COMPONENT PART NOTICE

THIS PAPER IS A COMPONENT PART OF THE FOLLOWING COMPILATION REPORT:

TITLE: Workshop on Assessing Uncertainty Held in Monterey, California on November 13-14, 1986.

TO ORDER THE COMPLETE COMPILATION REPORT, USE AD-A180 850.

THE COMPONENT PART IS PROVIDED HERE TO ALLOW USERS ACCESS TO INDIVIDUALLY AUTHORED SECTIONS OF PROCEEDING, ANNALS, SYMPOSIA, ETC. HOWEVER, THE COMPONENT SHOULD BE CONSIDERED WITHIN THE CONTEXT OF THE OVERALL COMPILATION REPORT AND NOT AS A STAND-ALONE TECHNICAL REPORT.

THE FOLLOWING COMPONENT PART NUMBERS COMPRISE THE COMPILATION REPORT:

AD#: P005 291 thru P005 301 AD#: ______

AD#: ______ AD#: ______

AD#: ______ AD#: ______

APPROXIMATING BELIEF FUNCTIONS IN A RULE-BASED SYSTEM

WILLIAM F. EDDY

Department of Statistics,

Carnegie-Mellon University

Pittsburgh, PA USA

November 2, 1986

Revised, December 1, 1986

To appear in:

The Proceedings of the First International Conference on Statistical Computing

1. INTRODUCTION

Rule-based expert systems have moved from a research activity in a small number of academic computer science departments to a growing commercial activity. This transition clearly indicates that the structure of a complex computer program enforced by a rule-based system (namely, the clear separation of the decision-making process, *the inference engine*, from the data on which the decisions are based, *the rule base*) is a useful step in the evolution of programming strategies. At the same time there has been a growing recognition that in most decision-making situations the data (namely, the rule base and the initial evidence used to start the decision-making process) are not known with certainty and consequently the inference procedures used in traditional rule-based systems are inappropriate. Over the last decade a number of inference procedures which use various numerical representations of uncertainty have been developed for use in rule-based systems. However, for a variety of reasons (including the fact that there is little logical basis for the representations) none of them has been widely successful.

In this paper we will describe the current state of an ongoing research project which is attempting to use probability as the mechanism for representing uncertainty in a rule-based system. A previous report was given in Eddy and Pei (1986). We have been constrained in our development of a probability-based expert system by a number of external considerations, the most important of which we delineate here.

The single most important constraint is that we are doing our development in the context of an existing rule-based expert system and as such we are constrained to limit the modifications we might wish to make to the system. In particular, we wish to limit our modifications to the inference engine only. This is not an overly serious constraint and it enforces a certain *locality* on the nature of the possible computations. Exactly this locality of computation will be required if the system is ever to be scaled up to a rule-base containing thousands or millions of rules. Barnett (1981) has induced the same kind of locality of computation in a system very similar to ours but at the cost of assuming unrealistic independence in various parts of the rule base: we discuss Barnett's work further in Section 5.

A second important constraint is that any numerical expressions of uncertainty about data are themselves quite uncertain, in practice, and as such we wish to allow for the expression of uncertainty about the uncertainties. There are a number of possible ways to do this: we have chosen what appears to us the simplest way to address this constraint. Precisely, we are using belief functions (Shafer, 1976) to represent sets of probability distributions. We discuss some of the details of belief functions in Section 4.

There are at least two parties involved in the development and use of a rule-based expert system: the *expert*, who expresses the rules, and the *user*, who causes the system to run by supplying it with some initial external evidence. An early planned use of the system we are developing was for game-playing to evaluate possible strategies. Initially we felt that it was important for the two players, the expert and the user, to be able to interchange roles without affecting the results. This turns out to be a quite complex constraint: a simplified version of this constraint requires that the system perfrom properly (i.e., get the "right" answer) if the expert and the user are one and the same individual.

