TOOLS AND TECHNIQUES

Uncertainty Management in Expert Systems

Keung-Chi Ng and Bruce Abramson
University of Southern California

One of the most popular current approaches to AI involves constructing programs that function as narrowly focused experts. The past decade has seen expert systems mature from experimental research projects into important contributors at many medical, engineering, commercial, and financial institutions. Nevertheless, problematic expert system design issues remain — uncertainty management is a case in point.

In this setting, the term "uncertainty" refers to a set of questions that human experts ask themselves almost daily: How do I fill the gaps in my own expertise? What do I do when not all of the specifics are known? How can I account for unpredictable events? How do I differentiate among events that appear to have identical causes? Since these (and related questions) represent issues which every human decision maker must constantly face, they are also issues that an automated expert system should be able to handle.

Uncertainty arises from a variety of sources, and confounds system designers in a variety of ways. As a result, the method with which a system handles uncertain information forms is a crucial component of its overall performance. Although every existing system copes differently, the schemes that have been implemented are all variations on a few major themes, with varying degrees of success. In an attempt to discover the relative merits of the more popular approaches, this survey overviews three complete theories of belief propagation — probability, Dempster-Shafer, and possibility theories, as well as some techniques devised specifically for expert systems — MYCIN's certainty factors, Prospector's subjective Bayesian method, and Cohen's theory of endorsements. Although most uncertainty research has focused on the strengths of one paradigm to the exclusion of both its shortcomings and its competitors, this survey will reveal both the benefits and limitations of the various schemes, present examples of expert systems within each school, and discuss some relevant open problems.

Before we begin, it is only fair to warn readers that both authors are believers in subjective probability theory. While our presentation may be biased towards probability, we will try to objectively present the basics of the various approaches. First we'll review basic expert system terminology, then detail several uncertainty management paradigms, followed by an assessment of their relative merits. We conclude with a discussion of some open problems and directions for future research.

Expert system terminology

An expert system is a computer program with three special characteristics:

0885/9000/90/0400-0029 $1.00 © 1990 IEEE

- Its database is designed with the help of a human expert.
- Its problem domain is quite narrow.
- It is expected to perform on par with a human expert.

Only one of these properties is not self-explanatory — the narrow problem domain. As an illustration, consider medical expert systems. Because medicine is such a broad domain, medical expert systems do not attempt to codify all medical knowledge; they focus on individual specializations and subspecializations. Among the existing medical expert systems, MYCIN[1] can only diagnose infectious blood diseases, while Internist[2] is applicable only to internal medicine. Like all programs, expert systems must be able to represent, manipulate, and communicate data. Three main system components incorporate these tasks — the knowledge base, the inference engine, and the user interface. Figure 1 shows these components' interconnections in an expert system.[1] While the role of the user interface is obvious, the precise nature of knowledge bases and inference engines may require further clarification.

A knowledge base contains expert-level information necessary to solve problems in a specific domain. This information is represented in several ways, but commonly as a set of rules. Other representation schemes include frames, semantic nets,[3] and belief networks.[4] Representing knowledge as rules will be assumed throughout this article, unless otherwise specified. A rule usually takes the form "If <if-part> Then <then-part>." The if-part of a rule is commonly referred to as its left-hand side, premise, or antecedent. Premises contain a collection of conditions that must be satisfied before the rule may be used. The then-part of a rule (its right-hand side, conclusion, or consequence) contains a set of actions to be performed when the rule is applied. Knowledge bases are inherently domain-specific, hence nontransferable — expert systems built for different domains require different knowledge bases.

An inference engine uses information stored in the knowledge base to solve problems. In addition to its ongoing interaction with the knowledge base, the inference engine also records the facts known about the current problem in a database called working memory, which gets updated as new information becomes available. In a rule-based expert system, the inference engine usually has a pattern matcher and a rule applier. The pattern matcher determines which rules are relevant — comparing information in working memory with the premises of every rule in the knowledge base, and recording the rules whose premises are satisfied. Then the rule applier selects a relevant rule to be applied. If it finds none, the rule applier does not act. On the other hand, when multiple rules apply, the applier selects and applies the most specific one (according to some heuristics called conflict resolution strategies). New information is created as the actions outlined in the selected rule's then-part are performed. The inference engine repeats this match-select-act cycle of interaction between working memory and knowledge base until nothing more can be done (that is, no rule matches). Since the knowledge representation scheme determines inference engine structure, different knowledge bases can use the same inference engine.

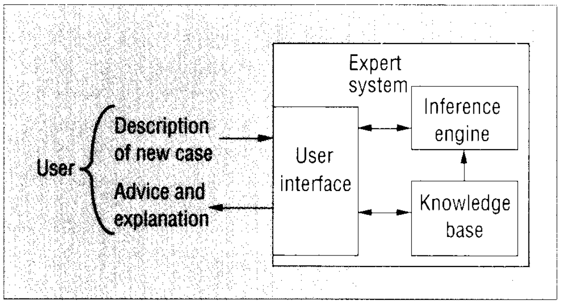


Figure 1. Major components of an expert system (arrows indicate information flow).

Uncertainty-related problems pervade the entire system. The design of expert knowledge bases, for example, must recognize that expertise is rarely complete or exact. If all information could be represented completely and exactly, any logically sound inferencing scheme would draw valid conclusions. Thus, knowledge bases constitute one major source of uncertain information in expert systems. We can trace this inherent uncertainty to the following four main sources: unreliable information, imprecise descriptive languages, inference with incomplete information, and poor combination of knowledge from different experts.[5]

- **Information can be unreliable.** This is usually due either to ill-defined domain concepts or to inaccurate data. In addition, rule-based systems often suffer from weak implications — when the expert is unable to establish a concrete correlation between a rule's premise and its conclusion. Many such systems treat weak implications by quantifying the degree of correlation; for example, MYCIN introduced numeric certainty factors (CF) for expressing correlations, and the resultant rules took the form "If <premise> Then (CF) <conclusion>."
- **Descriptive (or implementation) languages lack precision.** The numerous ambiguities in natural language are rarely clarified during translation to a formal language. As a result, rules that are not expressed precisely in the formal language can be misinterpreted. Thus, a perfect matching of facts with premises will rarely be adequate; the meaning of the facts must be approximately matched with those of the premises.
- **Inferences are sometimes drawn with incomplete information.** When the available information is incomplete, rule-based representations can't hope to be any better. The remedy resembles that for imprecision —

approximate pattern matching — except that the system accepts the value "unknown" while evaluating the premise's degree of certainty.

- **Experts sometimes disagree.** Combining the views of multiple experts into a consensus knowledge base is difficult, confusing, and frequently impossible. When all experts draw similar conclusions, consensus is generally derivable. When the experts have contradictory viewpoints, however, the combined conclusions are suspect. A rule-based system must resolve all conflicting rules before it can develop a consensus knowledge base. To derive a consensus, systems commonly attach a weight to each expert and calculate the composite conclusion. The weights, however, may be more of a problem than a solution, because no systematic way exists to obtain them. In addition, human beings generally do not have uniform levels of expertise throughout a domain.

In short, any good expert system must be able to handle uncertainty, because nearly every interesting domain contains data that are inherently inexact, incomplete, or unmeasurable. Numerous paradigms have been proposed to handle uncertainty in expert systems, some quantitative and some qualitative.[6] The popular quantitative or numeric methods include subjective probability theory,[7,8] Dempster-Shafer theory,[9] possibility theory,[10] certainty factors,[11] and Prospector's subjective Bayesian method.[12] The main qualitative or non-numeric approach is Cohen's theory of endorsements.[13] Each of these approaches has definite strengths and weaknesses. This survey will discuss the relative merits of these major schools as paradigms for uncertainty management, as well as their applicability to specific domains and their implementations in functioning expert systems.

Subjective probability theory

The basic concept of probability is so natural that it plays a significant role in our day-to-day lives. Conversations regarding the probability of rain or the odds that the Lakers will win the NBA championship are everyday occurrences (although only in Los Angeles is the latter event viewed as more likely than the former). Many popular games are also based on simple probabilistic calculations; the relative likelihood of rolling a natural (two dice summing to seven or 11) versus rolling craps (two dice totaling two, three, or 12) is crucial to developing proper payoffs in a craps game.

History. The concept of probability has been around for centuries; it may be traced back thousands of years to the introduction of words like "maybe," "probable," "chance," and "luck" (or more appropriately, their ancient equivalents) into spoken languages.[14] The mathematical theory of probability, however, was not formulated until relatively recently (around 1660). (For more details on the emergence of probability theory mathematics, see Hacking.[15]) An event's probability is classically defined as "the proportion of cases in which the given event occurs," but other definitions are possible.[14] In fact, several interpretations of the term are currently vying for supremacy. Since the turn of the century, three views have dominated:[16]

- **Objectivistic.** Probability measures the ratio of occurrences to observations for a proposition, in the long run. In other words, the law of large numbers guarantees that given enough observations, the percentage of an event's occurrence will approach objective probability.
- **Personalistic, subjectivistic, or judgmental.** Probability measures the confidence that an individual has about a particular proposition's truth. This view postulates that the individual concerned is in some way "reasonable," but it does not deny the possibility that two reasonable individuals faced with the same evidence may have different degrees of confidence in the same proposition's truth. The term "Bayesian" is often used as a synonym for subjective probability.
- **Necessary or logical.** Probability measures the extent to which one set of propositions, out of logical necessity and apart from human opinion, confirms the truth of another. In other words, probability measures inferential soundness or provability. Proponents generally regard this view as an extension of logic.

It is not surprising that these different probability interpretations use different inference schemes. However, only two main schools of probability calculus exist: Pascalian (or conventional) and Baconian (or inductive). Pascalian calculus uses Bayes' rule for belief revision; Baconian calculus uses the rules of logic to prove or disprove hypotheses. Thus, conventional probabilities cannot be derived from inductive probabilities, and vice versa. Proponents of the objectivistic and subjectivistic views follow Pascalian calculus; supporters of the logical view use Baconian calculus.

Most expert systems have used Pascalian probability theory; although Schum has studied Baconian calculus while analyzing legal evidence,[17] this probability theory has yet to appear in an expert system. (See our references for more information.[17,18]) Thus, we will assume Pascalian calculus throughout the rest of this article.

In expert systems, a knowledge base stores human knowledge. Thus, in representing an expert's knowledge with probability theory, the only appropriate interpretation of probability is as subjective belief.[19] As a result, most expert systems that have used probability theory were built by Bayesian researchers.

The basics. Before we introduce Bayes' Theorem, let's examine some fundamental probability theory terminology. Any standard probability text can provide a more detailed discussion.[7,20]

Let A be an event in the world. The collection of all possible elementary events, Ω, is called the sample space or the event space. The probability of an event A is denoted by $p(A)$, and every probability function p must satisfy three axioms:

(1) The probability of any elementary event A is nonnegative, or $\forall A \in \Omega : p(A) \geq 0$.

(2) The probability of the entire sample space is one, or $p(\Omega) = 1$.

(3) If k events $A_1, A_2, \ldots, A_k$ are mutually exclusive (that is, they cannot occur simultaneously), then the probability that at least one of these events will occur is the sum of the individual probabilities, or $p(A_1 \cup A_2 \cup \ldots \cup A_k) = \sum_{i=1}^{k} p(A_i)$.

