Structures of rule-based belief functions

by William F. Eddy
Gabriel P. Pei

Shafer's theory of evidential reasoning has recently received much attention as a possible model for probabilistic reasoning in expert system applications. This paper discusses the particular difficulties of implementing Shafer's belief functions in the context of the most common form of expert system, *rule-based* systems. The two most important problems are: the representation of the expert's subjective degrees of belief corresponding to his expressed rules, and the computational complexity of the inference mechanism for combining evidence. We argue that a potential approach for dealing with both problems is given by introducing constraints on the structure of the belief functions. These constraints, along with the expressed rules and the elicited belief values, form the expert's total knowledge.

1. Introduction

• *Uncertainty in rule-based systems*

Rule-based expert systems are collections of production rules which are linked or "chained" together to simulate a human expert's reasoning process. (A production rule is a statement of the form "If A then B," where A and B are logical propositions.) There is currently a great deal of interest in introducing uncertainty into the reasoning used in such production systems.

Probability theory has been used as the basis for combining numerical measures of uncertainty in several rule-based expert systems, for example, the SRI system PROSPECTOR. PROSPECTOR assigned independent expert-supplied conditional probabilities to propositions (see [1]). Users were permitted to input independent unconditional probabilities corresponding to observed evidence. Bayes' rule was used to compute the posterior probabilities.

Certainty factors are an alternative scheme for modeling uncertainties in rule-based expert systems. (Consider the MYCIN system for medical diagnosis described in [2].) Unlike probabilities, certainties are defined by and combined through an *ad hoc* set of rules.

These and other methods which attempt to introduce uncertainty into rule-based expert systems are reviewed more thoroughly in [3].

We can imagine at least two general methods for incorporating uncertainty into a rule-based expert system. In the first of these methods we suppose that an expert's opinions are *updated* by a user's opinions. We would expect the expert constructing the system to have a 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 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 the evidence nodes. Additionally, we would expect the system to update the expert's probability distribution with the user's probability distribution and then calculate the marginal (joint) probability distribution on the assignment of truth values to the propositions which are not antecedents of any rule (the goal nodes).

We see a variety of difficulties with this scheme:

1. It is unreasonable to expect anyone, user or expert, to express a joint probability distribution on the assignment of truth values to a large collection of propositions. This is because of
   a. The potential size of such a collection.
   b. The difficulty that most people have expressing probabilities even on collections with only one proposition.
2. There is an inherent asymmetry in the nature and use of the opinions supplied by the expert and the user. The user's opinion is used to update the expert's.
3. The amount of calculation required is so large as to prohibit implementation for systems with as few as fifty propositions.

In the second of the two general methods of incorporating uncertainty into a rule-based expert system, we suppose that a user's opinions are *pooled* with an expert's opinions. We expect the expert constructing the system to express uncertainty about the assignment of truth values to all the propositions in the system. Equally, we expect the user to express uncertainty about the assignment of truth values to all the propositions in the system. The expert and the user each express their uncertainty by a decomposition of their respective *universal* belief function; pieces of the decomposition are attached to the rules by the expert and are attached to the evidence nodes by the user. We use belief functions as a particular measure of uncertainty because of their intuitive appeal, but this general scheme allows the possibility of other measures.

This scheme overcomes some of the difficulties of the first scheme:

1. Neither expert nor user must express a "joint" opinion over the truth values of all the propositions in the system but must only express that portion of the opinion which is associated with a particular rule or evidence node.
2. The opinions of both expert and user are presumed to be subjective and are treated symmetrically. In particular, the roles of expert and user may be interchanged without affecting the result.

This paper is an attempt to use the belief functions described in [4] to overcome the difficulties of the first scheme and to achieve the goals of the second scheme. Current applications of Shafer's belief functions to rule-based expert systems face at least two major obstacles. First, in typical applications the set of epistemic possibilities termed the frame of discernment has previously been restricted to a mutually exclusive collection of propositions. This implies a fairly simple rule structure for the corresponding system. Second, there does not exist a standard methodology for attaching an expert's beliefs to the rules. Consequently, previous applications of belief functions to rule-based systems have been *ad hoc.*

This paper studies the consequences of the most common frames, that is, those with some mutually supporting propositions. We believe that it is more natural to base a frame of discernment on the expert-supplied rules associated with such mutually supporting propositions and that frames which are based on mutually exclusive propositions are artificially simple. We investigate how belief functions might be applied to mutually supporting propositions and give an example which demonstrates the construction of such a rule-based frame.

The frame of discernment is the critical element in the elicitation process. Rule-based expert systems imply "hidden" structures in the set of all possible subsets of propositions. Therefore, the expert only need express his degree of belief over the "visible" elements of the set of subsets. This can result in important computational savings (compared with the set of all possible subsets). This paper also demonstrates a decomposition of belief functions which allows a user, at the time of consultation with the system, to express his beliefs over the frame of discernment, which are then combined with the expert's expressed beliefs associated with the rules.

