AI Magazine Volume 4 Number 2 (1983) (© AAAI)

Techniques and Methodology

METHODOLOGICAL SIMPLICITY IN EXPERT SYSTEM CONSTRUCTION: The Case of Judgments and Reasoned Assumptions

Jon Doyle

Department of Computer Science
Carnegie-Mellon University
Pittsburg, PA 15213

Editors' Note: Many expert systems require some means of handling heuristic rules whose conclusions are less than certain Baysian techniques and other numerical scoring methods have been developed to combine and propagate certainty measures as the expert system draws inferences in solving different problems.

Doyle's paper argues that it is difficult for a human expert to produce reliable probabilities or numerical scoring factors for an inference rule, and that a radically different approach to the problem should be considered He essentially suggests that the expert be encouraged to think in terms of specific instances which would conflict with the general rule and to encode this knowledge explicitly.

Methodologically this seems to be very appealing, and helps to make both explicit and rigorous some of the techniques currently used by knowledge engineers when they encode and refine the expert's knowledge We would welcome comments and criticisms of this approach from those steeped in the practical issues of constructing large rule-based expert systems. — *Derek Sleeman and Jaime Carbonell*

Abstract

Probabilistic rules and their variants have recently supported several successful applications of expert systems, in spite of the difficulty of committing informants to particular conditional probabilities or "certainty factors," and in spite of the experimentally observed insensitivity of system performance to perturbations of the chosen values Here we survey recent developments concerning reasoned assumptions which offer hope for avoiding the practical elusiveness of probabilistic rules while retaining theoretical power, for basing systems on the information unhesitatingly gained from expert informants, and reconstructing the entailed degrees of belief later

I owe much to Gerald Sussman, Johan de Kleer, Guy Steele, Drew McDermott and Marvin Minsky for inspiration on these topics I also thank Joseph Schatz, Peter Szolovits, Randall Davis, Donald Kosy, and Mark Fox for valuable discussions, and Jaime Carbonell, John McDermott, and Edward Shortliffe for suggesting substantive improvements to this paper This research was supported by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597, monitored by the Air Force Avionics Laboratory under Contract F33615-81-K-1539 The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Government of the United States of America

The "Probability" Problem

Recent successes of "expert systems" stem from much hard work of designers and experts in elicting, encoding and examining the normally tacit rules of reasoning employed by the experts. These three tasks of elicting, encoding, and examining rules of reasoning influence each other. The designer must be able to encode the rules elicted, the encoding must facilitate examination of the rules in operation and the sorts of information elicted must seem important to the expert lest examination seem pointless. Much current practice employs

so-called judgmental rules for encoding the informant's rules of reasoning Judgmental rules, for example those in MYCIN, incorporate both propositional and "probabilistic" information. An investment advisor system, for instance, might be given a rule

IF: 1. The client's income bracket is 50%,
and 2. The client carefully studies market trends,
THEN: 3. There is evidence (0.8) that the investment should be in high-technology.

(from Davis, 1979) and would use the rule to draw conclusions whose "certainty factors" depend on the observed certainty factors of the hypotheses (lines 1 and 2) and the certainty factor (0.8) of the rule itself. Though few expert systems actually treat these numerical grades of certainty as Bayesian probabilities and conditional probabilities, their interpretation usually approximates that of Bayesian probabilities, namely as subjective degrees of belief. Although simple, these judgmental rules have supported development of many impressive applications.

In spite of the fruitfulness of this approach, its practitioners express discomfort with several of its requirements. One difficulty is that while it is relatively easy to elict tentative propositional rules from experts and from people in general, it is considerably harder to get commitment to particular grades of certainty That is, one's human informant might quickly suggest the propositional part of the above rule "IF 1 and 2, THEN 3" but might try to avoid assigning 0.8 or any other number to the rule. Worse still, individual informants frequently vary in their answers to a repeated question depending on the day of the week, their emotional state, the preceding questions, and other extraneous factors. This further aggravates the sensitivity of answers to the phrasing of questions noted by Tversky and Kahneman (1981).