Axioms 1 and 2 combine to yield

$$\forall A \in \Omega : 0 \leq p(A) \leq 1 \qquad \textbf{(1)}$$

Equation 1 shows that the probability of any event is between 0 and 1. By definition, when $p(A) = 0$, event A will never occur, and when $p(A) = 1$, event A must occur. A's complement $(\neg A)$ contains the collection of all elementary events in Ω except A. Since A and $\neg A$ are mutually exclusive, and $A \cup \neg A = \Omega$, Axiom 3 yields

$$\begin{aligned} p(A) + p(\neg A) &= p(A \cup \neg A) \\ &= p(\Omega) \\ &= 1 \end{aligned} \qquad \textbf{(2)}$$

Rewriting this equation as $p(\neg A) = 1 - \mathrm{p}(A)$ provides an easy way to compute $\mathrm{p}(\neg A)$ from $p(A)$.

Suppose that $B \in \Omega$ is another event. The probability that A will occur given that B occurs, written $p(A|B)$, is called the conditional probability of A given B. The probability that both A and B will occur, $p(A \cap B)$, is called the joint probability of A and B. By definition, the conditional probability $p(A|B)$ equals the ratio of the joint probability $p(A \cap B)$ to the probability of B (provided that the probability of B, $p(B)$, is nonzero), or

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \qquad \textbf{(3)}$$

Similarly, the conditional probability of B given A, $p(B|A)$, equals $p(B \cap A) / p(A)$, and thus $p(B \cap A) = p(B|A) \times p(A)$. Since joint probability is commutative, $p(A \cap B) = p(B \cap A)$. Hence, $p(A \cap B) = p(B \cap A) = p(B|A) \times p(A)$. Substituting this equality into Equation 3 yields Bayes' rule

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)} \qquad \textbf{(4)}$$

So far, no assumptions have been made about either A or B. If these two events are independent (one event's occurrence does not affect the other's occurrence), then by definition $p(A|B) = p(A)$ and $p(B|A) = p(B)$. This definition is inspired by the idea that if two events are truly independent, then the occurrence of the first should have no effect on the occurrence of the second. It also suggests a definite relationship between set theory and probability theory. If A and B are disjoint sets, set union corresponds to a sum of probabilities and set intersection corresponds to a product of probabilities. (Without the independence assumption, these relationships are not precise, and calculations must make use of the inclusion-exclusion formula.) In other words, if A and B are disjoint sets, $p(A \cup B) = p(A) + p(B)$ and $p(A \cap B) = p(A) \times p(B)$.

Continuing with set-theoretic notation for just a moment longer, recall that B can be written as $(B \cap A) \cup (B \cap \neg A)$. Since this union is clearly disjoint,

$$\begin{aligned} p(B) &= p((B \cap A) \cup (B \cap \neg A)) \\ &= p(B \cap A) + p(B \cap \neg A) \\ &= p(B \mid A) \times p(A) + p(B \mid \neg A) \times p(\neg A) \end{aligned} \qquad \textbf{(5)}$$

By switching notation from sets back to events, Equations 4 and 5 can be combined to yield

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B|A) \times p(A) + p(B|\neg A) \times p(\neg A)} \qquad \textbf{(6)}$$

Equation 6 lays the groundwork for using probability theory to manage uncertainty — it provides a way of obtaining the conditional probability of A given B from the conditional probability of B given A. This relationship allows expert systems to "turn rules around," so to speak. Consider a case in which all rules in an expert system are expressed in this form: "If <H is true> Then <E will be observed with probability p>." Clearly, if H is observed, this rule states that the probability that event E occurred is p. But what if the status of H is unknown, and E is observed? Equation 6 tells us how to compute the probability that H is true, as well.

This shift from A and B to H and E is not accidental. With Equation 6 we leave general probability theory and begin analyzing probabilistic calculations in expert systems. In this context, H usually represents a hypothesis and E denotes a piece of evidence. Thus we can rephrase Equation 6 in terms of hypotheses and evidence:

$$p(H|E) = \frac{p(E|H) \times p(H)}{p(E|H) \times p(H) + p(E|\neg H) \times p(\neg H)} \qquad \textbf{(7)}$$

Equation 7 relates a hypothesis to a piece of supporting evidence, while also relating observed evidence to an as-yet-unsubstantiated hypothesis. This interpretation also suggests defining the prior probability of hypothesis H, $p(H)$, as the probability assigned to H prior to the observation of any evidence.

In expert systems, the probabilities that are required to solve a problem are provided by human experts and stored in the knowledge base. These probabilities include the prior probabilities for all possible hypotheses ($p(H)$) and the conditional probabilities for observing a piece of evidence given a hypothesis ($p(E|H)$).

In medical diagnoses, for example, an expert physician must provide the prior probabilities of all possible diseases for a particular medical domain. Moreover, the conditional probability for observing a symptom given a specific disease must also be obtained for all pairs of symptoms and diseases (assuming that all symptoms are conditionally independent given a disease). (Two events E_1 and E_2 are conditionally independent if their joint probability given some hypothesis H equals the product of the conditional probabilities of each event given H, or $p(E_1E_2|H) = p(E_1|H) \times p(E_2|H)$.[21]) Users give the expert system information about observations (the presence of certain symptoms), and the system computes $p(H_i|E_j \ldots E_k)$ for all hypotheses ($H_1 \ldots H_m$) in light of the supplied symptoms ($E_j \ldots E_k$) and the stored probabilities. The probability $p(H_i|E_j \ldots E_k)$ is called the posterior probability of hypothesis H_i upon observing $E_j, \ldots, E_k$. These probabilities, calculated using Equation 7, provide comparative rankings for all possible hypotheses (that is, hypotheses with nonzero posterior probabilities).

Equation 7 is inherently limited in that each piece of evidence will affect only one hypothesis. We can generalize it as follows to account for both multiple hypotheses ($H_1, H_2, \ldots, H_m$) and multiple pieces of evidence ($E_1, E_2, \ldots, E_n$).

Single evidence, multiple (mutually exclusive and exhaustive) hypotheses follow:

$$p(H_i|E) = \frac{p(E|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E|H_k) \times p(H_k)} \qquad \textbf{(8)}$$

Multiple evidence, multiple (mutually exclusive and exhaustive) hypotheses follow:

$$p(H_i|E_1E_2 \ldots E_n) = \frac{p(E_1E_2 \ldots E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1E_2 \ldots E_n|H_k) \times p(H_k)} \qquad (9)$$

Unfortunately, Equation 9 causes pragmatism to rear its ugly head. The denominator requires us to know the conditional probabilities of all possible combinations of evidence for all hypotheses. This requirement makes Bayes' rule unworkable for most applications. To lighten the burden, we often assume conditional independence among pieces of evidence given a hypothesis. This reduces Equation 9 to

$$p(H_i|E_1E_2 \ldots E_n) = \frac{p(E_1|H_i) \times p(E_2|H_i) \times \ldots \times p(E_n|H_i) \times p(H_i)}{\sum_{k=1}^{m} p(E_1|H_k) \times p(E_2|H_k) \times \ldots \times p(E_n|H_k) \times p(H_k)} \qquad \textbf{(10)}$$

One can assume conditional independence among different pieces of evidence to suppress evidential subtleties in expert systems.

Belief propagation. Belief is propagated through a system by using Bayes' rule to compute all posterior probabilities of hypotheses, given observed evidence. These posterior probabilities provide ranking information about potentially true hypotheses.

Let's look at an example to illustrate the process. Suppose that in some system, three mutually exclusive and exhaustive hypotheses — H_1, H_2, and H_3 — have prior probabilities $p(H_1)$, $p(H_2)$, and $p(H_3)$, respectively. We then observe two conditionally independent pieces of evidence, E_1 and E_2, that support these hypotheses to differing degrees. Table 1 shows the conditional and prior probabilities for all hypotheses and evidence in this example. (These numbers illustrate a point and do not correspond to a specific problem.)

As evidence is collected, hypothesis belief will increase if evidence supports it and decrease if evidence opposes it. Suppose that we observe E_1 before E_2. Observing E_1, we compute the posterior probabilities for the hypotheses according to Equation 8,

$$p(H_i|E_1) = \frac{p(E_1|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k) \times p(H_k)} \quad , \ i = 1,2,3$$

Thus,

$$p(H_1|E_1) = \frac{0.4 \times 0.5}{0.4 \times 0.5 + 0.8 \times 0.3 + 0.3 \times 0.2} = 0.40$$

$$p(H_2|E_1) = \frac{0.8 \times 0.3}{0.4 \times 0.5 + 0.8 \times 0.3 + 0.3 \times 0.2} = 0.48$$

Table 1. The prior and conditional probabilities used in the above example.

i	1	2	3
$p(H_i)$	0.5	0.3	0.2
$p(E_1|H_i)$	0.4	0.8	0.3
$p(E_2|H_i)$	0.7	0.9	0.0

$$p(H_3|E_1) = \frac{0.3 \times 0.2}{0.4 \times 0.5 + 0.8 \times 0.3 + 0.3 \times 0.2} = 0.12$$

After E_1 is observed, belief in hypotheses H_1 and H_3 decreases while belief in H_2 increases. After observing E_2 as well, we compute the posterior probabilities by

$$p(H_i|E_1E_2) = \frac{p(E_1E_2|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1E_2|H_k) \times p(H_k)} \quad ,i = 1,2,3$$

Since E_1 and E_2 are conditionally independent given hypothesis H_i,

$$p(Hi|E_1E_2) = \frac{p(E_1|H_i) \times p(E_2|H_i) \times p(H_i)}{\sum_{k=1}^{3} p(E_1|H_k) \times p(E_2|H_k) \times p(H_k)} \quad ,i = 1,2,3$$

Hence,

$$p(H_1|E_1E_2) = \frac{0.4 \times 0.7 \times 0.5}{0.4 \times 0.7 \times 0.5 + 0.8 \times 0.9 \times 0.3 + 0.3 \times 0.0 \times 0.2} = 0.393$$

$$p(H_2|E_1E_2) = \frac{0.8 \times 0.9 \times 0.3}{0.4 \times 0.7 \times 0.5 + 0.8 \times 0.9 \times 0.3 + 0.3 \times 0.0 \times 0.2} = 0.607$$

$$p(H_3|E_1E_2) = \frac{0.3 \times 0.0 \times 0.2}{0.4 \times 0.7 \times 0.5 + 0.8 \times 0.9 \times 0.3 + 0.3 \times 0.0 \times 0.2} = 0.0$$

Although the initial ranking was H_1, H_2, H_3, only H_1 and H_2 remain under consideration after E_1 and E_2 were observed; H_2 is now considered more likely than H_1.

Expert system examples. To use probability theory to represent uncertainty, expert system designers must obtain all prior and conditional probabilities from human experts. Although they usually assume conditional independence to reduce the number of required probability assessments, they still need numerous assessments. Thus, it is not surprising that few expert systems have used subjective probability theory, and that many of these systems can only solve relatively uncomplicated problems.