• *Belief functions in production systems*

Let Θ be a set of mutually exclusive and exhaustive propositions about a problem domain. The set Θ is called the *frame of discernment*. Let 2^Θ be the set of all subsets of Θ; elements of 2^Θ may be interpreted as general propositions in the problem domain.

A *belief function* $Bel(\cdot)$ is a function from 2^Θ into $[0, 1]$ which satisfies

$$Bel(\varnothing) = 0,$$

$$Bel(\Theta) = 1,$$

and

$$Bel\left(\bigcup_{i=1}^{n} S_i\right) \geqq \sum_{I \subset [1, \cdots, n]} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} S_i\right),$$

where $|I|$ is the cardinality of the index set I. A *basic probability assignment* $m(\cdot)$ is a function from 2^Θ into $[0, 1]$, which satisfies

$$m(\varnothing) = 0,$$

$$\sum_{S \subset \Theta} m(S) = 1.$$

There is a one-to-one correspondence between belief functions and basic probability assignments given by

$$Bel(S) = \sum_{T \subset S} m(T) \qquad \text{for } S \subset \Theta,$$

and

$$m(S) = \sum_{T \subset S} (-1)^{|S-T|} Bel(T) \qquad \text{for } S \subset \Theta.$$

The correspondence is easily seen by substituting one equation into the other in either order. *Dempster's rule* for combining two belief functions is most easily given in terms of the corresponding basic probability assignments. Let m_1 and m_2 be two basic probability assignments on the same frame, Θ. Dempster's rule defines their orthogonal sum $m_{12} = m_1 \oplus m_2$ by

$$m_{12}(\varnothing) = 0$$

and

$$m_{12}(A) = K \sum_{S \cap T = A} m_1(S) m_2(T) \qquad \text{for } A \neq \varnothing,$$

where

$$K^{-1} = \sum_{S \cap T \neq \varnothing} m_1(S) m_2(T).$$

The belief function m_{12} contains the combined evidence of m_1 and m_2.

Our proposed method addresses some of the difficulties of the first scheme described in the section "Uncertainty in rule-based systems" as follows. First, we substitute Shafer's belief functions for probabilities. This has several immediate advantages:

1. An expert or user who finds it difficult to express opinions in probabilities can more easily express opinions with belief functions.
2. An expert or user who is unable (or unwilling) to use general belief functions can always use the special belief function which corresponds to "I don't know" (or "I won't tell").
3. An expert or user who wishes to use probabilities to express an opinion has that option available, since probabilities are a special case of belief functions.

The basic probability assignment corresponding to the special belief function which expresses total ignorance has

$$m(S) = 0 \qquad \text{for } S \neq \Theta$$

and

$$m(\Theta) = 1.$$

The basic probability assignment corresponding to a probability distribution has

$$m(S) = 0 \qquad \text{for } S \notin \Theta.$$

Second, we allow the expert to express his opinion through partial beliefs attached to each individual rule, rather than requiring a joint belief in the assignment of truth values to all the propositions. While this appears to be an assumption of independence, we argue in the section "Dependence and its implications for elicitation" that, in fact, there can still be strong dependence in the expert-supplied beliefs.

Third, to reduce the amount of computation involved, we eliminate from consideration certain possible assignments of truth values to the propositions; our choice of which assignments to eliminate is dictated by the expert-supplied rules.

We imagine that the expert has a joint belief function on the assignment of truth values to all the propositions in the system but only expresses an opinion as partial belief functions associated with individual rules. We also imagine that the user has a joint belief function on the assignment of truth values to all the propositions in the system but only expresses an opinion as partial belief functions associated with individual evidence nodes.

The system operates by chaining using Dempster's rule of combination as follows. As a rule is "fired," the current frame of discernment is *refined* to include the proposition(s) implied by the rule. The belief function is *extended* to the refined frame and combined with the expert-supplied belief attached to the rule (after it has also been extended to the same refined frame). Because of the commutativity of Dempster's rule, this process can be initiated with the user's belief associated with a single evidence node or with the expert's belief associated with an individual rule.

We emphasize that each person has a single universal belief function and that a partial belief attached to a rule or a node represents part of only one possible decomposition of the universal belief. The symmetry inherent in the combination rule suggests that the structure of both user and expert beliefs should be the same.