The remainder of this paper is organized as follows. In the next section we give a very brief introduction to the details of a rule-based expert system. In Section 3 we describe what, to us, are the most natural methods for incorporating uncertainty into a rule-based system. In Section 4 we provide a few of the formal details of belief functions and describe a few of their properties. In Section 5 we discuss various possible approximation techniques which will speed up the computations. In Sections 6 and 7 we provide detailed properties of the approximations we have developed. The essence of the approximations is to force the belief function to have a simpler form: an extreme form of this simplification occurs when the belief function represents a unique probability distribution. In Section 8 we briefly describe some of the features of our implementation of this theory in a LISP computer program.

2. RULE-BASED EXPERT SYSTEMS

A rule-based expert system (or production system) consists of a collection of production rules together with a system for linking or "chaining" the rules to simulate a human expert's reasoning process. A production rule (or, simply, a rule) is a statement of the form "If A then B," where A and B are logical propositions.

The mechanism used for chaining rules is generally one of two kinds: either *forward* chaining or *backward* chaining. In the forward chaining scheme the user of the system supplies some evidence, generally of the form "A is true," and the system then uses this evidence together with the rules to reason towards conclusions or goals. Forward chaining is generally described as *causal* or *deductive* reasoning. In the backward chaining scheme the system attempts to satisfy its goals by finding rules which, if true, would imply those goals. It repeats this process until it is compelled, by the lack of any rules implying its current goals, to ask the user if a particular one or more of those goals (the antecedents of certain rules) are true. If the user accedes this is deemed to be evidence that the rule is true. Backward chaining is generally referred to as *diagnostic* reasoning. One crucial computational problem in either form of reasoning is how to discover rules with given antecedents (forward chaining) or with given consequents (backward chaining) in the rule base. Currently the only general strategy is to search over the entire rule base. Some savings can be made by "remembering" the results of previous searches so they can be "looked up" in a table.

3. PROBABILITIES AND RULES

There is currently no generally accepted method for incorporating uncertainty into a rule-based expert system. One method which seems appealing at first glance is to treat the user's probabilities on the evidence as a prior opinion and the expert's probabilities on the rules as a likelihood and simply use Bayes rule. In this method, we would expect the expert constructing the system to have joint probability distribution on the assignment of truth values to the propositions which are consequents of all the rules in the system. This joint probability distribution would be conditional on the assignment of truth values to those propositions in the system which are antecedents of some rule and not consequents of any rule (the evidence nodes). Also, we would expect the user of the system to have a joint probability distribution on the assignment of truth values to these evidence nodes. There are a number of obvious difficulties with this scheme:

1. It is unreasonable to expect anyone to express a joint probability distribution on the assignment of truth values to a large collection of propositions for two reasons:

    a. The size of the collection of propositions;

    b. The inherent uncertainty in the expressed probability distribution.

2. The amount of calculation required is overwhelming, being exponential in the number of propositions in the system.

3. The symmetry constraint mentioned in the introductory section is obviously not satisfied.

There are also a few subtler problems:

1. The pooling of expert and user opinion via Bayes rule would appear to be inappropriate. More precisely, use of Bayes rule to pool the probability distributions of two individuals has no logical basis unless one of the individuals declares the probability distribution of the other individual to be his own.

2. Both the expert and the user can reasonably be expected to have a joint probability distribution on the assignment of truth values to all the propositions in the system. The use of lower-dimensional marginal and conditional probability distributions from these two higer-dimensional joint distributions appears to discard potentially useful information.

A second method for incorporating uncertainty into a rule-based expert system pools the opinions of the expert and the user. We would expect the expert constructing the system to have a joint probability distribution on the assignment of truth values to all the propositions in the system and we would expect the user to also have such a joint probability distribution. This second method can, by the appropriate choice of a pooling rule, satisfy the symmetry constraint mentioned above.

If it is possible to decompose each joint probability distribution so that a piece of the decomposition can be attached to a small number of propositions, and if this piece can be combined with another piece so that the entire joint distribution can be recovered then the difficulty of assigning a joint probability distribution on the assignment of truth values to a large collection of propositions may be overcome. One such decomposition is the conditional one; it would be desireable to have a decomposition that is symmetric so the order of composition is unimportant. Although Spiegelhalter (1986) has proposed a mechanism for allowing the conditional decomposition to be symmetric.