In the early 1970's, de Dombal and his associates at the University of Leeds used statistical data to develop a computer program for diagnosing acute abdominal pain.[22] (Unlike other expert systems, this system was never given a specific name.) This program avoids combinatorial explosion by restricting itself to seven diseases; the same group has worked recently on dyspepsia, similarly using a narrow range of diseases.[23]

A more recent expert system, Pathfinder also uses subjective probability theory.[24] Without assuming conditional independence among symptoms, Pathfinder diagnosed 63 lymph node diseases with around 110 symptoms. Pathfinder tracks dependencies among observed features using influence diagrams — graphical representations of the relationships and dependencies among aleatory variables and decisions. (For more information on influence diagrams, see our references.[25,26]) This relatively new tool enables Bayesian researchers and decision analysts to visualize probabilistic dependencies in decision analysis and to specify information states for which independencies can be assumed. IDES, another expert system based on influence diagramming, has recently been developed at the University of California at Berkeley.[27]

Analysis. The main difficulty in implementing subjective probability theory is the huge number of probabilities that must be obtained to construct a functioning knowledge base. If, for example, some medical diagnosis domain has 100 diseases and 700 relevant, observable symptoms, then at least 70,100 probability values (70,000 conditional probabilities and 100 prior probabilities) must be obtained — assuming that all the diseases are mutually exclusive, all symptoms are conditionally independent given a disease, and all evidence and variables are restricted to two values (true and false). Incorporating symptomatic dependencies and allowing the possibility of multiple diseases will bring the assessment number even higher. Unfortunately, assuming conditional independence is rarely valid,[28] and assuming mutual exclusivity and exhaustivity of disease categories is usually false — concurrent and overlapping categories are quite common.[29] Thus, a fundamental challenge facing probability proponents is removal of conditional independence and mutual exclusivity assumptions when dealing with real-world problems.

One effort to solve this problem deals with belief networks — probabilistic networks of conceptually related propositions, in which numbers regulate and propel information flow. A belief network is a special kind of influence diagram: While influence diagram nodes can represent decisions, chances, or values, belief networks contain only chance nodes. Recently, Pearl showed that by representing information in a knowledge base with Bayesian networks or belief networks, one can construct consistent probabilistic knowledge bases without imposing unnecessary conditional-independence assumptions.[4,30] These networks also ensure that evidence favoring a hypothesis will not be construed as partial support for its negation, and that sound explanations can be produced by tracing beliefs back to their sources. Efficient algorithms for belief propagation in a singly connected belief network (with only a single pathway from any node to any other node) are available.[4] However, it is unlikely that an efficient algorithmic solution exists for the general inference problem.[31] (Such "NP-hard" problems play a crucial role in complexity theory.[32]) No efficient algorithm exists for belief propagation in multiply connected networks, but Pearl has proposed a reasonable scheme for networks that are not highly connected.[4] Recently, Lauritzen and Spiegelhalter developed a belief propagation algorithm based on graph theory.[33] In addition to Pearl's detailed analysis of belief network theory, other researchers have proposed applying decision-theoretic concepts to AI and expert systems.[34] Decision theory and analysis may soon become relevant to the expert system community.[21]

Other challenges against using probability theory for belief propagation focus on the interpretation of prior probabilities provided by human experts. Prior probabilities, for example, are usually assigned equally across a set of items in which the expert is ignorant or indifferent.

If the expert has some information, however, the prior probabilities tend to be non-uniform. According to some, this method of representing ignorance assumes more information than is given; therefore, alternative schemes should be adopted. Probability theorists, on the other hand, contend that no extra information is assumed, and that the confusion simply arises from misunderstanding probability.[19]

Summary. Probability theory is a well-developed paradigm for representing events of unknown truths. As a result, it is not surprising that subjective probability theory was adopted by at least one school of researchers interested in expert system uncertainty management. Bayesian researchers view probabilities as measures of belief; the relationship between this interpretation and Bayes' rule coined the field's name. Expert system researchers, of course, have helped elevate Bayes' rule from a theoretical formula for calculating conditional probabilities to an integral part of any application that requires belief propagation with probabilities.

Probability theory has both strengths and weaknesses as an approach for uncertainty management. Probability proponents point mainly to its well-formalized methodology and to its nearly universal applicability. Its detractors, on the other hand, stress the inherent difficulties of eliciting and encoding expertise, as well as the prodigious number of calculations required for belief updating. Although such debates will undoubtedly continue throughout the foreseeable future, probabilistic computations would be ideal given reliable probabilities and extensive computer power.

Dempster-Shafer theory

Dempster-Shafer theory was developed by Arthur Dempster[35] in the 1960's and extended by Glen Shafer[9] in the 1970's. This theory was motivated by two difficulties these researchers had with probability theory: the representation of ignorance, and the idea that the subjective beliefs assigned to an event and its negation must sum to one.

Background. Ever since probability theory originated in the 17th century, several aspects of the field have been extensively disputed, including the representation of ignorance. The traditional method represents ignorance by indifference or by uniform probabilities. Some challenge this approach, because uniform probabilities seem to represent more information than is given — one can attribute equal prior beliefs to either complete ignorance or equal belief in all hypotheses. Furthermore, new evidence obscures the original ignorance expressed in the prior belief. Another heavily debated point involves fixing the probability of a hypothesis's negation once the probability of its occurrence is known, because $p(H) + p(\neg H) = 1$. Shafer claimed that, in many situations, evidence that only partially favors a hypothesis should not be construed as also partially supporting its negation.[9] DST tackles these problems by working with the power set of all possible hypotheses, and enables ignorance to be expressed explicitly and maintained throughout updating. Furthermore, DST does not fix hypothesis negation probability once occurrence probability is known.

The basics. Before we continue this discussion, it is only fair to warn readers that DST contains many unfamiliar definitions. As a result, the next few pages may make for slow reading.

Perhaps the most basic DST concept is the frame of discernment, Θ, defined as an exhaustive set of mutually exclusive events. In a simplified medical diagnosis on cholestatic jaundice,[36] for example, there may be four competing events: hepatitis (Hep), cirrhosis (Cirr), gallstones (Gall), and pancreatic cancer (Pan). Thus Θ has four elements. The role of Θ in DST resembles that of the sample space (Ω) in probability theory. The difference, however, is that in DST the number of possible hypotheses is $|2^\Theta|$, while in probability theory it is $|\Omega|$. In the above example, 2^Θ has 16 elements, representing all possible subsets of Θ. DST views observation of evidence against a hypothesis only as evidence supporting hypothesis negation. Thus, evidence disconfirming the hypothesis {Hep} (hepatitis and only hepatitis) is equivalent to evidence confirming the hypothesis {Cirr,Gall,Pan} (everything but hepatitis). (DST uses set notation when dealing with negation, different from the way it defines a hypothesis.) It does not, however, have a necessary impact on any other member of 2^Θ.

Let A be a subset of Θ. The basic probability number of A, denoted $m(A)$, is the probability assigned to the set A. The quantity $m(A)$ can be viewed as the portion of total belief assigned exactly to A. In many respects, this number can be treated like probability. The functions $p(A)$ and

$m(A)$ primarily differ in that A must be a singleton in probability theory, while A may contain several elements in DST. The function m maps the power set of Θ (2^{Θ}) to numbers between 0 and 1. It is called a basic probability assignment if it satisfies two properties: (1) the basic probability number of a null event is 0 ($m(\varnothing) = 0$); and (2) the sum of basic probability numbers for all subsets of Θ is 1 ($\Sigma_{A \subset \Theta} m(A) = 1$). In the above example, one possible basic probability assignment is $m(\{\text{Hep}\}) = 0.3$, $m(\{\text{Cirr}\}) = 0.2$, $m(\{\text{Gall}\}) = 0.1$, $m(\{\text{Hep,Cirr}\}) = 0.4$, and $m(A) = 0$ otherwise.

The belief of A, $Bel(A)$, measures the total amount of belief in A, not the amount assigned precisely to A by the basic probability assignment. Mathematically, this can be expressed as $Bel(A) = \Sigma_{B \subset A} m(B)$. The function Bel is called a belief function if it satisfies the following:

(1) The belief in a null hypothesis is 0, or $Bel(\varnothing) = 0$.

(2) The belief in Θ is 1, or $Bel(\Theta) = 1$.

(3) The sum of beliefs of A and $\neg A$ must be less than or equal to 1, or $Bel(A) + Bel(\neg A) \leq 1$. (The fact that $Bel(A) + Bel(\neg A) \leq 1$ is due to the definition of a hypothesis and its negation. In DST, A means "A and only A," while $\neg A$ means "everything but A.")

Thus, Bel equals m for singletons, but Bel is greater than or equal to m for sets that contain more than one element. For example, $Bel(\{\text{Hep}\}) = m(\{\text{Hep}\})$, but $Bel(\{\text{Hep,Cirr}\}) = m(\{\text{Hep,Cirr}\}) + m(\{\text{Hep}\}) + m(\{\text{Cirr}\}) \geq m(\{\text{Hep,Cirr}\})$.

The quantity $1 - Bel(\neg A)$ is called the plausibility of A ($Pl(A)$), providing the maximum amount of belief that can possibly be assigned to A. The functions Bel and Pl can be interpreted as the lower and upper probabilities induced by a multivalue mapping. Since $Bel(A) + Bel(\neg A) \leq 1$, $Pl(A) - Bel(A) \geq 0$.

In DST, only subsets of Θ with nonzero basic probability assignments are of interest, and each of these subsets is called a focal element of the belief function Bel over 2^{Θ}. The union of all the focal elements for a belief function is called its core. In the above example, given that $m(\{\text{Hep}\}) = 0.3$, $m(\{\text{Cirr}\}) = 0.2$, $m(\{\text{Gall}\}) = 0.1$, and $m(\{\text{Hep,Cirr}\}) = 0.4$, the core of the belief function is $\{\text{Hep,Cirr,Gall}\}$. Furthermore, $Bel(\{\text{Hep,Cirr}\}) = m(\{\text{Hep}\}) + m(\{\text{Cirr}\}) + m(\{\text{Hep,Cirr}\}) = 0.9$ is obtained from the basic probability numbers of the focal elements. The main difference between DST and probability theory is how probabilities are assigned to hypotheses. With the same set of possible hypotheses Θ, probability theory assigns probabilities to individual hypotheses, while DST assigns them to all possible subsets of Θ.

Table 2. An illustration of Dempster's combination rule.

M_1 \ M_2	{Cirr}(0.5)	{Hep,Cirr}(0.2)	Θ (0.3)
{Hep}(0.8)	∅(0.4)	{Hep}(0.16)	{Hep}(0.24)
Θ (0.2)	{Cirr}(0.1)	{Hep,Cirr}(0.04)	Θ (0.06)

Since DST development was partially motivated by the way that probability theory represents ignorance, DST expresses ignorance differently. In probability theory, uniform prior probability distributions represent complete ignorance. There is, however, no way to distinguish between instances of ignorance and instances in which known information suggests a uniform distribution. On the other hand, DST expresses ignorance explicitly. For example, if A and B are the only hypotheses, then probability theory would express ignorance about A and B with a probability of 1/2 for both A and B ($p(A) = p(B) = 1/2$). In DST, however, $m(\{A\}) = m(\{B\}) = 1/2$ indicates that the beliefs in A and B are the same, and no ignorance about their occurrences exists. The belief function for such a case is called a Bayesian belief function. Thus, if all the focal elements are singletons, then no ignorance exists regarding their occurrences; if any focal element contains more than one element, then some ignorance exists.