The computational burden is the most significant drawback to the use of belief functions in production systems. If we have a collection of m general propositions, then there are 2^m mutually exclusive possible propositions within the system. Direct application of Shafer's theory requires that we consider the set of all subsets of this frame of discernment, a collection of 2^{2^m} general propositions. Obviously, any reduction in the size of the exponent 2^m is highly desirable. Barnett [5] has studied this problem; by breaking the frame into independent partitions (a very strong assumption), Barnett reduces the computational complexity from exponential to linear. Here, we are interested in retaining the more natural possibility of dependence among the propositions in the system.

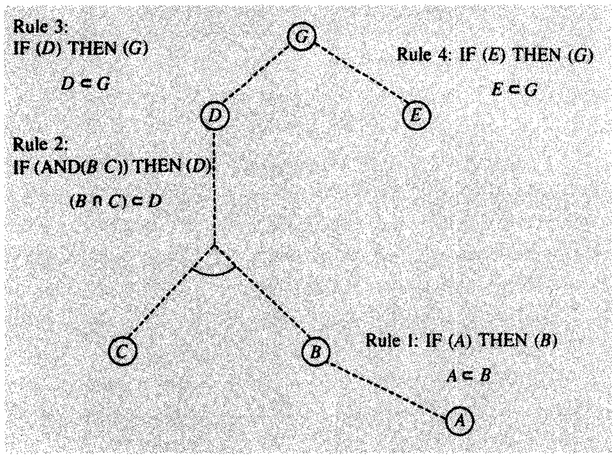

Figure 1

AND/OR representation of a small production system.

2. Mutually supporting propositions and rule-implied structure

• *Disjoint partitions*

Shafer's basic theory, presented in [4], requires that the frame of discernment (i.e., the set of possibilities) be composed of mutually exclusive propositions. This means that only one proposition can be true. In nearly all real situations, the propositions of concern are not mutually exclusive, and consequently direct application of Shafer's theory is impossible.

Let Ω be a set of mutually supporting propositions; that is, suppose that

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

By mutually supporting we merely mean that any assignment of truth values to the propositions is possible.

Let 2^Ω be a list of all the possible assignments of truth values to the elements of Ω. To simplify what follows, we use the symbol 1 to stand for *true*, the symbol 0 to stand for *false*, and the symbol $*$ to stand for *unspecified*. By using this notation any assignment can be written as a string of n symbols; for example, in a system Ω with four propositions, the string $\langle *1** \rangle$ would indicate that P_2 is true and $\langle *0** \rangle$ would indicate that P_2 is false. As a more general example, $\langle 110* \rangle$ would indicate that P_1 is true *and* P_2 is true *and* P_3 is false. Our standard interpretation of a completely specified string such as $\langle 1100 \rangle$ is the logical *and* of its components; that is, P_1 is true and P_2 is true and P_3 is false and P_4 is false. Our standard interpretation of a string such as $\langle 110* \rangle$ is that it represents the string $\langle 1100 \rangle$ *or* the string $\langle 1101 \rangle$.

Now, let $\mathcal{R}$ be a set of rules; that is, suppose that

$$\mathcal{R} = \{R_1, R_2, \cdots, R_m\},$$

where R_i is of the form $P_j \rightarrow P_k$. As an example, suppose we consider $\mathcal{R} = \{R_1\}$, where R_1 is $P_1 \rightarrow P_2$. We interpret this to mean that whenever P_1 is true, then P_2 must be true. Consequently, for any Ω, we have that in the presence of $\mathcal{R}$, the element $\langle 10* \cdots * \rangle$ cannot occur. We write $\langle 11* \cdots * \rangle = \langle 1** \cdots * \rangle$. We could equivalently write the dual equation $\langle 00* \cdots * \rangle = \langle *0* \cdots * \rangle$. This is simply a shorthand for the rule $R_1\colon P_1 \rightarrow P_2$ and means that for every belief function defined on 2^Ω the corresponding basic probability assignment $m(\cdot)$ assigns the value zero to $\langle 10* \cdots * \rangle$. Remembering that $\langle 10* \cdots * \rangle$ stands for a list of strings, we see that the rule R_1 has hidden 2^{n-2} elements of 2^Ω. We use the term *hidden* to indicate that conceptually the elements are present, but because every belief function assigns them a basic probability number of zero, they do not enter into any calculations. In practice, then, we take our list of propositions 2^Ω and use our list of rules $\mathcal{R}$ to determine which elements of 2^Ω remain visible.

To clarify the ideas we consider a production system based on the following six propositions:

A = Sleet is falling outside.
B = Harvey has mud on his shoes.
C = The door is unlocked.
D = Harvey's footprints are in the house.
E = Harvey's coat is in the house.
G = Harvey is in the house.

We suppose also that the expert-supplied rules are the following; the associated list of hidden elements is given with each rule.