We also allow the expression of uncertainty about probabilities by use of a belief functions as a lower bound on the probability. This will allow us to alleviate the first of the three obvious difficulties mentioned above. It does not seem possible to significantly reduce the computational requirements mentioned in the third difficulty; however, in Sections 5, 6, and 7 we discuss an approximation which provides some reduction in the computational burden (see Eddy and Pei, 1986, for an alternative scheme).

4. BELIEF FUNCTIONS

Following Shafer (1976), let Θ be a set of mutually exclusive and exhaustive propositions. Let 2^Θ be the set of all subsets of Θ; elements of 2^Θ can be interpreted as general propositions in the problem domain. A basic probability assignment is a function $m(\cdot)$ from 2^Θ into [0, 1] which satisfies

$0 \leq m(\cdot) \leq 1,$

$m(\phi) = 0,$

and

$\Sigma_{\text{S contained in } \Theta}\, m(S) = 1.$

There is a one-to-one correspondence between this basic probability assignment and the belief function, Bel(·), and plausability function, Pl(·), given by

$\mathrm{Bel}(S) = \Sigma_{\text{T contained in S}}\, m(T),$

$m(S) = \Sigma_{\text{T contained in S}}\, (-1)^{|S-T|}\mathrm{Bel}(T),$

and

$\mathrm{Pl}(S) = 1 - \mathrm{Bel}(\bar{S}).$

It is apparent that $\mathrm{Bel}(A) \leq \mathrm{Pr}(A) \leq \mathrm{Pl}(A)$ where Pr(A) is the probability of A. When Bel(A) equals Pl(A) for every element in 2^{Θ}, the values correspond to probabilities. This implies that the function m takes non-zero values on the singletons only.

There exist convex sets of probabilities, expressed only as a set of intervals of probability, which cannot be represented by belief functions. For example, suppose that the four events denoted by {1, 2, 3, 4} have the probabilities given by

$p_1 = (1 - 2\,q)/2$

$p_2 = (1 - 2\,q)/2$

$p_3 = q$

$p_4 = q$

where q ranges over the values $0 \leq q \leq 1/4$. Table 4-1 gives the values of the probability as a function of q, the belief Bel, the plausability Pl, and the implied basic probability number m for all the events in the algebra generated by these four events. The important point to notice is that $m(\cdot)$ is not positive for all events.

Table 4-1: Probability Intervals not Representable by a Belief Function

Event	Prob(·)	Bel(·)	Pl(·)	m(·)
1	(1 - 2 q)/2	1/4	1/2	1/4
2	(1 - 2 q)/2	1/4	1/2	1/4
3	q	0	1/4	0
4	q	0	1/4	0
12	1 - 2 q	1/2	1	0
13	1/2	1/2	1/2	1/4
14	1/2	1/2	1/2	1/4
23	1/2	1/2	1/2	1/4
24	1/2	1/2	1/2	1/4
34	2 q	0	1/2	0
123	1 - q	3/4	1	-1/4
124	1 - q	3/4	1	-1/4
134	q + 1/2	1/2	3/4	-1/4
234	q + 1/2	1/2	3/4	-1/4
1234	1	1	1	0

More generally, it can be shown that if a set of probability intervals are given for the elements of a partition as

$$0 \le L_i \le p_i \le U_i \le 1, \; i=1, \ldots, n$$

then for there to exist a corresponding belief function Bel(·), it is necessary that both

$$\sum_{i=1}^{n} L_i + U_j - L_j \le 1, \text{ for all } j$$

and

$$\sum_{i=1}^{n} U_i + L_j - U_j \ge 1, \text{ for all } j.$$

This provides a quick and dirty test whether or not an expressed set of probability intervals are in fact representable by a belief function. Unfortunately, the sufficient conditions are considerably more complex.

One particularly nice feature of the theory of belief functions is that it distinguishes between *indifference* and *ignorance.* Complete ignorance is represented by the *vacuous* belief function that assigns basic probability one to the set Θ and zero to every subset. Complete indifference assigns an equal amount to all singleton propositions and zero to every other subset; this is precisely a uniform probability distribution on the elements of the partition. Any degree of ignorance can be expressed quite naturally between the two extremes of complete ignorance and a well-defined probability distribution.