In probability theory, the probability of the negation of a hypothesis A is fixed once the probability of A is known, because $A \cup \neg A = \Omega$ and $p(A) + p(\neg A) = 1$. The same result in DST would require $Bel(A) + Bel(\neg A) = 1$, thereby implying that the belief in a hypothesis's negation is fixed once belief in the hypothesis is known. Hence, we cannot withhold belief from a proposition without increasing belief in its negation. No such restriction is present in DST; belief about the negation of a hypothesis does not depend on belief in the hypothesis itself, and the (weaker) restriction is $Bel(A) + Bel(\neg A) \leq 1$.

Combination of belief functions. When current evidence leads to multiple beliefs regarding the same hypothesis, the beliefs should be combined to provide an overall belief in the hypothesis. To propagate belief, DST usually combines different belief functions by computing their orthogonal sums with Dempster's rule of combination. Suppose that m_1 and m_2 are two basic probability assignments on a frame of discernment Θ. Then their orthogonal sum is

$$m_1 \oplus m_2(A) = K \sum_{X \cap Y = A} m_1(X) \times m_2(Y) \, ,$$

where K is a normalization constant ($K^{-1} = \Sigma_{X \cap Y \neq \varnothing} m_1(X) \times m_2(Y)$). If $A = \varnothing$, then by definition, $m_1 \oplus m_2(\varnothing) = 0$. If $K^{-1} = 0$, the orthogonal sum does not exist, and m_1 and m_2 are said to be totally contradictory. The quantity ($\log K$) is called the weight of conflict between Bel_1 and Bel_2. Thus, if Bel_1 and Bel_2 are not in conflict, $K = 1$; and if Bel_1 and

Bel_2 are totally contradictory, $K^{-1} = 0$. Orthogonal sums are both commutative and associative. Suppose two belief functions m_1 and m_2 are defined over a frame of discernment Θ, as shown in Table 2.[36] Then,

$$K = \frac{1}{1 - 0.4} = \frac{1}{0.6}$$

$$m_1 \oplus m_2 (\{\mathrm{Hep}\}) = \frac{0.16 + 0.24}{0.6} = 0.6667$$

$$m_1 \oplus m_2 (\{\mathrm{Cirr}\}) = \frac{0.1}{0.6} =0.1666$$

$$m_1 \oplus m_2(\{\mathrm{Hep,Cirr}\}) = \frac{0.04}{0.6} = 0.0667$$

$$m_1 \oplus m_2(\Theta) = \frac{0.06}{0.6} = 0.1$$

All other subsets of Θ have their combined belief equal to 0, and the sum of all the combined probability assignment for $m_1 \oplus m_2$ equals 1. For more combination examples using Dempster's rule, see Gordon and Shortliffe.[36] Since m_1 and m_2 represent two different belief functions on the same frame of discernment, combining the belief functions will provide a better picture of the belief of the hypotheses under consideration.

Analysis. One obvious problem with DST is its implementation complexity, inherent in the formalism — nearly all functions require exhaustively enumerating all possible subsets of the frame of discernment. However, by restricting the hypotheses of interest to mutually exclusive singletons and their negations, Barnett[37] has shown a linear-time algorithm for computing *Bel*, *m*, and other DST functions that enables rigorous applications of Dempster's rule.

If the frame of discernment's hypotheses are not mutually exclusive (such as the problem of multiple diseases in medical diagnosis), Θ may, of course, be redefined as the set of all possible subsets of all diseases. Unfortunately, the computational implications of this choice is horrifying — if there are 100 possible diseases, then $|\Theta| = 2^{100}$ and $|2^{\Theta}| = 2^{2^{100}}$. Since the evidence may actually focus on a small subset of 2^{Θ}, the computations need not always be intractable because DST need not explicitly enumerate subsets having zero belief.[36] Furthermore, the complexity of this problem can be reduced by using multiple frames of discernments, one for each set of mutually exclusive hypotheses.

DST lacks an effective decision procedure to draw inferences from belief functions — its main problem.[6,37] One can, however, use Dempster's rule for inference by having each belief function represent belief in a hypothesis given some evidence.[36] Shafer and Logan have proposed an efficient algorithm that computes degree of belief for more hypotheses and implements Dempster's combination rule for hierarchical evidence.[38]

Recently, some belief propagation schemes using extended DST have been proposed, and an experimental system named Gertis has been built.[39] However, no consensus exists on which scheme should be adopted as the (formal) belief propagation mechanism for DST. Thus, DST remains an unpopular approach for representing uncertainty in expert systems.

Summary. Dempster-Shafer theory enables ignorance to be expressed explicitly, and does not narrowly restrict belief in hypothesis negation once belief in its occurrence is known. However, these strong points lead to a combinatorial explosion because the hypothesis space is actually the power set of all possible hypotheses. To fill this huge space, human experts must provide all the beliefs on all subsets of possible hypotheses before an expert system can be built. Of course, the domain expert must provide basic probability assignments for the interesting subsets only, because all uninteresting subsets will have zero basic probability assignments. Although basic probability assignments can be computed based on the association between evidence and hypotheses, DST offers no guideline on how these assignments should be obtained. In addition, no effective procedure exists for drawing inferences from belief functions. Hence, it is not surprising that almost no expert system has ever been built using DST.

Possibility theory

Possibility theory was developed by Lotfi Zadeh as an extension of his theory of fuzzy sets.[10] The difficulties in representing inexact or vague information using probability theory motivated his research.

Background. In an expert system, the knowledge base contains a great deal of human knowledge, most of which is imprecise and qualitative. Often the boundary between competing hypotheses is vaguely defined. In expressing knowledge about such problems, human experts use terms such as "very likely" (or "probably") to describe occurrences of evidence and hypotheses (for example, "if the symptoms are . . ., then it is very likely that the disease is . . ."). When this type of expertise is encoded into probabilities, the "fuzziness" (or imprecision) is usually lost, and the idea is represented with specific (and frequently inaccurate) point values. Zadeh devised possibility theory to express these vague terms with precision and accuracy.

Possibility theory replaces the binary logic of probability with multivalued logic. In probability theory, either an event occurred or it did not. In possibility theory, shades of gray are permitted.

Apart from the imprecision in human knowledge, facts about the world are rarely known with certainty — another point in favor of possibility theory. Thus, these facts may only approximately match rule antecedents. Conventional rule-based systems usually evade this issue (or treat it in an ad hoc manner) because partial matching cannot be done with two-valued logic. Possibility theory makes partial matching natural and elegant by using compositional inference and interpolation.

The basics. Since possibility theory is based on the theory of fuzzy sets, we must begin by explaining fuzzy set terminology, key definitions, and properties. For a complete treatment, see Zadeh.[40]

Let U be a set of objects. A fuzzy set is a class of objects with a continuous grade of membership in U. Suppose that A is a fuzzy subset of U characterized by a membership function $\mu_A(u)$ which associates a real number in [0,1] with each element $u \in U$. The value of $\mu_A(u)$ represents the grade of membership of u in A, and the nearer the value of $\mu_A(u)$ to 1.0, the higher the grade of membership of u in A. Thus, the difference between a fuzzy set and a normal set (a set in which an object can either belong or not belong to the set) is the range of possible values for μ_A. In a normal set, the membership function can only take on values 0 and 1: $\mu_A(u) = 1$ denotes membership and $\mu_A(u) = 0$ denotes nonmembership.

These two theories also differ in fundamental number assignments, with no direct relationship between them — high possibility need not imply high probability, and vice versa. For example, if John can eat 1 to 3 eggs for breakfast, the possibility that John can eat 1, 2, 3 eggs may be set at 0.9, 1.0, 1.0, respectively. However, the probability that John will eat 1, 2, or 3 eggs on an arbitrary morning may be 0.1, 0.7, and 0.2.[10] Furthermore, no restriction exists on the sums of possibilities, while the sum of all probabilities must equal 1.

To further illustrate the properties of fuzzy sets, consider the following example:[10]

Let U be the set of integers, or $U = \{1,2,3,...\}$.
Let A be a fuzzy set of small numbers.
Then a subjective characterization of A could be:

u	1	2	3	4	5	6
$\mu_A(u)$	1.0	1.0	0.8	0.6	0.4	0.2

A fuzzy set is said to be empty if and only if its membership function is zero for all elements in U. Two fuzzy sets A and B are equal if and only if $\mu_A(u) = \mu_B(u)$ for all u in U. The complement of a fuzzy set A is denoted by $\neg A$, and its membership function is defined by $\mu_{\neg A} = 1 - \mu_A$. Other set operations also extend to their fuzzy counterparts: A fuzzy set A is said to be contained in a fuzzy set B (A is a subset of B) if and only if $\mu_A(u) \le \mu_B(u)$ for all $u \in U$. The union of two fuzzy sets A and B (with membership functions μ_A and μ_B, respectively) is a fuzzy set C, with $\mu_C(u) = \max[\mu_A(u), \mu_B(u)]$ for all $u \in U$. Similarly, the intersection of two fuzzy sets is another fuzzy set C, with $\mu_C(u) = \min[\mu_A(u), \mu_B(u)]$ for all $u \in U$. The fuzzy set operators $\cap$ and $\cup$ are both commutative and associative. The distributive and De Morgan's laws also hold for fuzzy sets.

Given these basics of fuzzy set theory, let's look at how to propagate belief. The paradigm that it suggests is known as possibility theory, which (like probability theory) deals mainly with distributions. Possibility distributions are directly related to fuzzy membership functions. If A is a fuzzy subset of U characterized by its membership function $\mu_A : U \to [0,1]$, the proposition "X is A" associates a possibility distribution Π_X such that $\Pi_X = A$ (also written as " X is $A \to \Pi_X = A$"). Correspondingly, the possibility distribution function associated with X, π_X, equals the membership function of A, that is, $\pi_X = \mu_A$.

Returning to the previous example of small integers, A can be rewritten as[10]

$$A = 1/1 + 1/2 + 0.8/3 + 0.6/4 + 0.4/5 + 0.2/6$$

where + denotes fuzzy set union, and the term 0.8/3 in the above expression signifies the 0.8 possibility that 3 is a small integer. Given these values for A, the proposition "X is a small integer" associates X with the possibility distribution $\Pi_X = A$. The possibility measure, $Poss\{x \in A\}$, is the possibility that a value x belongs to A, and is expressed as $Poss\{x \in A\} = \max_{u \in A}[\pi_X(u)]$. (If A is an infinite set, then $Poss\{x \in A\} = \sup_{u \in A}[\pi_X(u)]$. Thus, sup and inf will replace max and min for infinite sets.)

Since propositions can be both compounded and quantified (with terms like "very" or "not"), a theory designed to ease the transition from English to a more formal system must include rules for compounding and quantifying statements. Possibility theory deals with these modifications by the following set of rules:[10]

(1) Modifier rules. These rules define the impact on possibility distributions induced by English modifiers (for instance, Very, Not, and Rather) on a proposition of the form "X is A." The first step in defining these rules is to list all possible modifiers.

Consider, for example, a system in which the only modifiers allowed are Not, Very, and More-or-less. Some subjective analysis must be done to determine consistent functions assigned to these modifiers. Zadeh[8] has suggested squaring for Very, subtracting from one for Not, and taking the square root for More-or-less. Thus,

If X is $A \to \Pi_X = A$,
Then X is $mA \to \Pi_X = A^+$,

where A^+ is the modified possibility distribution induced by a legal modifier m. Thus, if m = Very, $A^+ = A^2$; if m = Not, $A^+ = \neg A = 1 - A$; if m = More-or-less, $A^+ = \sqrt{A}$.