Another difficulty stems from these. Noticing their informants' hesitancy, system designers test the sensitivity of the system's performance to the set of numbers used. Reported experiments show the numbers do not actually mean exactly what they seem to mean, for the performance of most systems remains constant under all sorts of small ($< 30\%$) perturbations in the precise values used. Understandably, expert system designers have difficulty justifying their use of the numerical judgments in face of these indications of psychological and pragmatic unreality. Unfortunately, they have had to stick to their guns, since no satisfactory alternative has been apparent.

One hope for giving expert system designers what they want is to investigate their practice, to seek simpler but equipotent sorts of information and operations which capture the "grain of truth" in the probabilistic approach. Something might explain both the designers' intuitions and their systems' successes. Since that something is clearly not standard Bayesian probability theory or its certainty factor variants, can we find what really underlies current practice?

This article answers "Yes." Recent developments concerning reasoned assumptions suggest an approach which allows familiar sorts of rules of reasoning, judgments of certainty of conclusions, and more besides — all without the unreality of the probabilistic approach. In the following, we explain the modified approach together with its practical and theoretical attractions. Related discussions of Bayesianism can be found (Minsky, 1975; Shortliffe & Buchanan, 1975; Duda, Hart, & Nilsson, 1976; Szolovits, 1978; Szolovits & Pauker, 1978).

Reasoned Assumptions

The approach of reasoned assumptions modifies the standard treatment of uncertainty, not the treatment of propositional information Thus in the investment rule above, we change only the certainty factor information, and leave the propositional information intact. Although our approach accommodates representational systems based on "frames" and "units" as easily as systems based on "assertions," for expository simplicity we pretend the database consists of elementary logical sentences, and leave restatement of the approach in one's favorite representational system as an exercise for the reader. For terminological simplicity, we label all the numerical approaches as "Bayesianism," and use MYCIN as our standard of comparison Strictly speaking, MYCIN is not Bayesian, but the differences are immaterial in the following discussion.

We abbreviate the parts of IF-THEN rules by writing A to mean the set of antecedent sentences of the IF part, and by writing C to mean the set of conclusion sentences of the THEN part If the rule simply relates concrete (ground) sentences, A and C are sets of concrete sentences If the rule expresses general or schematic information in terms of variables $\vec{x} = x_1, \ldots, x_n$, we can indicate this generality by writing $A(\vec{x})$ and $C(\vec{x})$ instead. We also write $\neg A$ to mean the set of negations of statements in A, that is, $\neg A = \{\neg a \mid a \in A\}$.

We express uncertain rules of reasoning in expressions of the form

$$A \,||\, B \,|\!\!-\, C$$

and

$$\forall \vec{x} \quad [A(\vec{x}) \,||\, B(\vec{x}) \,|\!\!-\, C(\vec{x})].$$

We read the former as "A without B gives C," an expression informally interpreted as "conclude every sentence in C if every sentence in A has been concluded and no sentence in B has been concluded." We interpret the latter expression as a schema implying all concrete instances of the form $A(\vec{g}) \,||\, B(\vec{g}) \,|\!\!-\, C(\vec{g})$ for ground terms $\vec{g}$ We call these expressions *reasons*, and the conclusions derived from them *reasoned assumptions*. We connect reasons with IF-THEN rules by noting that A and C act as antecedents and consequents as before, and that B contains qualifications on the inference

Thus, ignoring uncertainty, we can rewrite the propositional part of "IF A, THEN C" as $A \parallel \emptyset \mid\!\!\sim C$.