R_1:	$A \rightarrow B$	$\langle 10**** \rangle$,
R_2:	$B \,\&\, C \rightarrow D$	$\langle *110** \rangle$,
R_3:	$D \rightarrow G$	$\langle ***1*0 \rangle$,
R_4:	$E \rightarrow G$	$\langle ****10 \rangle$.

Figure 1 gives a schematic representation of these rules in the form of an AND/OR tree. We refer to this example in Section 3.

• *Naive approaches to selection of belief functions*

There are several possible forms of belief functions which an expert might attach to a rule, and there are several possible forms of a belief function which a user might attach to evidence. In this section, we consider some of the possible forms of each in the context of a trivial system.

Let $\Theta = \{A, B\}$, and let $\mathcal{R} = \{R\}$, where the rule R is $A \rightarrow B$. One approach to selecting a belief function to attach to R would argue that the only situation that is important occurs when A is true *and* the rule R is correct. This leads to a choice of the form

$$m_1\{\langle 11\rangle\} = p_1,$$

$$m_1\{\Theta\} = 1 - p_1,$$

which can be interpreted as a belief p_1 that the proposition *AandB* is true. A value of $1 - p_1$ is assigned to Θ to reflect the uncommitted belief.

One possible approach to selecting a belief function to reflect evidence on A might choose

$$m_{A_1}\{\langle 1*\rangle\} = p_A,$$

$$m_{A_1}\{\Theta\} = 1 - p_A.$$

This could be interpreted as a belief p_A that the A is true and a belief $1 - p_A$ that is uncommitted. Applying Dempster's rule and letting $Bel_{A_11} = Bel_{A_1} \oplus Bel_1$, we find

$$Bel_{A_11}\{\langle *0\rangle\} = 0,$$

$$Bel_{A_11}\{\langle *1\rangle\} = p_1,$$

and Bel_{A_11} is zero for all other propositions except $\langle **\rangle$. We interpret the proposition $\langle *1\rangle$ as "B is true," and this proposition has belief p_1, which does *not* depend on the strength p_A of our evidence on A. No matter what our belief on A, the belief on B is the same. It appears then that at least one of the two choices Bel_1 or Bel_{A_1} is not sufficiently general.

A different approach to selecting a belief to attach to the rule R corresponds to a *logical* interpretation of the rule. If the rule is correct, then any of the states $\langle 00\rangle$, $\langle 01\rangle$, or $\langle 11\rangle$ is possible and the state $\langle 10\rangle$ is impossible. This leads to the choice

$$m_2\{\langle 00\rangle \text{ or } \langle 01\rangle \text{ or } \langle 11\rangle\} = p_2,$$

$$m_2\{\Theta\} = 1 - p_2.$$

This can be interpreted as a belief p_2 that the rule is correct and a belief $1 - p_2$ that is uncommitted. Combination of this belief with the belief function m_{A_1} given above to express evidence on A yields $Bel_{A_12} = Bel_{A_1} \oplus Bel_2$,

$$Bel_{A_12}\{\langle *0\rangle\} = 0,$$

$$Bel_{A_12}\{\langle *1\rangle\} = p_A p_2,$$

and Bel_{A_12} is zero for all other propositions except $\langle **\rangle$. In this case, the proposition "B is true" has a belief which does depend on p_A. However, because of the structure of m_{A_1}, the additional evidence can only reduce the initial belief in B. We again conclude that one of the two choices m_2 or m_{A_1} has a structure which is too limited to generate the kind of behavior we seek. In the next section we suggest that, in fact, all of Bel_1, Bel_2, and Bel_{A_1} are too limited.

Our goal is to choose a structure for both rule beliefs and user beliefs which will allow belief to shift from one set to another as additional evidence is pooled.

• *Attaching beliefs to rules*

The forms of the belief functions considered in the previous section are limited in their ability to represent expert information about the rules and user information about the evidence. We prefer the following form for expert-supplied opinions:

$$m_3\{\langle 11\rangle\} = p_3,$$

$$m_3\{\langle 0*\rangle\} = q_3,$$

$$m_2\{\Theta\} = 1 - (p_3 + q_3).$$

This is related to the *conditional* interpretation of the rule R. We assign belief p_3 to the proposition *AandB* and belief q_3 to the proposition *notA*.

We first note that because the element $\langle 10\rangle$ is hidden, as discussed in Section 2, we assign zero belief to it. We interpret the value of p_3 as expressing the strength of the necessity of the rule R. Similarly, we interpret q_3 as expressing the strength of the sufficiency of the rule. We also note that m_1 is a special case of m_3 letting $q_3 = 0$. Additionally, m_3 can be thought of as a generalization of m_2, taking the belief p_2 and splitting it into the two components p_3 and q_3.