(2) Composition rules. If both A and B are propositions, then $A \times B$ denotes a proposition that is a composition of A and B. The three most commonly used composition modes are logical And, logical Or, and logical Implication.

(a) Conjunction. Propositions are related by means of And.

If the proposition is
X is A And Y is B
Then
$\mu_{A*B}(u,v) = \min[\,\mu_A(u), \mu_B(v)]$

(b) Disjunction. Propositions are related by means of Or.

If the proposition is
X is A Or Y is B
Then
$\mu_{A+B}(u,v) = \max[\mu_A(u), \mu_B(v)]$

(c) Implication. Propositions are related by means of "If . . . Then . . .."

If the proposition is
If X is A Then Y is B
Then
$\mu_{B|A}(u,v) = \min[1, (1 - \mu_A(u) + \mu_B(v))]$

(3) Truth qualification rules. These rules define the possibility distribution modifications induced by truth quantifiers (for example, Very True, Quite True, and More-or-less True). A proposition "It is τ that X is A," where τ denotes the truth quantifier on the proposition, can be expressed as

X is A is $\tau \rightarrow \Pi_X = A^+$
and
$\mu_{A+}(u) = \mu_\tau(\mu_A(u))$

The definition of μ_τ is similar to the definition of English modifiers discussed above. For example, if τ = Very True, then $\mu_\tau(v) = v^2$.

As an example on how to use the above rules, consider the proposition: "If X is Not Very large And Y is More-or-less small, Then Z is Very Very large." Then, the conditional possibility distribution is described by

$$\mu_{(Z|X,Y)}(u,v,w) = \min[1, (1 - \min[1 - \mu^2_{large}(u), \mu_{small}(v)] + \mu^4_{large}(w))]$$

Evidence propagation. In possibility theory, inferences drawn upon observing new evidence are performed by means of generalized modus ponens.[10,41] In standard logic, modus ponens is a natural deduction rule stating that if hypotheses "$A \rightarrow B$" and "A" are both true, then hypothesis "B" is true. GMP differs from the classical version in two respects — matching is not required to be exact, and predicates are not required to be crisp (they can be fuzzy). For example, if A, A^*, B, and B^* are fuzzy statements where the operator can be an arbitrary modifying adjective, then GMP will come up with "Y is B^*" from the given premise and implication as follows:

Premise:	X is A^*.
Implication:	If X is A then Y is B.
Conclusion:	Y is B^*.

For instance,

Premise:	This tomato is very red.
Implication:	If a tomato is red, then the tomato is ripe.
Conclusion:	This tomato is very ripe.

From the definition of implication in composite rules, $\mu_{Y|X}(u,v) = \min[1,(1 - \mu_A(u) + \mu_B(v))]$ and $\Pi_X = A^*$. Using GMP, we conclude that "Y is B^*." Hence, the grade of membership of Y can be obtained from μ_A and μ_B as follows:

$$\mu_y(v) = \max_{u\in U}[\min[\mu_{A^*}(u),\mu_{Y|X}(u,v)]]$$
$$= \max_{u\in U}[\min[\mu_{A^*}(u),(1 - \mu_A(u) + \mu_B(v))]]$$

Moreover, other inference rules may also be extended to possibility theory. From "X is $A \rightarrow Y$ is B" and "Y is B^*," for example, we can conclude "X is A^*." This is called fuzzy modus tollens.

Expert system examples. Existing expert systems that use possibility theory to codify uncertainty are mostly rule-based systems in which the rules' premises and conclusions contain fuzzy quantifiers. Several such systems have been developed for medical diagnosis, because it is rarely possible to work with exact definitions, descriptions, and assertions in medical domains. The Cadiag-2 system[42] diagnoses rheumatic, hepatic, and pancreatic diseases and coagulation defects. It has a knowledge base of about 2500 symptoms and 308 diseases, expressed as a set of rules in which the premises and conclusions may contain linguistic variables and fuzzy terms. The system uses fuzzy inference to propagate and track belief. Sphinx,[43] an internal-medicine expert system, represents knowledge in both rules and trees, and manages uncertainty with fuzzy sets.

Fuzzy sets and possibility theory have also been used in other domains. Ishizuka et al.[44] have built a rule-based expert system for assessing structural damages to buildings. Their system uses both fuzzy sets and Dempster's rule for belief combination. In addition, fuzzy set programming languages (for example, Pruf) and expert system tools (for instance, Flops and SPII-1) are available.[41]

Analysis. Possibility theory enables the world's inherent fuzziness to be represented explicitly and easily. Nevertheless, this approach does have its share of problems. Cheeseman[45] points out that the strong "truth compositionality" property of fuzzy sets creates inconsistency. Truth composition refers to the requirement that the truth measure associated with a logical compound is strictly a function of its component. One example is the fuzzy set "min" rule ($\mu_{A*B}(u,v) = \min[\mu_A(u), \mu_B(v)]$). This rule, however, is only reasonable in the special case of mutual exclusivity, and is not true in general. Wise and Henrion[46] point out that the "min" rule is equivalent to probabilistic inference by assuming maximum correlation between events A and B — if A and B are mutually exclusive events, the possibility that they will occur simultaneously should be zero, but the "min" rule may not recognize it.

Of course, many other difficulties with possibility theory remain. One stems from the interpretation and definition of fuzzy quantifiers. The subjective analysis that was used to generate functional relationships ($A^+ = \sqrt{A}$ if m = More-or-less, or $A^+ = A^2$ if m = Very) may enhance knowledge base consistency, but does not necessarily increase its accuracy. Nuances are frequently lost because similar linguistic terms are assigned the same function. Furthermore, possibility theory lacks formal semantics. In an attempt to resolve this difficulty, Giles[47] suggested an interpretation for basic terms (including grade of membership and possibility) and logical operations (including And, Or, and Not), but he also admitted that further work is needed.

In addition to these problems, a long-standing debate persists concerning the necessity of fuzziness theories; Cheeseman[45] and Stallings[48] argue that probability theory addresses all the proper issues, while Zadeh[49] disagrees. Nevertheless, fuzzy sets and possibility theory have attracted, and continue to attract, a great deal of attention, research, and controversy.

Other methods

Probability theory, Dempster-Shafer theory, and fuzzy sets were all developed before expert systems became popular. When expert system research began in earnest, researchers attempted to develop uncertainty management schemes specifically tailored to knowledge-intensive rule-based systems. In the early 1970's, Shortliffe developed the certainty factor approach — among the first and most notable of these schemes — to represent uncertain information in MYCIN.[1] Another expert-system-specific scheme, Prospector's subjective Bayesian method was developed by Duda et al.[12] in the 1970's, motivated by the difficulty of representing Prospector's uncertain information with other numeric paradigms. In the 1980's, Paul Cohen developed the theory of endorsements[13] — the primary non-numeric paradigm for uncertainty management — as an attempt to represent uncertain information qualitatively, rather than quantitatively.

Certainty factors. The probability-theoretic approach to uncertainty management requires a prodigious amount of data. Hence, some weakly substantiated approximations and assumptions are usually used to reduce the requisite number of probability assessments. When Shortliffe began work on MYCIN, he felt that probability theory would not be appropriate to medicine because of the following:[29]

- Often, not enough good data exists for a particular medical problem to create a statistical knowledge base.
- Medical knowledge and the heuristics for solving medical problems must be explicitly represented. Probability cannot easily accomplish this.
- Explaining the line of reasoning is very important if physicians are to accept the program.

In view of these potential problems, Shortliffe proposed a new approach called certainty factors.[11]

The basics. In an expert system using certainty factors, the knowledge base consists of a set of rules in the following form: "If <evidence> Then (CF) <hypothesis>," where CF denotes hypothesis belief given observed evidence.

Before any combination or propagation of evidence can be performed, two intermediate functions must be calculated. These functions, $MB[h,e]$ and $MD[h,e]$, measure the degrees to which belief in hypothesis h would be increased if e were observed, and the degree to which disbelief in h would be increased by observing the same evidence e, respectively. They are defined as follows:

$$MB[h,e] = \begin{cases} 1 & \text{if } p(h) = 1 \\ \dfrac{\max[p(h|e),p(h)] - p(h)}{\max[1,0] - p(h)} & \text{otherwise} \end{cases}$$

$$MD[h,e] = \begin{cases} 1 & \text{if } p(h) = 0 \\ \dfrac{\min[p(h|e),p(h)] - p(h)}{\min[1,0] - p(h)} & \text{otherwise} \end{cases}$$

The values of $MB[h,e]$ and $MD[h,e]$ range between 0 and 1. If more than one piece of evidence supports a hypothesis, a combination function for MB and MD (indicating the total strength of hypothesis belief and disbelief) is used in computing CF (see Shortliffe and Buchanan for more information[11]). Evidence is then propagated by computing CF from MB and MD, where

$$CF = \frac{MB - MD}{1 - \min[MB,MD]}$$

The value of CF can range from -1 to +1; -1 indicates the confirmation of h's negation, and +1 indicates h's confirmation.

When two or more rules affect the same hypothesis, the individual *CF*s obtained from these rules are combined to give a combined *CF* for the hypothesis.

$$CF_{combine}(X,Y) = \begin{cases} X + Y \times (1 - X) & \text{if both } X, Y>0 \\ \dfrac{X+Y}{1 - \min[|X|,|Y|]} & \text{one of } X, Y<0 \\ -CF_{combine}(-X,-Y) & \text{if both } X, Y<0 \end{cases}$$

Van Melle[1] suggested this $CF_{combine}$ function as an improvement to the original one — it avoids the possibility that a single piece of strong evidence against a hypothesis will cancel out the effects of several strong pieces of supporting evidence (and vice versa).

Expert system examples. Several expert systems have been built using the certainty factor approach, including MYCIN, a diagnostic program for infectious blood disease, and VM,[50] a ventilator-monitoring program. Moreover, the expert system shell EMYCIN uses MYCIN's framework and has no specific domain knowledge.[51] Puff, built using EMYCIN, interprets pulmonary function tests and has been ported to minicomputers for routine use in hospitals.[52] S.1 is a newer expert system shell based on EMYCIN's framework and uncertainty representation, and is commercially available to expert system builders.[3]

Analysis. Ever since certainty factors were introduced in MYCIN, they have been used to represent uncertainty in rule-based expert systems. At the same time, their theoretical aspects have been investigated. Adams found that some unstated (or implicit) assumptions made by certainty factors may not always be valid, such as the assumption of independence among hypotheses. He concluded that the "weakness of certainty factors is the inobvious interdependence restriction placed on parameter estimation by assumptions of independence."[53]

In a recent study, Heckerman showed that the original definition of certainty factors is inconsistent with its combination function (for example, the combination function for certainty factors is not commutative — the belief in a hypothesis given evidence E_1 and E_2 will depend on the order of update).[54] He further provided a probabilistic interpretation for certainty factors, and showed that the new interpretation provides useful insight into necessary and sufficient conditions for propagating certainty factors through an inference network. He also suggested methods for relaxing the model's implicit assumptions that are rarely valid in practice. Grosof pointed out that Heckerman's revised certainty factor is a special case of Dempster-Shafer theory.[55] Clancey analyzed the sensitivity of MYCIN's decisions to certainty factors.[1] By modifying the number of values that *CF* can take in MYCIN rules, Clancey showed that the program is not very sensitive to *CF* changes. He further showed that by turning off the *CF* engine, MYCIN can still provide correct diagnoses to most test cases — indicating that a rule's contents are more important than the certainty factors associated with it.