Reasoned assumptions express uncertainty in terms of the non-statistical notions of *typicality* and *defeasibility* (English for "liability to defeat"), notions concerned with the preferences of the agent about how to adopt assumptions in order to resolve ambiguities in its information about the problem. Rules of typicality express the usual, normal or initial conclusions for consideration, conclusions which may be individually defeated if circumstances warrant. (See Doyle (1982) for more discussion of these notions.) Completely certain inference rules can be written as $A \parallel \emptyset \mid\!\!\sim C$, in which case the absence of qualifying statements means the conclusions are always drawn from the antecedents. Rules stating the normal or usual conclusions following from antecedents may be written either as *default reasons* or as *defeasible reasons*. Default reasons are of the form $A \parallel \neg C \mid\!\!\sim C$, meaning that the conclusions are inferred from the antecedents only if their negations are not already known. Defeasible reasons are of the form $A \parallel \{Defeated(R)\} \mid\!\!\sim C$, where R is the name of the reason itself. Thus $R = [A \parallel \{Defeated(R)\} \mid\!\!\sim C]$ means that the conclusions are inferred from the antecedents only if the use of the reason has not been ruled out by the presence of the statement $Defeated(R)$ in the database. Default and defeasible reasons are usually used in concert with other reasons expressing special cases, exceptions and other overriding conditions. If a particular application (instantiation) of a schematic reason produces unwanted conclusions, we defeat the particular application, not the schematic reason itself. In other words, we ordinarily correct errors on a case-by-case basis. If the overriding reasons cover most circumstances, the overridden reasons serve as "catch-all" rules, guidelines for what to do about "everything else" not covered by the specific case reasons. Moreover, independently formulated default reasons sometimes conflict on instances, and so may require conflict resolution reasons defeating one instance in favor of another in ambiguous circumstances. (See Reiter (1978) and Reiter & Criscuolo (1981) for examples.)

Comparison of the Approaches

One cannot simply reformulate probabilistic rules as reasons according to their certainty factors To re-express a database of expertise, we require the knowledge acquisition process carried a bit further than usual. The approach of reasoned assumptions supposes that numerical judgments of certainty often hide more specific information not yet made explicit by the expert informant. When the expert says that "To degree 0.3, IF A, THEN C," this really means that many exceptional cases are familiar to the expert. One might ask the informant to list these exceptions as a set B, in order to qualify the rule by writing it as $A \parallel B \mid\!\!\sim C$, but it is often as difficult to think of exceptions offhand as it is to think of ordinary heuristic rules. Instead, we apply the same technique to articulating expertise as that already practiced, namely the informant expresses what is clear, and then formulates and reformulates the missing cases, exceptions, and generalizations by repeatedly examining the system's performance on test problems. At bottom, we always have rules of the form "Usually, IF A, THEN C" or "Usually, IF A, THEN $\neg C$," which we express as defeasible or default reasons, and we express the intermediate degrees of uncertainty by case analysis and reasons stating exceptions to generalities. This of course requires more work in articulating expertise than the probabilistic approach, since one may have to formulate several cases and conflict resolution reasons that could be hidden in a single number, but in the long run, improving the performance of a probability-based system requires the same sort of case and conflict analyses. That is, use of probabilities may make the initial database smaller, but by the time expert performance has been molded from the initial approximation, about the same information should be present in the one approach as in the other. Since the same hard questions must be addressed in either case, and since the non-probabilistic rules can be had with less hesitation, we conclude the total work of the reasoned assumptions approach should not exceed that of the probabilistic approach.

Our assessment of the relative amounts of work needed to formulate a body of expertise assumes that the two sorts of encodings can express the same information. While the mathematical details are inappropriate for pursuit here, recent theoretical treatment of reasoned assumptions shows how subjective probabilities or certainty factors may be derived from sets of reasons, a project reminiscent of Savage's (1972) construction of quantitative subjective probabilities from qualitative subjective probabilities. The converse derivation seems unlikely, so from a purely theoretical viewpoint, the information expressed in reasons may be more powerful or fundamental than that expressed in probabilistic rules We briefly sketch these ideas.

MYCIN's probabilistic rules are interpreted by computing the "deductive closure" of the rules together with problem specific information. This results in a single probability distribution on all statements of interest.

In contrast, sets of schematic and problem specific reasons are interpreted by finding their *admissible extensions*. The admissible extensions $AExts(S)$ of a set S of reasons are "closed and grounded" supersets E of S. "Closed" means that if $A \parallel B \mid\!\!\sim C$ is in E, and if every element of A is in E, and if no element of B is in E, then every element of C is in E. "Grounded" means that every statement in E either is in S or is supported by a noncircular argument from S and qualifiers not in E. (See Doyle (1982) for the precise formulation.) We reconstruct subjective probabilities from admissible extensions by using a probabilistic algorithm to derive admissible extensions from the initial set of reasons. Probabilistic algorithms compute perfectly definite structures, in our case some $E \in AExts(S)$, but make deliberately randomized choices whenever there is more than one construction possible Probabilities enter into our observations of the computation, but not into the