Letting $Bel_{A_13} = Bel_{A_1} \oplus Bel_3$, we get

$$Bel_{A_13}\{\langle *0\rangle\} = 0,$$

$$Bel_{A_13}\{\langle *1\rangle\} = p_3/(1 - q_3 p_A),$$

and Bel_{A_13} is zero for all other propositions except $\langle **\rangle$. In this case, the proposition "B is true" has a belief which also does depend on p_A, but the belief can only increase with the additional evidence introduced by p_A. A more general structure for m_A is necessary in order to permit the pooling of both confirmatory and disconfirmatory evidence. We prefer to let the evidence on A be represented by a belief function that has structure identical to m_3. In this way, both user and expert opinions are treated symmetrically through Dempster's rule. We choose m_{A_2} as follows:

$$m_{A_2}\{\langle 11\rangle\} = p_A,$$

$$m_{A_2}\{\langle 0*\rangle\} = q_A,$$

$$m_{A_2}\{\Theta\} = 1 - (p_A + q_A).$$

One should note that m_{A_2} contains evidence about the consequent B. This may give rise to the legitimate concern that the user must somehow have knowledge of the likelihood of B given A. While this is a natural consequence of m_{A_2}, we note that the user may declare his ignorance by choosing p_A and q_A small.

In our view each user (and expert) has a single universal belief function over all the propositions in the system. We cannot expect this universal belief to be expressed but can rather expect expression of a partial belief function attached to a small number of closely related propositions. This partial belief is part of a (nonunique) decomposition of the larger universal belief function. All belief functions (including m_{A_2}) are already defined over the frame which implicitly includes the proposition B. The belief function m_{A_2} merely makes this explicit.

So that the evidence contained in m_{A_2} can be combined with that contained in m_3, it is necessary that

$$k = p_3 q_A + q_3 p_A < 1.$$

This simply measures the *weight of conflict* between the two beliefs. The weight of conflict can be determined during the chaining process.

Combination of the expert-supplied belief m_3 with the user-supplied belief function m_{A_2} given above yields $Bel_{A_23} = Bel_{A_2} \oplus Bel_3$:

$$Bel_{A_23}\{\langle *0 \rangle\} = 0,$$

$$Bel_{A_23}\{\langle 0* \rangle\} = 1 - (1 - q_A)(1 - q_3)/(1 - k),$$

$$Bel_{A_23}\{\langle *1 \rangle\} = 1 - (1 - p_A)(1 - p_3)/(1 - k),$$

and Bel_{A_23} is zero for all other propositions except $\langle ** \rangle$. It can be shown that Bel_{A_23} has the desired property; i.e., $Bel_{A_23}\{\langle *1 \rangle\}$ will increase or decrease from $Bel_3\{\langle *1 \rangle\}$ depending on the ratio p_A/q_A.

3. Refinement of frames and extension of beliefs

• *Refinement, extension, and compatibility*

Let

$$\Omega_1 = \{P_1, P_2, \cdots, P_n\}$$

be a collection of mutually supporting propositions, and let Θ_1 be the frame of discernment related to Ω_1 through the rule base R by the construction described in Section 2. Let

$$\Omega_2 = \{P_1, P_2, \cdots, P_n, P_{n+1}\},$$

include the additional proposition P_{n+1}, and let Θ_2 be the related frame. Then Θ_1 is a *coarsening* of Θ_2, and Θ_2 is a *refinement* of Θ_1. Now let p_i be the truth value of P_i; that is, let

$$p_i = \begin{cases} 1 & \text{if} \quad P_i \text{ is true,} \\ 0 & \text{if} \quad P_i \text{ is false,} \\ * & \text{if} \quad P_i \text{ is undetermined.} \end{cases}$$

If we have a basic probability assignment $m(\cdot)$ on 2^{Θ_1}, then the *minimal extension*, m_E, of m to 2^{Θ_2} is given by

$$m_E(\langle p_1 \cdots p_n * \rangle) = m(\langle p_1 \cdots p_n \rangle).$$

Obviously, m is a *marginal* basic probability assignment of m_E.

More generally, Shafer defines an extension when there is a compatibility relationship $\mathbf{r}$ between two frames $\Theta_1 = \{S_i\}$ and $\Theta_2 = \{T_j\}$ as follows:

$$Bel_E(A) = K\, Bel\{S_i \mid S_i \;\mathbf{r}\; T_j \text{ and } T_j \in A\} \qquad \text{for } A \subset \Theta_2,$$

where

$$K^{-1} = \sum_{S_i \mathbf{r} T_j \text{ and } T_j \subset \Theta_2} m(S_i).$$

It should be noted that minimal extension and marginal are not exact dual operations. The extension of the margin of a belief function is not necessarily equal to the original belief function. Nevertheless, the margin of the extension of a belief function is always exactly equal to the original belief function.