The basic theory of belief functions requires that the frame of discernment be composed of mutually exclusive propositions. This means that only one proposition at a time can be true. In an expert system this condition is explicitly not satisfied; consequently, direct application of the the theory is impossible. We overcome this problem as follows. Let Ω be a set of mutually supporting propositions; that is, suppose that

$$\Omega = \{P_1, P_2, \ldots, P_n\}.$$

By mutually supporting we mean that any assignment of truth values to the propositions is possible. Let 2^{Ω} be a list of the possible assignments of truth values to the elements of Ω. If we now let 2^{Ω} be the frame of discernment then it is possible to use the theory of belief functions.

As originally proposed this theory used a rule of combination (now widely known as *Dempster's Rule of Combination*) for two basic probability assignments, m_1 and m_2, of the form

$$m_1 \otimes m_2 (A) = K \Sigma_{S \text{ intersect } T = A} m_1(S)\, m_2(T) \text{ for } A \neq \phi \qquad (4.1)$$

where the normalization constant K is chosen so the combined basic probabilities add to one. We have found this rule to be unsatisfactory and are currently exploring some alternative possibilities. Consider repeated application of this rule of combination, viz.,

$$m_1 \otimes m_2 \otimes \ldots \otimes m_n. \qquad (4.2)$$

What are the possible limits as n increases? It is fairly easy to see that both the uniform probability distribution and any belief function with a single focal element (including the vacuous belief function) are solutions and there are no others. It is unreasonable to expect that any rule of combination, when iterated in this manner, would yield every belief functions as a possible limit; on the other hand the observed behavior of the combination rule give in Equation 4.1 appears too restricted.

Typically, two different belief functions will not be defined over the same frame of discernment and a combination rule such as Equation 4.1 can not be directly applied. One frame is *compatible* with another if it can be obtained from it by splitting some of its possibilities into finer possibilities. The frame of the finer analysis is called a *refinement* of the original; the former is called a *coarsening* of the latter. Before application of a rule of combination it may be necessary to refine one or both of the frames in order to obtain a common frame of discernment.

5. REDUCING THE COMPUTATIONAL COMPLEXITY

There are considerable computational difficulties in using this theory. An initial assignment of 2^n basic probability assignments must be made, where n represents the number of propositions in the frame of discernment Θ. The required number of evaluations in using any combination rule increases exponentially as more propositions are included. It seems reasonable that intelligent exploitation of some structure could result in computational savings.

One way to reduce computational complexity is to assume that each piece of evidence either confirms or denies a single proposition rather than a disjunction. This is the approach that Barnett takes in his work (Barnett, 1981). While this will reduce the number of calculations from exponential to linear, it also means that the frame must be broken into independent partitions. This is a very strong assumption and not likely to be satisfied in practice. Here, we are interested in retaining the more natural possibility of dependence among the propositions in the system.

Another possible approach would discount, at an early stage of the calculations, sets with zero, or very small, basic probability assignments. Yet another approach is to ignore those sets with a cardinality higher than a predetermined threshold. This is the approach we take here. It is possible to reduce the computational problem from one of exponential time to one of polynomial time, and the degree of the polynomial can be set in advance by suitable choice of the threshold.

A belief function provides both a lower bound and an upper bound for the probability. The narrower the range of this interval the more definite the knowledge about the probability. It seems reasonable to require that any approximation to an m-function should preserve the properties of an m-function. This produces one of the following three possibilities:

1. a less definite assignment of uncertainty (a wider interval);

2. a more definite assignment of uncertainty (a narrower interval);

3. no change.

Suppose the cardinality of Θ is n (that is Θ contains n propositions). The approximations to be used involve neglecting m-function values attached to elements of 2^{Θ} with cardinality greater than a threshold value k. To restore the approximation to an m-function requires some form of renormalization. To produce the first case (above) it is proposed that the m-function is restored by moving all the ignored basic probability mass to the element Θ. To produce the second effect the excess basic probability mass should be added to the elements of 2^{Θ} with cardinality less than (or equal to) the threshold value k in proportion to their original values.