One of the most attractive features of the certainty factor model is that it represents, combines, and propagates the effects of multiple sources of evidence in terms of joint beliefs or disbeliefs in each hypothesis. Since the model was originally designed only to represent change in belief induced by evidence, rather than an absolute degree of belief,[56] certainty factors avoid the need for prior probabilities, thereby addressing one of the more contentious challenges hurled at probabilistic belief propagation. Heckerman and Horvitz, however, showed that certainty factors cannot represent some particular classes of dependencies among uncertain beliefs efficiently and naturally.[57] They further contended that AI researchers will find belief networks an expressive representation for capturing the complex dependencies associated with uncertain knowledge.

Prospector's subjective Bayesian method. One geological expert system, Prospector, relied on an uncertainty management scheme based on Bayesian decision theory and on Bayes' rule. Instead of using the prior and conditional probabilities directly, as a more standard subjective-probability-theoretic system might, it used "odds-likelihood" functions. In this method, an inference net can represent graphically the collection of rules specifying all knowledge.[12] Each rule takes the form: "If <evidence> Then <hypothesis> (*LS*,*LN*)," where *LS* measures support favoring a hypothesis and *LN* measures the support against it. For example,[3] "If the region is a hypabyssal region environment, Then the region has a favorable level of erosion (200,0.0002)." Basically, this rule strongly supports the hypothesis "favorable erosion" if the region is a hypabyssal environment (*LS* = 200), and it also strongly opposes the hypothesis if the region is not a hypabyssal region environment (*LN* = 0.0002).

The advantage that likelihood ratios have over prior and conditional probabilities is that human experts do not have to provide "exact" probability assessments. The standard reliance on perfect encoding creates difficulties for human experts; people feel far more comfortable providing likelihood ratios. Since the original prior and conditional probabilities can be recovered from the two likelihood ratios (*LS* and *LN*), this approach can use Bayes' rule to propagate evidence.

The basics. The two numbers associated with each rule are called likelihood ratios, representing a rule's perceived "strength." The likelihood ratio of observing *E* given that *H* is true, $LS(H,E)$, is defined as the ratio of $p(E|H)$ to $p(E|\neg H)$. This formula looks suspiciously like a representation of odds. In general, the odds of a proposition *A* is given by $p(A)/p(\neg A)$. In terms of hypotheses and evidence, the prior odds of *H*, $O(H)$, is the ratio of prior probabilities $p(H)/p(\neg H)$, while the posterior odds of H given E, $O(H|E)$, is the ratio of posterior probabilities $p(H|E)/p(\neg H|E)$. When we combine these definitions, the relationship between odds and likelihood ratios becomes

$$O(H|E) = LS(H,E) \times O(H)$$

This equation tells us how to update H's odds, given that E was observed. In a rule-based system, a domain expert provides likelihood ratios. High LS values ($LS \gg 1$) indicate that observing E would greatly increase the posterior probability H (in other words, the rule supports the hypothesis).

Similarly, the likelihood ratio of observing $\neg E$ given H, $LN(H,E)$, is defined as the ratio of $p(\neg E|H)$ to $p(\neg E|\neg H)$. In an analogous manner,

$$O(H|E) = LN(H,E) \times O(H)$$

Note that LN cannot be derived from LS — the domain expert must provide it independently. Low LN values ($0 \leq LN \ll 1$) decrease the posterior probability of H given $\neg E$. Moreover, if we exclude the extreme cases that $p(E|\neg \mathrm{H})$ is 0 or 1, it can be observed that ($LS > 1$) implies ($LN < 1$), and ($LS < 1$) implies ($LN > 1$).

Evidence is combined using the following two rules:

(1) Conjunctive. The probability of observing n pieces of evidence is the same as the probability of witnessing the least likely one among them, or

$$p(E_1 \wedge E_2 \wedge \ldots \wedge E_n) = \min[p(E_1), p(E_2), \ldots, p(E_n)]$$

(2) Disjunctive. The probability of observing at least one of a set of n pieces of evidence is the same as the probability of witnessing the most likely one of them, or

$$p(E_1 \vee E_2 \vee .. \vee E_n) = \max[p(E_1), p(E_2), \ldots, p(E_n)]$$

Belief is propagated by computing the posterior odds of a hypothesis (H) given the observation of some evidence (E). Using posterior odds, we can recover posterior probability $p(H|E)$ by $O(H|E)/(1 + O(H|E))$. If H supports G, then the ratio of increase in prior probability of H to the original prior probability of $\neg H$ is multiplied by the LS of the rule "If H Then G (LS,LN)." We can use the resulting value to compute the posterior odds on G. The following example shows how this scheme works.[58] In Figure 2, (20,1) represents the LS and LN values in the rule "If Rcib Then Smir (LS,LN)." Now suppose the user indicates that Rcib is present. This evidence increases the odds of Smir by a factor of 20, thereby raising its probability from 0.03 to 0.382. (The prior odds on Smir is 0.03/(1 - 0.03) = 0.030927, giving the posterior odds on Smir equal to 20 × 0.030927 = 0.61855, which corresponds to a probability of 0.61855/(1 + 0.61855) = 0.382.) This increases the odds on Hype by a factor of 300, weighted by the degree to which Smir has increased from its prior probability, by a factor 300 × (0.382 - 0.03) / (1 - 0.03) = 108.866. Hence, the posterior probability of Hype is 0.52373. Propagation continues in this manner "upwards" through the network. If, however, the evidence (Rcib in the above example) is not known to be definitely present or absent, then a linear interpolation between the two extremes will be used for inferencing.[12]

Examples of expert systems. To use this approach, the knowledge base must satisfy many assumptions, including the conditional independence of evidence under both a hypothesis and its negation. Since these assumptions are rarely satisfied, few systems have been built with the subjective Bayesian approach. Prospector, developed for mineral exploration, is the best known expert system employing this approach for handling uncertain information.[58] Expert system shells Kas[59] and AL/X[60] are commercially available. The accounting expert system Auditor[61] was built using AL/X.

Analysis. To use Prospector's subjective Bayesian approach, several assumptions (explicit and implicit) must be satisfied — primarily that of conditional independence. All pieces of evidence are assumed explicitly to be conditionally independent given a hypothesis; all pieces of evidence are assumed implicitly to be conditionally independent given the negation of a hypothesis. Glymour,[62] Johnson,[63] and Pednault et al.[64] have studied these assumptions independently. Their investigations concluded that the restriction added by assuming conditional independence between pieces of evidence and the negation of hypotheses renders the scheme effectively useless (because only very few real-world problems can satisfy both assumptions).

Furthermore, many expert systems that used Prospector's scheme overlooked the constraints underlying the mathematical properties of LS and LN; consequently, these systems performed poorly.[65]

Theory of endorsements. This qualitative approach represents uncertainty as a body of richly structured knowledge about situations. In most real-world settings, the "strength of evidence" is actually a summary of several factors; numerical approaches inevitably summarize all supporting and opposing evidence into a single number. Since an intelligent reasoner usually discriminates among these factors, the summary represented by numeric paradigms is inadequate. Furthermore, the semantics of the numbers that represent knowledge about uncertain information are often unclear. The theory of endorsements is therefore centered around the idea of dealing with reasons for believing (or disbelieving) a hypothesis. By making explicit the knowledge about uncertainty and evidence that would otherwise be summarized in a number, endorsements show how to reason with the knowledge directly, rather than indirectly (that is, through a numerical calculus).

The basics. The primary motivation underlying the theory of endorsements is that knowledge about uncertain situations should influence system behavior. An important step towards this goal is resolution task design — deciding what to do when required evidence is lacking. A resolution task is one whose execution reduces uncertainty by obtaining more information. Consonant with this goal is the criterion that uncertainty should affect control decisions, an objective achieved by determining whether a task would contribute anything as a prerequisite for allocating resources to that task. Furthermore, when an endorsements-based system schedules tasks on its agenda, it does so by asking what it expects them to contribute to its certainty. Thus, these control decisions may not solve uncertain problems, but they do exploit knowledge about uncertainty to facilitate resolutions.

All the reasons for believing (or disbelieving) a hypothesis are represented in structures called endorsements. Endorsements are frame-like knowledge structures representing reasons to believe (positive endorsements) or disbelieve (negative endorsements) a particular hypothesis. They are associated with propositions and inference rules at various times during reasoning. Five endorsement classes are important for reasoning about uncertainty in rule-based systems:[13]

- **Rule endorsements.** Reasons to believe (or disbelieve) inference rules (for example, a clause in a premise is endorsed as maybe-too-restrictive when the premise might occasionally fail due to this clause when the conclusion is in fact valid).
- **Data endorsements.** Reasons to believe (or disbelieve) raw data (for example, a statement about one's own risk tolerance is often conservative).
- **Task endorsements.** Arguments about the evidence that executing tasks are likely to produce, used for scheduling tasks (for example, a task is worth doing because it may produce a corroborating conclusion).
- **Conclusion endorsements.** Reasons to believe (or disbelieve) conclusions. These are combinations of a priori rule endorsements and detected relationships, such as corroboration between conclusions (for example, a conservative conclusion about one's risk tolerance is corroborated by other evidence).
- **Resolution endorsements.** Records of applying methods to resolve uncertainty (for example, no rules conclude a desired goal, but after eliminating a maybe-too-restrictive clause from a rule, the desired conclusion is achieved).

The reasoning performed by these endorsements resembles the goal-directed reasoning of many expert systems. Thus, it may start out by trying to conclude a goal, and then backchain through its rule base. As reasoning proceeds, new bodies of endorsements develop from reasons to believe (or disbelieve) its conclusions. These new endorsements provide justifications for the conclusion. It is important that endorsements affect agenda control because the theory of endorsements is oriented towards the effects of uncertainty on behavior. These effects are two-fold. First, the system uses endorsements to decide whether a proposition at hand is certain enough, asking whether the endorsements of a subgoal conclusion were good enough to warrant using the conclusion to assert its parent goal. Second, the system uses resolution tasks to reduce uncertainty by obtaining more information through endorsements. The principle of these tasks is that negative endorsements are viewed as problems to be solved, thus providing the reasoning required to lessen uncertainty.