computation itself. We apply this idea by supposing that all admissible extensions with the same number of elements are equally likely to be derived from the initial set of reasons, and that the probability of derivation decreases with the number of elements in the admissible extension. Specifically, we suppose $\mathrm{pr}(E \mid S)$, the probability of deriving extension E from S, to be proportional to $2^{-|E|}$. The "degree of belief" $\mathrm{pr}(a \mid S)$ of a statement a given S is then the normalized probability of deriving admissible extensions containing a. That is, if

$$N = \sum_{E \in AExts(S)} 2^{-|E|},$$

we can define

$$\mathrm{pr}(E \mid S) = \frac{2^{-|E|}}{N},$$

and

$$\mathrm{pr}(a \mid S) = \sum_{a \in E \in AExts(S)} \mathrm{pr}(E \mid S).$$

For example, if S is the set of reasons

$$\emptyset \parallel \{\neg c_1\} \mid\!\!\vdash \{c_1\}$$

$$\emptyset \parallel \{c_1\} \mid\!\!\vdash \{\neg c_1\}$$

$$\{c_1\} \parallel \{\neg c_2\} \mid\!\!\vdash \{c_2\}$$

$$\{c_1\} \parallel \{c_2\} \mid\!\!\vdash \{\neg c_2\}$$

then S has three admissible extensions

$$E_1 = S \cup \{\neg c_1\}$$

$$E_2 = S \cup \{c_1, c_2\}$$

$$E_3 = S \cup \{c_1, \neg c_2\}$$

so that $\mathrm{pr}(c_1 \mid S) = \frac{1}{2}$, $\mathrm{pr}(\neg c_1 \mid S) = \frac{1}{2}$, $\mathrm{pr}(c_2 \mid S) = \frac{1}{4}$, and $\mathrm{pr}(\neg c_2 \mid S) = \frac{1}{4}$. If the dependence of c_2 and $\neg c_2$ on c_1 is removed, we have instead four admissible extensions

$$S + \{c_1, c_2\}, \{c_1, \neg c_2\}, \{\neg c_1, c_2\}, \{\neg c_1, \neg c_2\}$$

so that statements in S have probability 1 and $c_1, \neg c_1, c_2, \neg c_2$ each have probability $\frac{1}{2}$.

The preceding suggests, both practically and theoretically, that reasoned assumptions express a more fundamental notion of uncertainty than numerical stipulations of certainty. One incompleteness in the preceding comparison concerns objective probabilities. It may not always be reasonable to expect that all probabilities may be analyzed into cases. The probabilities may be the results of formal physical measurements of phenomena, or informal observations by the informant of past experience with success of rules or reliability of data. Moreover, even if such probabilities might ultimately succumb to analysis, their analysis may cost too much.

We cannot yet offer a complete solution to this difficulty. While one obvious approach is to assume the more likely outcome as a default and expect to work to correct the outcome when wrong, another possibility is to combine reasons and objective probabilities by making the probabilistic extension-computing algorithm sensitive to statements of fundamental objective probabilities. Unfortunately, this combination has not yet been adequately explored.

Even if a workable combination of reasoned assumptions and objective probabilities exists, this need not sanction continued reliance on purely probabilistic systems. The approach of reasoned assumptions has other attractions for knowledge acquisition, attractions which recommend probabilities only as a tool of last resort. We mention two.

One of the most important requirements placed on reasoning systems by knowledge acquisition is that of explicability of conclusions, since informants must see how the system uses its information and arrives at its conclusions in order to criticize and correct it. Most expert systems based on probabilities keep track of the applications of rules used in computing derived probabilities, and so can explain the probability of a conclusion in terms of its computation. Similarly, the reasoned assumption approach explicitly involves a notion of explanation in the groundedness requirement on admissible extensions. But the two sorts of explanations are not of comparable power since computational histories need not be illuminating explanations of conclusions. This is an *important* difference, for explanations in terms of applied rules simply explain numbers with more *numbers*, while explanations in terms of assumptions explain unlikely conclusions in terms of the *assumptions* needed to get them, or from the critical viewpoint, in terms of their possible counterexamples. Compared with the probability-combination rules used in Bayesianism, explicit computation of admissible extensions presents the informant with more fine-grained examples and counterexamples, explanations which more readily guide formulation of special cases, exceptions, and conflict resolution reasons. While both approaches permit explanation, those of reasoned assumptions are more useful.