• *Dependence and its implications for elicitation*

The Dempster combination mechanism applies when two independent bodies of evidence over a frame are combined. It is important to notice that this is not necessarily an assumption of statistical independence if the combined beliefs are extended to a different but compatible frame of reference.

Following an example of Shafer [7], we suppose there is a frame $\Theta_1 = (A, B)$ and an assignment of belief

$$m(\langle 10 \rangle) = p_{10},$$

$$m(\langle 11 \rangle) = p_{11},$$

$$m(\langle 00 \rangle) = p_{00},$$

$$m(\langle 01 \rangle) = p_{01}.$$

Now suppose we are told that an event X, which is impossible if the true state is $\langle 11 \rangle$, occurs. Then our revised assignment of belief (conditional on X) becomes

$$m_x(\langle 11 \rangle) = 0,$$

$$m_x(\langle 10 \rangle) = p_{10}/(p_{10} + p_{01} + p_{00}),$$

$$m_x(\langle 01 \rangle) = p_{01}/(p_{10} + p_{01} + p_{00}),$$

$$m_x(\langle 00 \rangle) = p_{00}/(p_{10} + p_{01} + p_{00}).$$

Finally, suppose that Θ_2 is a compatible frame which concerns the event X; in the notation of Section 3, $\Theta_1 \;\mathbf{r}\; \Theta_2$, so that

$$m_x(\langle 11 \rangle) = m_2(\varnothing),$$

$$m_x(\langle 10 \rangle) = m_2(X),$$

$$m_x(\langle 01 \rangle) = m_2(\text{Not } X),$$

$$m_x(\langle 00 \rangle) = m_2(\text{Either } X \text{ or not } X).$$

It can be shown that there are an infinite number of probability assignments m_* and m_{**} so that we can write

$$m_2 = m_* \oplus m_{**}$$

and m_2 over Θ_2 is compatible with m_X over Θ_1.

This provides a basis for eliciting expert opinions. Suppose we wish to obtain a consensus opinion m_2 over Θ_2 from a pair of experts. We assume that one expert has an opinion over the frame Θ_2 and the other has an opinion over a compatible frame Θ_1, which is more complicated because this expert has observed an event X. We assume that it is possible to compute certain elements for the frame Θ_2 given the event X; for example, we suppose that it is possible to compute $m_x(\langle 10 \rangle)$ in the example. We elicit an opinion m_* from the first expert. Then we elicit the opinion m_{**} with respect to the frame Θ_2 from the second expert who has observed the event X. We then compute $m_2 = m_* \oplus m_{**}$ and compare the compatible elements from Θ_2 with those "computable" elements of Θ_1. This provides a check on the internal consistency of the experts.

We should also point out that the particular form of belief m_3 that we have chosen in Section 2 to represent the expert-supplied opinion allows a particularly simple method of elicitation. We merely ask the expert to supply r, the ratio of belief that the rule is correct to belief that the rule is incorrect. We also need to know u, the expert's total amount of uncertainty. Then the system can solve the two simultaneous equations to determine

$$p_3 = r(1 - u)/(1 + r)$$

and

$$q_3 = (1 - u)/(1 + r).$$

- *Chaining of beliefs*

The basic mechanisms for propagating beliefs through a system are extension of the belief function to the refined frame and Dempster's rule of combination. In summary, as each rule is fired two events occur:

1. The current frame is refined, the visible elements are identified (remembering that the rule base has increased), and the current belief is extended to the visible elements.
2. Dempster's rule is used to combine the extension of the current belief with the extension of the expert-supplied belief attached to the rule.

Let Ω be a collection of propositions, and let $\Theta = 2^\Omega$ be the corresponding frame of discernment prior to the firing of a rule. We further assume that every rule exists as a conjunctive normal form. That is,

$$R = (P_{i_1} \wedge \cdots \wedge P_{i_m}) \rightarrow P_C.$$

If a possible rule has a disjunctive antecedent, we merely split the rule into two or more rules with single (or possibly conjunctive) antecedents and the same consequent. If a possible rule has a conjunctive consequent, we merely split the rule into two or more rules with single (or possibly disjunctive) consequents and the same antecedent. We note that this assumption implies that the underlying graph is a Chow tree [8]. A Chow tree is a directed (and connected) graph with the property that there are no cycles in the corresponding undirected graph.