6. AN OUTER APPROXIMATION

Denoting the approximations by $m^{*}(\cdot)$, $\mathrm{Bel}^{*}(\cdot)$ and $\mathrm{Pl}^{*}(\cdot)$ and dealing with the conservative approach first, the desired results are as follows:

$\mathrm{Bel}^{*}(A) \leq \mathrm{Bel}(A)$, A contained in Θ

$\mathrm{Pl}^{*}(A) \geq \mathrm{Pl}(A)$, A contained in Θ. (6.1)

The remaining requirement is that the function $m^{*}(\cdot)$ does not violate the rules for an m-function, no matter what the value of k. The three requirements that a function must satisfy to be an m-function are simply

$m(\phi) = 0$

$0 \leq m(A) \leq 1$

and

$\Sigma_{A \text{ contained in } \Theta} m(A) = 1.$

We define m^* to be an *order k outer approximation* to m as follows:

$$m^*(\phi) = 0$$

$$m^*(A) = m(A) \qquad \text{if } |A| \leq k \text{ and A contained in } \Theta$$

$$m^*(A) = 0 \qquad \text{if } |A| > k \text{ and } A \neq \Theta$$

$$m^*(\Theta) = 1 - \Sigma_{A \text{ contained in } \Theta}\, m^*(A)$$

$$= \Sigma_{A \text{ contained in } \Theta,\ |A| > k}\, m(A) \tag{6.2}$$

where k is the threshold cardinality and $|\cdot|$ represents the the number of elements in the set. The first requirement for $m^*(\cdot)$ to be an m-function is trivially satisfied, and the second requirement is clearly satisfied for all the above parts (the latter simply because the sum must be less than or equal to the sum of all the $m(A)$, which is one). All that remains is to verify the third condition for an m-function.

$$\begin{aligned}\Sigma_{A \text{ contained in } \Theta}\, m^*(A) &= m^*(\phi) + \Sigma_{A \text{ contained in } \Theta,\ |A| \leq k}\, m(A) \\ &\qquad + \Sigma_{A \text{ contained in } \Theta,\ |A| > k,\ A \neq \phi}\, m(A) + m^*(\Theta) \\ &= 0 + \Sigma_{A \text{ contained in } \Theta,\ |A| \leq k}\, m(A) \\ &\qquad + 0 + \Sigma_{A \text{ contained in } \Theta,\ |A| > k}\, m(A) \\ &= \Sigma_{A \text{ contained in } \Theta}\, m(A) \\ &= 1.\end{aligned} \tag{6.3}$$

The range of possible values for k is given by

$$0 \leq k \leq n-1.$$

The value $k=0$ always yields tho vacuous probability assignment and the value $k=n-1$ always yields the original probability assignment. It is clear that the smaller the value of k the more information is being neglected and the approximation becomes more vague (the interval widens). The higher the value of k the less information is being neglected so the approximation should be closer to the original specification. Clearly there is also a possibility that the new m^*-function will not be different than the original m-function. This can happen when for a specific value of k, all the elements of 2^{Θ} with greater cardinality have m-function values of zero.

It's now necessary to prove the assertions made in Equation 6.1. First consider the belief $Bel^{*}(\cdot)$. It's easier to carry out the proof in four parts corresponding to the Equation 6.2. Clearly $Bel^{*}(\phi) = 0$, hence the first part is satisfied. The second part is satisfied as $Bel^{*}(A) = Bel(A)$ if the cardinality of A is less than or equal to k. The third part follows from

$$\begin{aligned} Bel^{*}(A) &= \Sigma_{B \text{ contained in } A}\ m^{*}(B) \\ &= \Sigma_{B \text{ contained in } A,\ |B| \leq k}\ m(B) \\ &\leq \Sigma_{B \text{ contained in } A}\ m(B) \\ &= Bel(A). \end{aligned} \tag{6.4}$$

Recall $Bel(\Theta) = 1$ is one of the requirements of an $Bel(\cdot)$ function. For the final part of the proof it is required to show that $Bel^{*}(\Theta) = 1$ (This actually follows automatically since $m^{*}(\cdot)$ satisfies the conditions of an m-function.).