Another attraction of reasoned assumptions is that they facilitate additive updating of databases and supply "audit trails" pinpointing the history of rule revisions. If the informant decides a rule needs changing in MYCIN, the rule must be deleted and replaced by the informant in cooperation with a bookkeeping system (We need not consider the grotesque alternative of adding an identical rule with the equally certain but contrary conclusion.) In contrast, with defeasible reasons, the required bookkeeping is explicitly part of the system, so that the informant can indicate reasoned retractions and replacements of faulty reasons, with the history of database changes made explicit in the reasons themselves. For example, if the informant decides to base a revision on a further hypothesis about the problem situation, he need not replace the faulty reason alone, but can make the hypothesis

explicit in the qualifiers of the new reason and in the defeater of the old reason, so that the old reason will again be used when circumstances void the informant's hypothesis.

Conclusion

Compared with numerical judgments of subjective probability, reasoned assumptions offer both closer correspondence with the expertise easily obtainable in practice, and comparable theoretical power. This suggests exploring expert systems based on reasoned assumptions instead of resting content with current systems. A variety of practicable systems for manipulating and interpreting reasons currently exist for use, but they do not yet implement facilities for computing degrees of belief, facilities which may be necessary for summarizing the structure of large sets of admissible extensions as well as for quantifying confidence levels Considerable theoretical work seems needed to identify those notions of admissible extensions and derived probabilities which may be feasibly computed, and this may temporarily temper the attractions of reasoned assumptions Nevertheless, the likely practical rewards seem to justify the pursuit.

Whatever the ultimately preferred treatment of uncertainty, I hope this article encourages further analysis of practice aiming for increased methodological simplicity. More than in some periods of its history, current artificial intelligence involves many people building on the work of others, especially by taking existing tools and doing something with them rather than devising speculative inventions for each new application. But one can also build on the achievements of others by analyzing their experience to extract the essence of their ideas and results, rather than simply assuming that the first successful formulation is best In such circumstances, one analyzes demonstrably practicable ideas and techniques, not to patch them, but to rethink, understand, and possibly improve them in fundamental ways. I believe much of artificial intelligence theory and practice is ripe for this sort of analysis, not just its treatments of uncertainty. In many other areas, artificial intelligence offers numerous alternative, seemingly incomparable theories Are there further "grains of truth" awaiting discovery which might connect and explain these alternatives?

References

Davis, R , 1979 Interactive transfer of expertise: acquisition of new inference rules, *Artificial Intelligence* 12, 121-157

Doyle, J., 1982. Some theories of reasoned assumptions: an essay in rational psychology, Pittsburgh: Department of Computer Science, Carnegie-Mellon University.

Duda, R , Hart, P., and Nilsson, N , 1976. Subjective Bayesian methods for rule-based inference systems, *Readings in Artificial Intelligence* (B. L Webber and N J. Nilsson, eds.), Palo Alto: Tioga (1981), 192-199.

Minsky, M., 1975 A framework for representing knowledge, *The Psychology of Computer Vision* (P Winston, ed.), New York: McGraw-Hill.

Reiter, R , 1978. On reasoning by default, *Proc. Second Conf. on Theoretical Issues in Natural Language Processing*, 210-218

Reiter, R , and Criscuolo G , 1981 On interacting defaults, *Proc Seventh International Joint Conference on Artificial Intelligence*, 270-276

Savage, L. J , 1972. *The Foundations of Statistics*, 2nd rev. ed , New York: Dover

Shortliffe, E H , and Buchanan, B G , 1975. A model of inexact reasoning in medicine, *Mathematical Biosciences* 23, 351-379

Szolovits, P , 1978 The lure of numbers: how to live with and without them in medical diagnosis, *Proc Coll Computer-Assisted Decision Making using Clinical and Paraclinical (Laboratory) Data* (B E. Statland and S. Bauer, eds), Tarrytown: Technicon, 65-76.

Szolovits, P , and Pauker, S G , 1978 Categorical and probabilistic reasoning in medical diagnosis, *Artificial Intelligence* 11, 115-144

Tversky, A , and Kahneman, D., 1981 The framing of decisions and the psychology of choice, *Science* 211, 453-458