Suppose the rule R is fired. Let

$$\Omega^* = \Omega \cup P_{i_1} \cup \cdots \cup P_{i_m} \cup P_C,$$

and let Θ^* be the corresponding frame. The rule R determines a list of hidden elements according to the method described in Section 2 and hence a reduced frame of visible elements Θ^{**}. Before the firing of the rule R, current beliefs are represented by the basic probability assignment m on Θ. We construct the minimal extension of m to Θ^* and the margin of this extension on Θ^{**}. In each case we denote the basic probability assignment m. This should engender no confusion since the same beliefs are represented by m in each case; the beliefs are referred to different frames. Finally, let m_R stand for the basic probability assignment attached to the rule R. The effect of R can be written as

$$R\colon (\Omega, \Theta, m) \rightarrow (\Omega^*, \Theta^{**}, m \oplus m_R).$$

This notation is intended to describe the refinement and subsequent coarsening of the frame implied by the rule R and the simultaneous updating (by Dempster's rule) of m to the combination of m and m_R.

Recall now the example introduced in Section 2. First, we note that the four rules do in fact form a Chow tree as depicted in Fig. 1 using standard AND/OR notation. To begin the actual propagation of evidence through the network in the example, suppose the user supplies evidence on A. Then R_1 fires and the beliefs represented by m_A are extended to the smallest common refinement with the expert-supplied belief m_1 attached to the rule, namely the frame (A, B). Next, the frame is coarsened because of the one hidden element described by $\langle 10 \rangle$. Then m_A and m_1 are combined to get m_{A1}. In the notation of the previous paragraph,

$$R_1\colon (\Omega_A, \Theta_A, m_A) \rightarrow (\Omega^*_{A1}, \Theta^{**}_{A1}, m_{A1} = m_A \oplus m_1).$$

Suppose now the user supplies evidence on C; then R_2 fires. The beliefs represented by m_{A1} and m_C are extended to their smallest common refinement with the expert-supplied belief m_2 attached to the rule, namely (A, B, C, D) with the hidden elements $\langle 10{*}{*} \rangle$ and $\langle {*}110 \rangle$; there are six hidden elements and ten visible. All three of m_{A1}, m_C, and m_2 are combined by Dempster's rule to yield m_{A1C2}. In the notation above,

$$R_2\colon (\Omega_{A1}, \Theta_{A1}, m_{A1}) \rightarrow (\Omega^*_{A1C2}, \Theta^{**}_{A1C2}, m_{A1C2} = m_{A1} \oplus m_C \oplus m_2).$$

Figure 2 shows the state of the system at this point. The equation determining the hidden elements associated with each rule is displayed next to the rule. The ten visible elements are indicated by the Venn diagram.

Continuing the chaining process, R_3 fires and m_{A1C2} and m_3 are extended to the frame based on (A, B, C, D, G) with

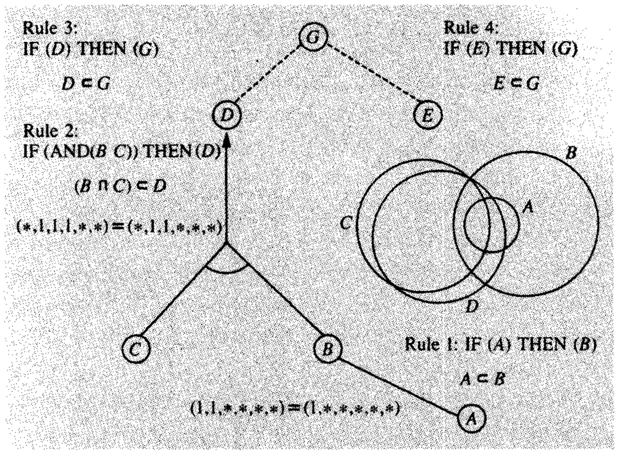

Figure 2

State of the production system after Rules 1 and 2 are fired. The hidden elements are summarized below each corresponding rule. The Venn diagram indicates the visible elements.

the hidden elements ⟨10∗∗∗⟩, ⟨∗110∗⟩, and ⟨∗∗∗10⟩; there are sixteen hidden elements and sixteen visible elements. The beliefs are combined to yield m_{A1C23}. We write

$$R_3\colon (\Omega_{A1C2}, \Theta_{A1C2}, m_{A1C2}) \rightarrow (\Omega^{*}_{A1C23}, \Theta^{**}_{A2C23}, m_{A1C23} = m_{A1C2} \oplus m_3).$$

If the user does not supply evidence on E, then the system will provide the marginal belief m_G on the coarsened frame (G). On the other hand, suppose the user does supply evidence on E. The beliefs represented by m_{A1C23}, m_E, and m_4 will be extended to the common refinement based on (A, B, C, D, E, G) with hidden elements ⟨10∗∗∗∗⟩, ⟨∗110∗∗⟩, ⟨∗∗∗1∗0⟩, and ⟨∗∗∗∗10⟩. There will be a total of 40 hidden elements and 24 visible ones. The beliefs m_{A1C23}, m_E, and m_4 will be combined to yield $m_{A1C23E4}$. That is,

$$R_4\colon (\Omega_{A1C23}, \Theta_{A1C23}, m_{A1C23}) \rightarrow (\Omega^{*}_{A1C23E4}, \Theta^{**}_{A1C23E4}, m_{A1C23E4} = m_{A1C23} \oplus m_E \oplus m_4).$$

Finally, the system computes the marginal belief on the coarsened frame (G). There are no hidden elements in this final coarsened frame, and in general applications there are no hidden elements in the final marginal frame on the goal nodes.