$$\begin{aligned} Bel^{*}(\Theta) &= \Sigma_{B \text{ contained in } \Theta}\ m^{*}(B) \\ &= \Sigma_{B \text{ contained in } \Theta,\ |B| \leq k}\ m(B) + m^{*}(\Theta) \\ &= \Sigma_{B \text{ contained in } \Theta,\ |B| \leq k}\ m(B) + \Sigma_{B \text{ contained in } \Theta,\ |B| > k}\ m(B) \\ &= \Sigma_{B \text{ contained in } \Theta}\ m(B) \\ &= Bel(\Theta). \end{aligned} \tag{6.5}$$

Hence the condition on the $Bel^{*}(\cdot)$ has been satisfied. The condition on the $Pl^{*}(\cdot)$ now follows immediately.

$$Pl^{*}(A) = 1 - Bel^{*}(\bar{A}) \geq 1 - Bel(\bar{A}) = Pl(A). \tag{6.6}$$

It has now been shown that this form of approximation gives the desired effect of widening the interval between the belief and the plausability. The computational saving is made because of all the zeros used to replace the original assessments for sets with cardinality greater than k. Clearly these sets can now be ignored when performing a combination. This form of approximation could prove very useful in large systems: however. there is a danger that the approximation may not be very good. The best results will undoubtedly come when small basic probability numbers are assigned to sets with high cardinality. It may prove to be a worthwhile exercise to increment the value of k on successive iterations until two successive iterations yield close results. This sort of numerical exercise is a task for the future.

7. AN INNER APPROXIMATION

In a similar manner to Section 6 the opposite effect of narrowing the interval between the belief and the plausability can be achieved. Denoting these approximations by $m_*(\cdot)$, $\mathrm{Bel}_*(\cdot)$ and $\mathrm{Pl}_*(\cdot)$, the desired results are now as follows:

$$\mathrm{Bel}_*(A) \geq \mathrm{Bel}(A)\ ,\ A \text{ contained in } \Theta$$

$$\mathrm{Pl}_*(A) \leq \mathrm{Pl}(A)\ ,\ A \text{ contained in } \Theta. \tag{7.1}$$

Again the function $m_*(\cdot)$ must not violate the rules for an m-function. It is convenient to set up an intermediary function for ease of presentation:

$$M_k(A) = \Sigma_{B \text{ contained in } A,\ |B| \leq k}\ m(B). \tag{7.2}$$

We define m_* to be an *order k inner approximation* to m as follows:

$$m_*(\phi) = 0$$

$$m_*(A) = 0 \text{ if } |A| > k,\ A \text{ contained in } \Theta$$

$$= m(A) + \Sigma_{A \text{ contained in } D,\ |D| > k}\ m(A) \times m(D)\ /\ M_k(D), \text{ otherwise.} \tag{7.3}$$

Once again the first requirement of an m-function is trivially satisfied. As all the component parts of Equation 7.3 are non-negative it is sufficient to verify the third clause of an m-function. That is, we must verify that the component parts of $m_*(\cdot)$ sum to one.

$$\Sigma m_*(A) = \Sigma_{|A| \leq k} \{m(A) + \Sigma_{|D| > k,\ A \text{ contained in } D}\ m(A) \times m(D) \ / \ M_k(D)\}$$
$$= \Sigma_{|A| \leq k}\ m(A) + \Sigma_{|D| > k}\ m(D) \times M_k(D) \ / \ M_k(D))$$
$$= \Sigma_{|A| \leq k}\ m(A) + \Sigma_{|D| > k}\ m(D)$$
$$= \Sigma_{A \text{ contained in } \Theta}\ m(A)$$
$$= 1. \tag{7.4}$$

The conditions for an m-function are thus satisfied. The range of possible values for k is

$$1 \leq k \leq n\text{-}1.$$

The value k=1 corresponds to approximating the belief function by a probability distribution and the value k=n-1 yields the original probability assignment.

