation (Noisy OR requires only n numbers)



I_1	I_2	I_3	$P(V=a)$	$P(V=b)$
T	T	T	.99	.01
T	T	F	.50	.50
	...		...	

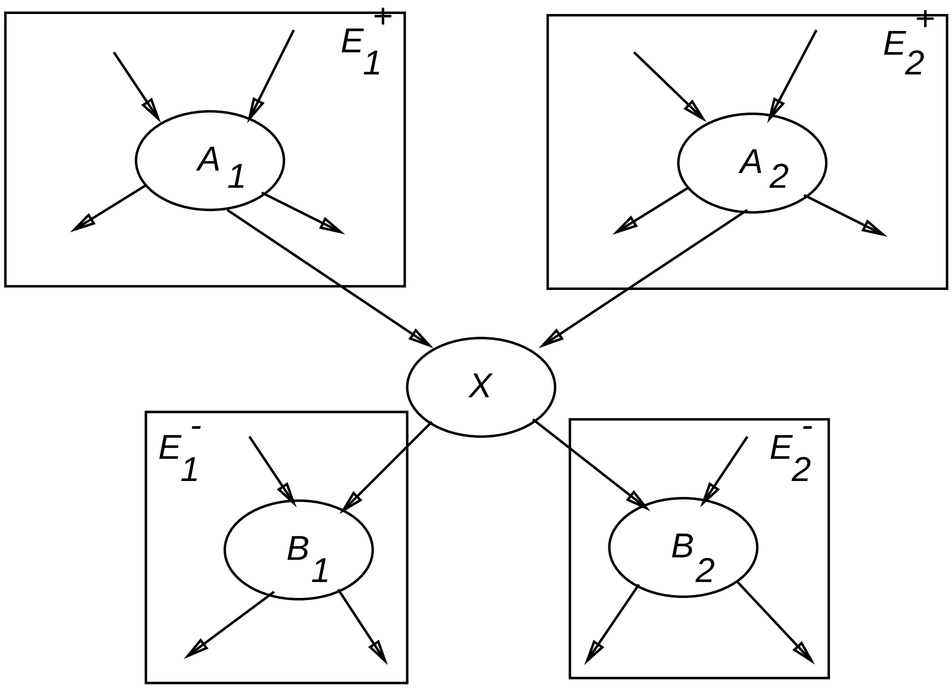
Evaluating Belief Networks

- Given value(s) of any variable(s), can compute any other(s).
- For singly-connected networks (polytrees), evaluation is $O(Diameter)$.
- Multiply-connected networks are reduced to singly-connected by:
  - Clustering nodes (generating larger tables)
  - Instantiating nodes (generating a forest)
  - Stochastic simulation (giving an approximation)

Current work is in efficient algorithms for approximating, learning, partially evaluating, and incrementally modifying belief networks.

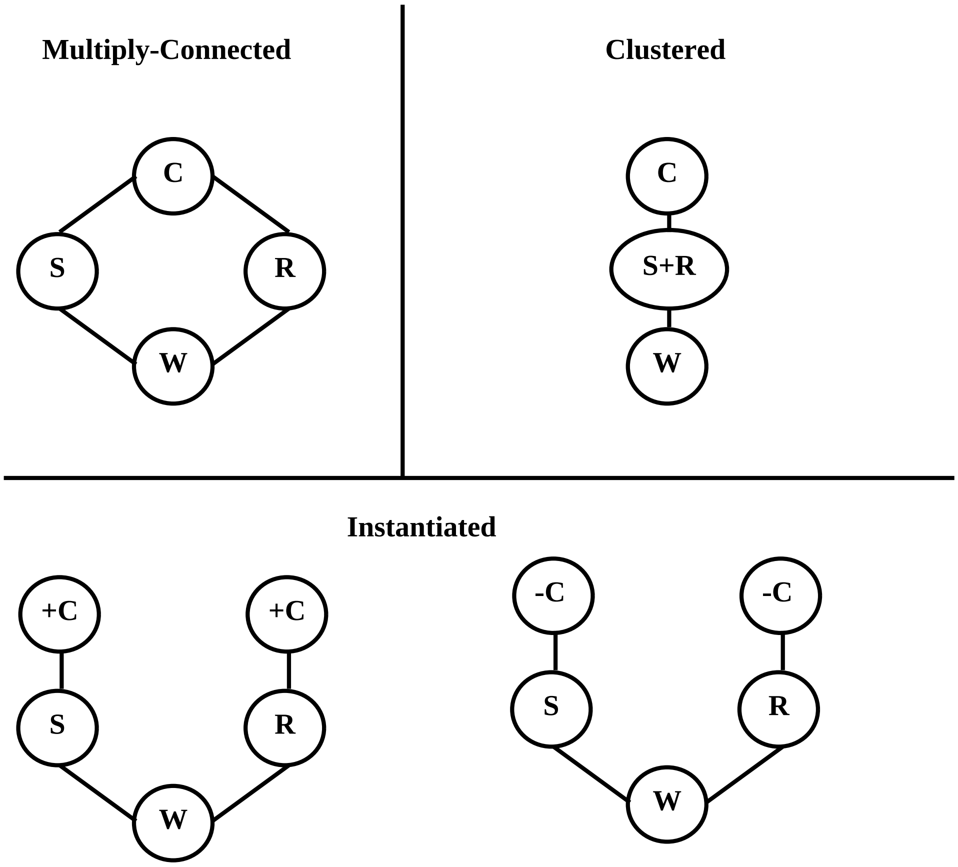
Conclusion (4): Belief networks are an efficient means of representing and computing belief.