Suppose that an additional rule is appended to the previous system:

$$R_5\colon A \rightarrow B^C.$$

It should be noted that there are no hidden elements for the frame $\Theta = (A, B)$. If the network "fans out" in this manner, there can be no computational savings. On the other hand, if the system instead contains the additional rule

$$R_5\colon B \rightarrow E,$$

our methods will not work. If the system does not represent a Chow tree, then we have difficulty interpreting hidden and visible elements. We mention this to point out that if the network "fans in," then one must be particularly careful in applying our scheme.

There are two essential aspects of the system we have described here. First, we interpret Dempster's rule as a pooling mechanism. It is not an updating mechanism such as Bayes' rule. This world view of a belief function is essentially static. We assume the existence of an underlying universal frame of discernment with all epistemic possibilities contained in it. We are allowed to contemplate coarser versions of this frame and refinements of these coarsenings, but we are not allowed to augment the frame. Dempster's rule shifts belief over the different members of the power set of the frame.

Second, we believe that the propositions contained in the rules define the frame. However, the rules identify certain elements which must be structural zeros in the expert's opinion. Of course, the expert may assign general values to all the other elements. The expert's total belief may be decomposed in an infinite number of possible sums. We have proposed one such decomposition which is tractable to compute and can be interpreted in correspondence with the expert's rules.

Acknowledgment

Discussions with Gail Gong during the early stages of this work helped us formulate our ideas. We would also like to thank the reviewers for their comments; the criticism of one referee was particularly helpful in clarifying our ideas.

References

1. R. O. Duda, P. E. Hart, and N. J. Nilsson, "Subjective Bayesian Methods for Rule-Based Inference Systems," *AFIPS Conference Proceedings*, 1976, pp. 1075–1082.
2. E. H. Shortliffe and B. G. Buchanan, "A Model of Inexact Reasoning in Medicine," *Math. Biosci.* **23,** 351–379 (1975).
3. P. K. Black and W. F. Eddy, "Models of Inexact Reasoning," *Technical Report 351*, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA, 1985.
4. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
5. J. A. Barnett, "Computational Methods for a Mathematical Theory of Evidence," *Proceedings of the National Conference on Artificial Intelligence*, 1981, pp. 868–875.
6. J. G. Gordon and E. H. Shortliffe, "A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space," *Technical Report STAN-CS-84-1023*, Department of Computer Science, Stanford University, Stanford, CA, 1984.
7. G. Shafer, "Probability Judgment in Artificial Intelligence and Expert Systems," *Technical Report*, University of Kansas, Lawrence, KS, 1984; prepared for the Conference on the Calculus of Uncertainty in Artificial Intelligence and Expert Systems.
8. C. K. Chow and C. N. Liu, "Approximating Discrete Probability Distributions with Dependence Trees," *IEEE Trans. Info. Theory* **IT-14,** 462–467 (1968).

Received March 22, 1985; revised August 12, 1985

William F. Eddy *Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.* Professor Eddy is Associate Professor of Statistics at Carnegie-Mellon University, and a consultant to government and industry on the application of statistics. His research interests center on the theory of random sets and computational methods in statistics. He received the A.B. from Princeton University, Princeton, New Jersey, in 1971, and the M.A., the M.Phil., and the Ph.D. from Yale University, New Haven, Connecticut, the latter in 1976. He is a Fellow of the American Statistical Association and the Royal Statistical Society, and an elected member of the International Statistical Institute. Professor Eddy was the editor of *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, published in 1981.

Gabriel Pei *IBM Federal Systems Division, 9500 Godwin Drive, Manassas, Virginia 22110.* Dr. Pei joined IBM in 1983 and is currently working on an artificial intelligence project in combat control systems. His major research has been in the application of probabilistic methods to uncertain reasoning models for expert systems and development of knowledge acquisition techniques. Prior to joining IBM he worked for Daniel H. Wagner Associates developing mathematical search techniques and algorithms for the Office of Naval Research. Dr. Pei received his B.A. in mathematics in 1974 from the University of Virginia, Charlottesville, and the M.S. in 1975 and Ph.D. in 1980 in statistics from Carnegie-Mellon University, Pittsburgh, Pennsylvania. Dr. Pei is a member of the American Association for Artificial Intelligence, the American Mathematical Society, and the American Statistical Association.