It is now necessary to prove the assertions made in Equation 7.1. The proof for the $Pl_*(\cdot)$ function part will follow in a similar fashion to that for the $Pl^*(\cdot)$ function above. But it is necessary to prove the belief part first. Clearly $Bel_*(\phi) = 0$, and $Bel_*(\Theta) = 1$ (as $m_*(\cdot)$ satisfies the conditions for an m-function.). Now it is necessary to prove the assertion in the cases where for any subset A of Θ, $|A|$ is either greater than k or less than or equal to k. In the latter case the following relationships hold:

$$Bel_*(A) = \Sigma_{E \text{ contained in } A}\ m_*(E)$$
$$= \Sigma_{E \text{ contained in } A,\ |E| \leq k}\ m(E)$$
$$+ \Sigma_{E \text{ contained in } A,\ |E| \leq k}\ \Sigma_{|D| > k,\ E \text{ contained in } D}\ m(E) \times m(D) \ / \ M_k(D))$$
$$= \Sigma_{E \text{ contained in } A,\ |E| \leq k}\ m(E) + c \quad \text{(say)}. \tag{7.5}$$

But since the cardinality of A is assumed to be less than or equal to k, then the cardinality of E is already determined, such that

$$\Sigma_{E \text{ contained in } A,\ |E| \leq k}\ m(E) = \Sigma_{E \text{ contained in } A}\ m(E)$$
$$= Bel(A). \tag{7.6}$$

Hence $Bel_*(A) \geq Bel(A)$ for the case where $|A| \leq k$. Now a proof for the other case ($|A| > k$) is needed. Equation 7.5 still holds and serves as the starting point here.

$$\mathrm{Bel}_*(A) = \Sigma_{E \text{ contained in } A,\ |E| \le k}\ m(E) + c \tag{7.7}$$

$$= \mathrm{Bel}(A) - \Sigma_{E \text{ contained in } A,\ |E| > k}\ m(E) + c.$$

If the terms in the constant c are expanded and collected in a suitably different way it becomes apparent that c contains the summation over $|E| > k$. That is

$$c = \Sigma_{E \text{ contained in } A,\ |E| > k} + c_1 \quad \text{(say)}. \tag{7.8}$$

Hence the conditions are satisfied as now it is clear that

$$\mathrm{Bel}_*(A) = \mathrm{Bel}(A) + c_1. \tag{7.9}$$

The conditions on the plausability function now follow immediately.

$$\mathrm{Pl}_*(A) = 1 - \mathrm{Bel}_*(\bar{A}) \le 1 - \mathrm{Bel}(\bar{A}) = \mathrm{Pl}(A). \tag{7.10}$$

It has now been shown that this form of approximation gives the desired effect of narrowing the interval between the belief and the plausability. A slightly better computational saving is achieved with this inner approximation than with the outer approximation because one additional value of m is known to be zero. The effects of the approximations are summarized in Table 7-1.

The terms 'increased' and 'decreased' in Table 7-1 should not be interpreted strictly: that is, they include the possibility of no change.

Both of these approximations set basically the same elements to zero, for a given value of k, to achieve a computational saving (the one exception is Θ.) It may be possible to combine the two approaches. As one approximation achieves a wider interval and the other achieves the opposite effect it should be possible to find some optimal combination of the two approximations. There are obviously many possible measures to optimize. A particularly simple one is to choose the proportionality constant β to minimize

$$\sum_{A \neq \Theta} \rho(\beta \mathrm{Bel}_*(A) - \mathrm{Bel}(A)).$$

If $\rho(x) = x^2$ the solution is

Table ?-1: Summary of the Approximations

Bel(·)	Cardinality	Approximation	
		$Bel^{*}(\cdot)$	$Bel_{*}(\cdot)$
	> k	fixed	increased
	≤ k	decreased	increased

Pl(·)	Cardinality	Approximation	
		$Pl^{*}(\cdot)$	$Pl_{*}(\cdot)$
	> k	increased	decreased
	≤ k	fixed	decreased

$$\beta_{*} = \sum_{A \neq \Theta} Bel_{*}(A)\ Bel(A) \,/ \sum_{A \neq \Theta} Bel_{*}(A)\ Bel_{*}(A).$$

We do not yet have any numerical experience with this approximation and we are examining other measures of distance.

We note that one of the primary motivations for the use of belief functions is the uncertainty attached to the probability assessments of the expert and the user. An order 1 inner approximation to a belief function is a probability distribution. An interesting question occurs: Is there any sense in which the order 1 inner approximation is an optimal approximation (estimate?) of the uncertain probability distribution which is represented by the belief function?