Singly-Connected (Polytrees)



- Summary of all influence comes through neighboring nodes
- Therefore, local computations suffice
- Parallel implementations are possible

Multiply-Connected Networks



The Nature of Desire

- Traditional programs have no explicit desires; just epiphenomenal ones
- AI programs usually use *goals* as desires (and intentions).
- Decision theory uses *preferences* over states, which are more general.
- We can define p as a goal iff p is preferred to $\neg p$.
- If we want to reason about composite desires and partial satisfaction we can use *lotteries*.

Some notation:

- Outcome A is preferred to B: $A \succ B$
- Outcome A and B are indifferent: $A \sim B$
- Outcome A is preferred or indifferent to B: $A \succsim B$
- Outcome A occurs with probability p, else B:
  The lottery $[p, A;\ B]$

The Axioms of Utility Theory

1. **Orderability** An agent should know what it wants.

$$(A \succ B) \vee (B \succ A) \vee (A \sim B)$$

2. **Transitivity** No money pump.

$$(A \succ B) \wedge (B \succ C) \Rightarrow (A \succ C)$$

3. **Continuity** There's always a trade-off point

$$A \succ B \succ C \Rightarrow \exists p \ [p, A; \ C] \sim B$$

4. **Substitutability** of equals for equals

$$\sim B \Rightarrow [p, A; \ C] \sim [p, B; C]$$

5. **Monotonicity** Better odds are better

$$A \succ B \Rightarrow (p \geq q \Leftrightarrow [p, A; \ B] \succ [q, A; \ B])$$

6. **Decomposability** No fun in gambling.

$$[p, A; \ [q, B; \ C]] \sim [p, A; \ (1-p)q, B; \ C]$$

The Utility Principle

If an agent's preferences obey the axioms of utility theory then there *must* exist a function U from states to real numbers that is consistent with preferences in that:

$$U(A) > U(B) \Leftrightarrow A \succ B$$

$$U(A) = U(B) \Leftrightarrow A \sim B$$

$$U(A) \geq U(B) \Leftrightarrow A \succsim B$$

Conclusion (5): Before you can decide how to act, you have to know what you desire. Utilities are a way of specifying desires.

Decision theory concentrates on how to elicit utilities and make sure they are coherent.

AI (Doyle, Wellman) concentrates on flexible ways of representing and combining atomic preference judgements and policies.

Trivia: A micromort is worth about $20, across several studies.

The Nature of Decisions

> "To judge what one must do to obtain a good or avoid an evil, it is necessary to consider not only the good and the evil in itself, but also the probability that it happens or does not happen; and to view geometrically the proportion that all these things have together."
> — Arnauld, *The Port-Royal Logic*, 1662 (Pascal's school)

The principle of Maximum Expected Utility:

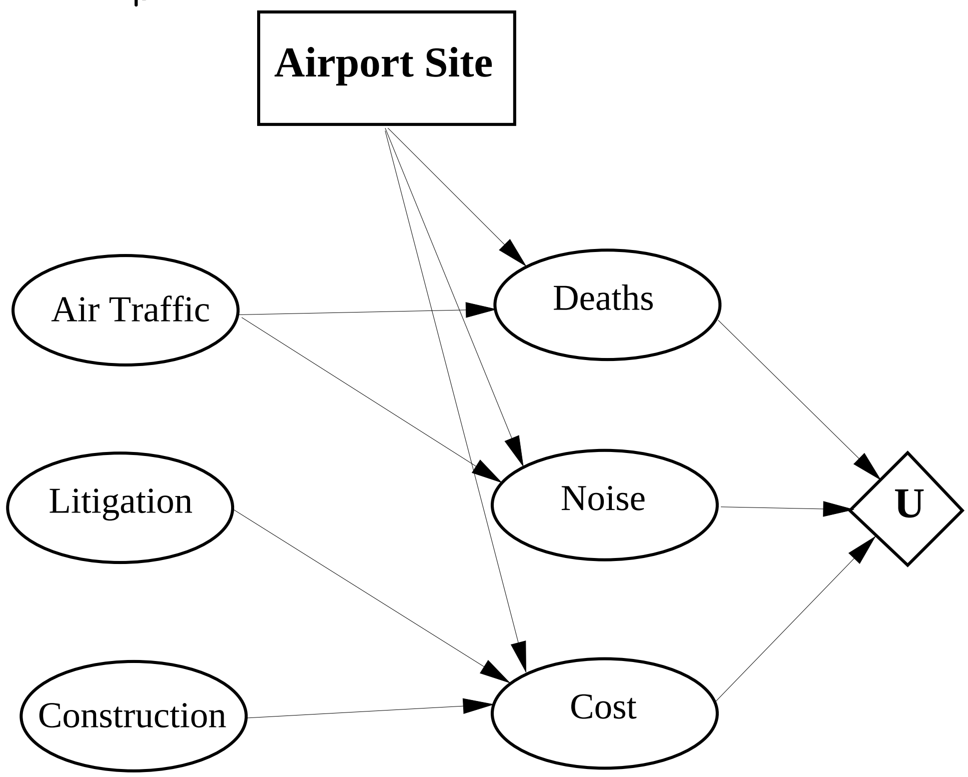
Choose the action with the maximum expected utility.