In addition to rules for deciding when a proposition is certain enough for a task and rules for resolving uncertainty, the theory of endorsements also has a simple rule for combining endorsements and propagating them over inferences. A conclusion inherits all its premise's endorsements, plus any that result from posting the conclusion (such as a contradiction between this conclusion and another). This rule, however, has two problems — reasons to believe or disbelieve a premise are not always endorsements of the conclusion, and the rule may lead to large bodies of endorsements after only a few inferences. Cohen and his group at the University of Massachusetts are currently exploring new methods, such as semantic combination rules for endorsements, that address these and other related problems.[13]

Examples of expert systems. Since the theory of endorsements has only recently entered the field of expert system uncertainty management, few existing systems use it. This approach was initially tested in Folio[66] using expert heuristics from the portfolio management domain. The Solomon system[13] reasons about the uncertainty associated with heuristics and their use. Cohen points out that Solomon is actually not an expert system, but is intended eventually as an "E-system" (shell) in the style of EMYCIN. In addition, development environments and systems for reasoning with uncertainty in medicine, such as Mum and Mu, are also available.[67,68]

Analysis. The theory of endorsements, the first qualitative paradigm for uncertainty management, has been developed only recently. Although combining evidence and ranking propositions are important in controlling inference, these operations are not easy with the theory of endorsements — for example, the simple rule outlined earlier has the serious problem that in only a few steps, a large body of endorsements can be created. This problem inhibits the scaling of programs using this theory into realistic settings. Nevertheless, the theory of endorsements does offer one major advantage over numeric paradigms — the line of reasoning and the semantics of endorsements generated during inference are both explicit and clear.

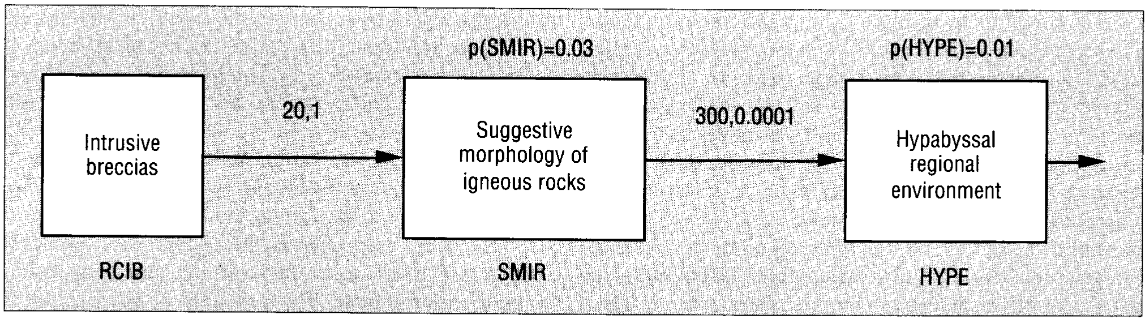

p(SMIR)=0.03
p(HYPE)=0.01
Intrusive breccias
20,1
Suggestive morphology of igneous rocks
300,0.0001
Hypabyssal regional environment
RCIB
SMIR
HYPE

Figure 2. A simple example of evidence propagation in Prospector.

Comparison of methods

The previous four sections outlined the popular paradigms for uncertainty management in expert systems. Comprehensive examinations of the various formalisms are available.[4,69] Subjective probability theory, as the oldest and best established of the group, has attracted the most attention. Nevertheless, many researchers have found Bayesian methods of belief propagation somewhat lacking and inefficient. This dissatisfaction, in turn, motivated the design of the newer theories.

Not surprisingly, these newer paradigms appear to outperform subjective probability theory in the specific areas for which they were designed: Dempster-Shafer theory represents ignorance better, possibility theory does a better job dealing with fuzzy information, and certainty factors give better explanations of the control flow through a rule-based system.

Nothing comes free, however. Both DST and possibility theory suffer from the use of unfamiliar terminology, and they lack formal semantics and implementational clarity. Certainty factors and Prospector's approach require too many assumptions that are not true in general. Many proponents of probability have made rather bold claims about the theory's universality — Cheeseman[19] and Pearl[30] have both claimed that probability theory can be applied to any problem. Lindley said that anything that can be done with fuzzy logic, belief functions, or other alternative techniques can be done better with probability.[8] Needless to say, other researchers disagree.

Recently, Horvitz et al. proposed a framework that can be used in identifying the fundamental differences between probabilistic and non-probabilistic methodologies, and used this framework to compare several non-probabilistic schemes (fuzzy set theory, Dempster-Shafer theory, and certainty factors).[70] Wise and Henrion compared the performance of different schemes (subjective probability theory, fuzzy set theory, certainty factors) using the same set of rules and data.[46] In their tests, Bayesian networks outperformed competitors while fuzzy sets and certainty factors, although not on par with that of the Bayesian approach, performed comparably to each other. More recently, Heckerman empirically compared three scoring mechanisms (Bayes' theorem, odds likelihood, and Dempster-Shafer theory), and concluded that the Bayesian approach is best.[71]

Although the various uncertainty management schemes are different from one another, they do share some common problems. A human expert willing to quantify expertise is needed when implementing any theory. Finding an expert able to accurately quantify personal, subjective, and qualitative information, however, is no mean feat. It has been observed that humans are easily biased,[72] and thus the quality of the knowledge extracted from experts depends greatly on the method used for assessment. Nevertheless, expert system researchers have expended surprisingly minimal effort on studying and deriving appropriate assessment techniques. Most of the work done on these (for example, see our references[73]) has gone relatively unnoticed by the AI community.

The choice of uncertainty management scheme is strongly dependent on existing conditions and domains. If an expert is available for serious decision-analytic conversations, probability theory is likely to be the most appropriate — at least in the opinion of the authors. In the absence of such expertise, however, probability assignments may be too arbitrary to be meaningful. An expert system whose knowledge base is constructed primarily from textual material may need to propagate information using one of the more complicated paradigms.

Expert system designers must face one reality — large knowledge bases cannot be searched quickly; the best they can hope for is relative efficiency. Detractors of all these methods have hurled the derogatories of "inefficient" and "intractable" at the targets of their venom. At this point, probability theory appears to be the most efficient general-case method, although instances exist in which some of the others can outperform it. Even so, Bayesian belief propagation techniques are of exponential complexity, and are thus impractical for sufficiently large databases.

In short, all of the approaches described in this article have their particular strengths and weaknesses. However, many interesting open problems remain whose resolutions should lead to powerful, general, and implementable techniques for coping with uncertainty management in expert systems.

This survey has examined the merits and demerits of popular paradigms for representing and managing uncertain information in expert systems. Only some of the problems mentioned here are currently under investigation; others either have gone unnoticed or have been placed on permanent hold by the research community. The following lists (in no particular order) significant issues that should be investigated further:

- **Application of techniques from other fields.** Researchers in other fields have explored similar issues discussed in the expert system community. Very little effort, however, has been expended in incorporating known techniques from other fields into AI systems. Researchers in decision analysis, game theory, operations research, financial forecasting, psychology, and management science have all investigated issues related to uncertainty and decision making. Which of these techniques are applicable to uncertainty management in AI?
- **Consistency checking.** Relatively little has been done in consistency checking on the knowledge base. It is very important that a knowledge base is consistent, because inconsistent knowledge bases may provide inconsistent conclusions. Is it possible to perform consistency checking on knowledge bases mechanically? If so, how?
- **Consensus of knowledge bases.** It is frequently desirable to have a knowledge base derived from several experts, rather than from a single expert. How can weights be assigned to different experts so as to accurately reflect their expertise? What can be done if different knowledge bases are represented with different representation schemes? What can be done if the viewpoints of different experts are contradictory?
- **Knowledge refinement.** Expert system designers face the problem of knowledge obsolescence in a knowledge base. This problem could be effectively solved if an automatic means of incorporating new knowledge into the system could be devised. Methods on refining and enhancing knowledge bases have been proposed,[74,75] but they work only with rule-based systems. Are there refinement methods for other representation schemes?
- **Assessment methods.** All the popular paradigms share the common problem of human bias and subjectivity. Can knowledge assessment be automated? Can it be accomplished without the intervention of a decision analyst or a knowledge engineer? Do ways exist to improve the quality of knowledge assessments? Can it be done more efficiently?

With luck, the next few years will see significant progress on one or more of these issues. The answers to these questions should lead to more powerful AI programs, both in the area of expert systems and elsewhere.

Acknowledgments

Research for this article was supported in part by the National Library of Medicine under grant R01LM04529, and in part by a grant from the USC Faculty Research Initiation Fund.

We would like to thank Ward Edwards, Thomas Eppel, David Heckerman, Eric Horvitz, and John Yen for providing useful feedback on this manuscript.

References

1. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, B.G. Buchanan and E.H. Shortliffe, eds., Addison-Wesley, Reading, Mass., 1984.

2. R.A. Miller, H.E. Pople, Jr., and J.D. Myers, "Internist-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine," *New England J. Medicine*, Vol. 307, No. 8, 1982, pp. 468-476.

3. D.A. Waterman, *A Guide to Expert Systems*, Addison-Wesley, Reading, Mass., 1986.

4. J.Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, Palo Alto, Calif., 1988.

5. P.P. Bonissone and R.M. Tong, "Reasoning with Uncertainty in Expert Systems," *Int'l J. Man-Machine Studies*, Vol. 22, No. 3, 1985, pp. 241-250.

6. R.K. Bhatnagar and L.N. Kanal, "Handling Uncertain Information: A Review of Numeric and Non-Numeric Methods," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier North-Holland, New York, N.Y., 1986, pp. 3-26.

7. B. de Finetti, *Theory of Probability*, John Wiley & Sons, New York, N.Y., 1974.

8. D.V. Lindley, "The Probability Approach to the Treatment of Uncertainty in AI and Expert Systems," *Statistical Science*, Vol. 2, No. 1, 1987, pp. 17-24.

9. G.Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, N.J., 1976.

10. L.A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems*, Vol. 1, No. 1, 1978, pp. 3-28.

11. E.H. Shortliffe and B.G. Buchanan, "A Model of Inexact Reasoning in Medicine," *Mathematical Biosciences*, Vol. 23, Nos. 3/4, 1975, pp. 351-379.

12. R.O. Duda, P.E. Hart, and N.L. Nilsson, "Subjective Bayesian Methods for a Rule-Based Inference System," *Proc. Nat'l Computer Conf.*, Vol. 45, 1976, pp. 1075-1082.

13. P.R. Cohen, *Heuristic Reasoning about Uncertainty: An AI Approach*, Pitman, Boston, Mass., 1985.

14. I.J. Good, "Kinds of Probability," *Science*, Vol. 129, No. 3347, 1959, pp. 443-447.

15. I. Hacking, *The Emergence of Probability*, Cambridge Univ. Press, New York, N.Y., 1975.

16. L.J. Savage, *The Foundations of Statistics*, Dover Publications, New York, N.Y., second ed., 1972.

17. D.A. Schum, *Evidence and Inference for the Intelligence Analyst*, University Press of America, Lanham, Md., 1987.

18. L.J. Cohen, *The Probable and the Provable*, The Clarendon Press, Oxford, U.K., 1977.

19. P. Cheeseman, "In Defense of Probability," *Proc. Ninth IJCAI*, Morgan Kaufmann, Palo Alto, Calif., 1985, pp. 1002-1009.

20. W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, New York, N.Y., 1957.

21. D. von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research*, Cambridge Univ. Press, New York, N.Y., 1986.

22. F.T. de Dombal et al., "Computer-Aided Diagnosis of Acute Abdominal Pain," *British Medical J.*, Vol. 2, Apr. 1972, pp. 9-13.

23. J.C. Horrocks and F.T. de Dombal, "Computer-Aided Diagnosis of 'Dyspepsia'," *The American J. Digestive Diseases*, Vol. 20, No. 5, 1975, pp. 397-406.