8. COMPUTER IMPLEMENTATION

We have developed a number of computer programs to use belief functions with rule-based systems. The following material discusses the algorithm to be followed when using forward chaining. The steps for backward chaining are similar.

The basic mechanism for propogating beliefs through the system are extension of the belief function to a refined frame and combination with another belief function. The user of the system is asked to provide evidence in the form of a belief function. If there is a match of the preconditions, then a rule will fire (become instantiated). Note that all of the preconditions for a rule must be matched before a rule will actually fire. Therefore, a user may be asked to input a number of beliefs before a rule does fire.

When a rule fires the current frame is refined and the current belief is extended to the rest of the elements. The extension of the current belief is combined with the extension of the expert-supplied belief attached to the rule. This process is then be repeated until a desired goal is reached.

An expert will have previously supplied his beliefs concerning each of these rules and these beliefs will be attached to the rules. Rules may have a number of precondition clauses but must only have one resultant clause. If a possible rule has a disjunctive precondition the rule is split into two or more rules with single (or possibly conjunctive) preconditions and the same resultant clause. If a possible rule has a conjunctive resultant clause the rule is split into two or more rules with a single (or possible disjunctive) result and the same preconditions. Note that this structure implies that the underlying graph is a Chow tree (Chow and Liu, 1968). A Chow tree is a directed (and connected) graph with the property that there are no cycles in the corresponding undirected graph.

Each rule base requires that an expert supply belief functions for each of the rules. These expert-supplied beliefs attached to the rules, in most cases, will not change from one use of the system to the next.

The system allows both forward and backward chaining. In a typical chaining program without belief functions, when a user supplies the fact(s) for a rule, the rule will fire and a conclusion will be reached with certainty. In this system, the user supplies evidence in the form of a belief function. The expert-supplied belief function for that rule is retrieved. All of the precondition clauses of the rule must be checked because they, too, may have attached belief functions. This is because previous rule instantiations may have created a belief function for these *if* clauses. Also, an *if* clause may have a belief function attached to it from a previous use as an evidence node.

From a computer programming standpoint this means that many belief functions must be created and stored and additional checking must be performed to determine if these belief functions are to be used with the current rule. This is mainly determined by looking at the active-set for each belief function. The active-set is a list of the propositions that a belief function pertains to. When compared, rules may have some of the same members of the active-set list, but no two rules should have exactly the same members. The procedure that takes two belief functions and defines them on a compatible frame of discernment is called *refinement*.

After all of the belief functions associated with a rule firing have been combined into one overall belief function, control is returned to the chaining program. The resulting belief function is stored for further use and is output to the user along with the conclusion (result of the instantiated rule). The user can then begin this process again by introducing more new evidence.

9. ACKNOWLEDGEMENT

I am grateful to my former student Gabriel Pei for initiating and fostering my interest in this topic, to my former student Perry Zalevsky for much of the computer programming, and to my current student Paul K. Black for continuing the work on this project. The computer facilities of the Department of Statistics at Carnegie-Mellon University used in the preparation of this report were acquired with partial financial support from the National Science Foundation (under Grant MCS 82-05126 and Grant DMS 84-04927), from the Office of Naval Research (under Contract N00014-84-G-0137), and from Digital Equipment Corporation. Preparation of this report was partially supported by the Office of Naval Research (under Contract N00014-84-K-0588).

10. REFERENCES

Barnett, J.A. (1981). Computational methods for a mathematical theory of evidence. *Proceedings of the National Conference on Artificial Intelligence*, 868-875.

Chow, C.K. and C.N. Liu (1968). Approximating discrete probability distributuions with dependence trees.] *IEEE Transactions on Information Theory*, it-14, 462-467.

Eddy, W.F and G.P. Pei (1986). Structures of rule-based belief functions. *IBM Journal of Research and Development*, **30**, 93-101.

Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton University Press, Princeton.

Spiegelhalter, D.J. (1986). Coherent evidence propogation in expert systems. Talk presented at the International Workshop on Bayesian Methodology, Stresa, Italy.