- Each action is represented as a lottery of possible outcomes
- The expected utility of an action is the utility of the lottery:
  $U([p_1, S_1;\ \ldots\ p_n, S_n]) = \sum_i p_i U(S_i)$

Conclusion (6): If you want to act successfully, then you must obey the rules of Decision Theory. If you don't you will do worse in the long run.

Influence Diagrams

An influence diagram is a belief network with one additional node (diamond) representing the final utility, and zero or more nodes (squares) representing choice or decision points.



Evaluating Influence Diagrams

- Consider all possible values of the decision node(s)
- Evaluate the diagram as a belief network for each setting
- Choose the decision that gives the best utility
- This means a rational agent looks like:
  - Update probabilities based on current percept
  - Evaluate the resulting influence diagram
  - Select the action with best utility
  - Execute it
  - Available actions include changing the network's topology or tables
- Current research is in *sequential* decision-making
- What and when to learn is also interesting

The biggest subroutine is still evaluating a belief network.

Conclusion (7): To reason properly is still NP-complete, but there are new representations and algorithms that perform well in special cases.

Conclusion (8): Because of these algorithms, many practitioners are moving away from the lamp post of computational tractability towards the keys of semantic clarity.

An Example: Pathfinder

The Pathfinder program (Heckerman 1991) diagnoses lymph-node diseases. Over 60 diseases and 100 findings.

Four versions:

- I: Rule-based (MRS); no uncertainty
- II: Tried several methods: Certainty factors, Dempster-Shafer, Idiot Bayes (best)
- III: Idiot Bayes, expert interview protocol pays attention to low probabilities (10% of errors on II due to P=0)
- IV: Belief network

Comparison of III and IV:

- III rated at 7.9/10; IV at 8.9/10 (significant)
- At $20/micromort, IV makes $6,000/patient.

Conclusion (9): The semantically clear systems outperform the old ones.

Bibliography

In Statistics and Decision Theory:

- Cardano (1565), Pascal and Fermat (1654), Huygens (1657), Bayes (1763), Laplace (1812).
- Cox, *Probability, Frequency, and Reasonable Expectation*, American Journal of Physics, 1946.
- Good, *Rational Decisions*, Journal of the Royal Statistics Society, 1952.
- Jeffrey, *The Logic of Decision*, Chicago, 1983.
- Howard and Matheson, *Influence Diagrams*, in "Applications of Decision Analysis," 1984.
- Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, 1985.
- Spiegelhalter, *A Statistical View of Uncertainty in Expert Systems*, in "Artificial Intelligence and Statistics, Gale, Addison Wesley 1986.
- *Statistical Science*, Feb. 1987 (Special Issue on Uncertainty in AI).

Bibliography

In AI (1970's and 80's):

- Feldman and Sproull, *Decision Theory and AI II: The Hungry Monkey*, Cognitive Science, 1977.
- Pearl, *Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach*, AAAI, 1982.
- Pearl and Kim, *A Computational Model for Causal and Diagnostic Reasoning in Inference Systems*, IJCAI, 1983.
- Cheeseman, *In Defense of Probability*, IJCAI, 1985.
- *Proceedings of the* N*th Uncertainty in AI Conference*, 1985-1993.
- Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988.

Bibliography

In AI (1990's):

- Neapolitan, *Probabilistic Reasoning in Expert Systems*, Wiley, 1990.
- Shafer and Pearl, *Readings in Uncertain Reasoning*, Morgan Kaufmann, 1990.
- Dean and Wellman, *Planning and Control*, Morgan Kaufmann, 1991.
- Heckerman, *Probabilistic Similarity Networks*, MIT Press (ACM Doctoral Dissertation Award), 1991.
- Charniak, *Bayesian Networks without Tears*, AI Magazine, Winter 1991.
- Henrion, Breese & Horvitz, *Decision Analysis and Expert Systems*, AI Magazine, Winter 1991.
- Elkan, *The Paradoxical Success of Fuzzy Logic*, AAAI (Best paper award) 1993.