24. D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani, "Update on the Pathfinder Project," *Proc. 13th Symp. Computer Applications in Medical Care*, IEEE Computer Society Press, Los Alamitos, Calif., 1989, pp. 203-207.

25. R.A. Howard and J.E. Matheson, "Influence Diagrams," in *The Principles and Applications of Decision Analysis*, R.A. Howard and J.E. Matheson, eds., Strategic Decision Group, Menlo Park, Calif., 1984, pp. 721-762.

26. R.D. Shachter, "Evaluating Influence Diagrams," *Operations Research*, Vol. 34, No. 6, 1986, pp. 871-882.

27. A.M. Agogino and A. Rege, "IDES: Influence Diagram-Based Expert System," *Mathematical Modeling*, Vol. 8, 1987, pp. 227-233.

28. P. Szolovits and S.G. Pauker, "Categorical and Probabilistic Reasoning in Medical Diagnosis," *AI*, Vol. 11, Nos. 1-2, 1978, pp. 115-144.

29. E.H. Shortliffe, B.G. Buchanan, and E.A. Feigenbaum, "Knowledge Engineering for Medical Decision Making: A Review of Computer-Based Clinical Decision Aids," *Proc. IEEE*, Piscataway, N.J., Vol. 67, No. 9, 1979, pp. 1207-1224.

30. J. Pearl, "How to Do with Probabilities What People Say You Can't," *Proc. Second Conf. AI Applications*, IEEE Computer Society Press, Los Alamitos, Calif., Dec. 1985, pp. 6-12.

31. G.F. Cooper, "The Computational Complexity of Probabilistic Inference using Belief Networks," Tech. Report KSL-87-27, Stanford Univ., Stanford, Calif., 1987.

32. M.R. Garvey and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, Calif., 1979.

33. S.L. Lauritzen and D.J. Spiegelhalter, "Local Computations with Probabilities on Graphical Structures and their Applications to Expert Systems," *J. Royal Statistical Society*, Series B, Vol. 50, No. 2, 1988, pp. 157-224.

34. E.J. Horvitz, J.S. Breese, and M. Henrion, "Decision Theory in Expert Systems and AI," *Int'l J. Approximate Reasoning*, Vol. 2, No. 3, 1988, pp. 247-302.

35. A.P. Dempster, "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals Math. Statistics*, Vol. 38, No. 2, 1967, pp. 325-339.

36. R.A. Gordon and E.H. Shortliffe, "A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space," *AI*, Vol. 26, No. 3, 1985, pp. 323-357.

37. J.A. Barnett, "Computational Methods for a Mathematical Theory of Evidence," *Proc. Seventh IJCAI*, Morgan Kaufmann, Palo Alto, Calif., 1981, pp. 868-875.

38. G. Shafer and R. Logan, "Implementing Dempster's Rule for Hierarchical Evidence," *AI*, Vol. 33, No. 3, 1987, pp. 271-298.

39. J.Yen, "Gertis: A Dempster-Shafer Approach to Diagnosing Hierarchical Hypotheses," *Comm. ACM*, Vol. 32, No. 5, 1989, pp. 573-585.

40. L.A. Zadeh, "Fuzzy Sets," *Information and Control*, Vol. 8, No. 3, 1965, pp. 338-353.

41. H.J. Zimmermann, *Fuzzy Sets, Decision Making, and Expert Systems*, Kluwer Academic Publishers, Boston, Mass., 1987.

42. K.-P. Adlassnig and G.Kolarz, "Cadiag-2: Computer-Assisted Medical Diagnosis using Fuzzy Subsets," in *Approximate Reasoning in Decision Analysis*, M.M. Gupta and E.Sanchez, eds., Elsevier North-Holland, New York, N.Y., 1982, pp. 219-247.

43. M. Fieschi et al., "Sphinx: An Interactive System for Medical Diagnosis Aids," in *Approximate Reasoning in Decision Analysis*, M.M. Gupta and E. Sanchez, eds., Elsevier North-Holland, New York, N.Y., 1982, pp. 269-275.

44. M. Ishizuka, K.S. Fu, and J.T.P. Ya, "A Rule-Based Inference with Fuzzy Set for Structural Damage Assessment," in *Approximate Reasoning in Decision Analysis*, M.M. Gupta and E. Sanchez, eds., Elsevier North-Holland, New York, N.Y., 1982, pp. 261-268.

45. P. Cheeseman, "Probabilistic vs. Fuzzy Reasoning," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier Science Publishers, New York, N.Y., 1986, pp. 85-102.

46. B.P. Wise and M. Henrion, "A Framework for Comparing Uncertain Inference Systems to Probability," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier Science Publishers, New York, N.Y., 1986, pp. 69-83.

47. R. Giles, "Semantics for Fuzzy Reasoning," *Int'l J. Man-Machine Studies*, Vol. 17, No. 4, 1982, pp. 401-415.

48. W. Stallings, "Fuzzy Set Theory versus Bayesian Statistics," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 7, No. 3, 1977, pp. 216-219.

49. L.A. Zadeh, "Is Probability Theory Sufficient for Dealing with Uncertainty in AI: A Negative View," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier Science Publishers, New York, N.Y., 1986, pp. 103-116.

50. L.M. Fagan et al., "Extensions to Rule-Based Formalism for a Monitoring Task," in *Rule-Based Expert Systems*, B.G. Buchanan and E.H. Shortliffe, eds., Addison-Wesley, Reading, Mass., 1984, pp. 397-423.

51. W. van Melle, E.H. Shortliffe, and B.G. Buchanan, "EMYCIN: A Knowledge Engineer's Tool for Constructing Rule-Based Expert Systems," in *Rule-Based Expert Systems*, B.G. Buchanan and E.H. Shortliffe, eds., Addison-Wesley, Reading, Mass., 1984, pp. 302-313.

52. J.S. Aikins et al., "Puff: An Expert System for Interpretation of Pulmonary Function Data," *Computers and Biomedical Research*, Vol. 16, 1983, pp. 199-208.

53. J.B. Adams, "A Probability Model of Medical Reasoning and the MYCIN Model," *Math. Biosciences*, Vol. 32, Nos. 1/2, 1976, pp. 177-186.

54. D.E. Heckerman, "Probability Interpretation for MYCIN's Certainty Factors," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier Science Publishers, New York, N.Y., 1986, pp. 167-196.

55. B.N. Grosof, "Evidential Confirmation as Transformed Probability: On the Duality of Priors and Updates," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier, New York, N.Y., 1986, pp. 153-166.

56. E.J. Horvitz and D.E. Heckerman, "The Inconsistent Use of Measures of Certainty in AI Research," in *Uncertainty in AI*, L.N. Kanal and J.F. Lemmer, eds., Elsevier Science Publishers, New York, N.Y., 1986, pp. 137-152.

57. D.E. Heckerman and E.J. Horvitz, "On the Expressiveness of Rule-Based Systems for Reasoning under Uncertainty," *Proc. Sixth Nat'l Conf. AI*, Morgan Kaufmann, Palo Alto, Calif., 1987, pp. 121-126.

58. R.O. Duda, J. Gaschnig, and P.E. Hart, "Model Design in Prospector Consultant System for Mineral Exploration," in *Expert Systems in the Micro-Electronic Age*, D. Michie, ed., Edinburgh Univ. Press, Edinburgh, U.K., 1979, pp. 153-167.

59. *Building Expert Systems*, F. Hayes-Roth, D.A. Waterman, and D.B. Lenat, eds., Addison-Wesley, Reading, Mass., 1983.

60. A. Paterson, *AL/X User Manual*, Intelligent Terminals Ltd., Oxford, U.K., 1981.

61. C.W. Dungan and J.S. Chandler, "Auditor: A Microcomputer-Based Expert System to Support Auditors in the Field," *Expert Systems*, Vol. 2, No. 4, 1985, pp. 210-221.

62. C. Glymour, "Independence Assumptions and Bayesian Updating," *AI*, Vol. 25, No. 1, 1985, pp. 95-99.

63. R.W. Johnson, "Independence and Bayesian Updating Methods," *AI*, Vol. 29, No. 2, 1986, pp. 217-222.

64. E.P.D. Pednault, S.W. Zucker, and L.V. Muresan, "On the Independence Assumption underlying Subjective Bayesian Updating," *AI*, Vol. 16, No. 2, 1981, pp. 213-222.

65. D.E. O'Leary and N.A. Kandelin, "Validating the Weights in Rule-Based Systems," *Int'l J. Expert Systems*, Vol. 1, No. 3, 1988, pp. 253-279.

66. P.R. Cohen and M.D. Lieberman, "A Report on Folio: An Expert Assistant for Portfolio Managers," *Proc. Eighth IJCAI*, Morgan Kaufmann, Palo Alto, Calif., 1983, pp. 212-215.

67. P.R. Cohen et al., "Management of Uncertainty in Medicine," Tech. Report 86-12, Univ. of Massachusetts, Amherst, Mass., 1986.

68. P.R. Cohen, M. Greenberg, and J. DeLisio, "Mu: A Developed Environment for Prospective Reasoning Systems," Tech. Report 87-46, Univ. of Massachusetts, Amherst, Mass., 1987.

69. D.A. Schum, "Probability and the Process of Discovery, Proof, and Choice," Boston Univ. Law Review, Boston, Mass., Vol. 66, Nos. 3-4, 1986, pp. 825-876.

70. E.J. Horvitz, D.E. Heckerman, and C.P. Langlotz, "A Framework for Comparing Alternative Formalisms of Plausible Reasoning," *Proc. Fifth Nat'l Conf. AI*, 1986, pp. 210-214.

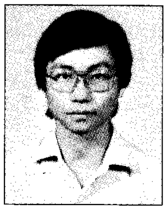
71. D.E. Heckerman, "An Empirical Comparison of Three Scoring Schemes," *Proc. Fourth Workshop Uncertainty in AI*, 1988, pp. 158-169.

72. *Judgement under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, eds., Cambridge Univ. Press, New York, N.Y., 1982.

73. C.S. Spetzler and C-A.S. Stäel von Holstein, "Probability Encoding in Decision Analysis," *Management Science*, Vol. 22, No. 3, 1975, pp. 340-358.

74. L.-M. Fu, *Learning Object-Level and Meta-Level Knowledge in Expert Systems*, PhD thesis, Stanford Univ., Stanford, Calif., 1985.

75. A. Ginsberg, *Automatic Refinement of Expert Knowledge Bases*, Morgan Kaufmann, Palo Alto, Calif., 1988.



Keung-Chi Ng is a doctoral student in computer science at the University of Southern California, where he received his MS in computer science in 1987. He received his BS in electrical engineering (with honors) from the University of Hong Kong in 1980. His research interests include expert systems, knowledge acquisition, decision analysis, and uncertainty representation. His dissertation concerns the use of multiple experts in an expert system. He is a student member of IEEE, AAAI, and ACM.

Bruce Abramson received his BA in 1983, his MS in 1985, and his PhD in 1987 (all in computer science) from Columbia University. While pursuing his doctorate, he spent two years in residence at UCLA. Since 1987, he has been assistant professor of computer science at the University of Southern California. He is currently working on AI decision, inference, diagnosis, and forecasting systems, and is a member of the IEEE, AAAI, ACM, and IIF.

The authors can be reached at the Computer Science Department, University of Southern California, Los Angeles, CA 90089.