8

Computational Methods for A Mathematical Theory of Evidence *†

Jeffrey A. Barnett

Abstract. Many knowledge-based expert systems employ numerical schemes to represent evidence, rate competing hypotheses, and guide search through the domain's problem space. This paper has two objectives: first, to introduce one such scheme, developed by Arthur Dempster and Glen Shafer, to a wider audience; second, to present results that can reduce the computation-time complexity from exponential to linear, allowing this scheme to be implemented in many more systems. In order to enjoy this reduction, some assumptions about the structure of the type of evidence represented and combined must be made. The assumption made here is that each piece of the evidence either confirms or denies a single proposition rather than a disjunction. For any domain in which the assumption is justified, the savings are available.

1 Introduction

How should knowledge-based expert systems reason? Clearly, when domain-specific idiosyncratic knowledge is available, it should be formalized and used to guide the inference process. Problems occur either when the supply of easy-to-formalize knowledge is exhausted before our systems pass the "sufficiency" test or when the complexity of representing and applying the knowledge is beyond the state of our system building technology. Unfortunately, with the current state of expert-system technology, this is the normal, not the exceptional case.

At this point, a fallback position must be selected, and if our luck holds, the resulting system exhibits behavior interesting enough to qualify as a success.

* From *Proceedings of the Seventh International Conference on Artificial Intelligence*, 1981, 868–875.

† This research is supported by the Defense Advanced Research Projects Agency under Contract No. DAHC15 72 C 0308. Views and conclusions contained in this report are the author's and should not be interpreted as representing the official opinion or policy of DARPA, the US Government, or any person or agency connected with them.

Typically, a fallback position takes the form of a uniformity assumption allowing the utilization of a non-domain-specific reasoning mechanism: for example, the numerical evaluation procedures employed in mycin [17] and internist [14], the simplified statistical approach described in [10], and a multivalued logic in [18]. The hearsay-ii speech understanding system [13] provides another example of a numerical evaluation and control mechanism—however, it is highly domain-specific.

Section 2 describes another scheme of plausible inference, one that addresses both the problem of representing numerical weights of evidence and the problem of combining evidence. The scheme was developed by Arthur Dempster [3, 4, 5, 6, 7, 8, 9], then formulated by his student, Glen Shafer [15, 16], in a form that is more amenable to reasoning in finite discrete domains such as those encountered by knowledge-based systems. The theory reduces to standard Bayesian reasoning when our knowledge is accurate but is more flexible in representing and dealing with ignorance and uncertainty. Section 2 is a review and introduction. Other work in this area is described in [12].

Section 3 notes that direct translation of this theory into an implementation is not feasible because the time complexity is exponential. However, if the type of evidence gathered has a useful structure, then the time complexity issue disappears. Section 4 proposes a particular structure that yields linear time complexity. In this structure, the problem space is partitioned in several independent ways and the evidence is gathered within the partitions. The methodology also applies to any domain in which the individual experiments (separate components of the evidence) support either a single proposition or its negation.

Section 5 and 6 develop the necessary machinery to realize linear time computations. It is also shown that the results of experiments may vary over time, therefore the evidence need not be monotonic. Section 7 summarizes the results and notes directions for future work in this area.

2 The Dempster-shafer Theory

A theory of evidence and plausible reasoning is described in this section. It is a theory of evidence because it deals with weights of evidence and numerical degrees of support based upon evidence. Further, it contains a viewpoint on the representation of uncertainty and ignorance. It is also a theory of plausible reasoning because it focuses on the fundamental operation of plausible reasoning, namely the combination of evidence. The presentation and notation used here closely parallels that found in [16].

After the formal description of how the theory represents evidence is presented in Sect. 2.1, an intuitive interpretation is given in Sect. 2.2, then a comparison is made, in Sect. 2.3, to the standard Bayesian model and similarities and differences noted. The rule for combining evidence, Dempster's orthogonal sum, is introduced in Sect. 2.4 and compared to the Bayesians' method of conditioning in Sect. 2.5. Finally, Sect. 2.6 defines the simple and separable support functions. These functions are the theory's natural representation of actual evidence.

2.1 Formulation of the Representation of Evidence

Let Θ be a set of propositions about the exclusive and exhaustive possibilities in a domain. For example, if we are rolling a die, Θ contains the six propositions of the form 'the number showing is i' where $1 \leq i \leq 6$. Θ is called the *frame of discernment* and 2^{Θ} is the set of all subsets of Θ. Elements of 2^{Θ}, i.e., subsets of Θ, are the class of general propositions in the domain; for example, the proposition 'the number showing is even' corresponds to the set of the three elements of Θ that assert the die shows either a 2, 4, or 6.

The theory deals with refinings, coarsenings, and enlargements of frames as well as families of compatible frames. However, these topics are not pursued here—the interested reader should see [16] where they are developed.

A function $\text{Bel}\colon 2^{\Theta} \to [0,1]$, is a *belief function* if it satisfies $\text{Bel}(\emptyset) = 0$, and for any collection, $A_1, \ldots, A_n$, of subsets of Θ,

$$\text{Bel}(A_1 \cup \cdots \cup A_n) \geq \sum_{\substack{I \subseteq \{1 \ldots n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \text{Bel}\Big(\bigcap_{i \in I} A_i\Big).$$

A belief function assigns to each subset of Θ a measure of our total belief in the proposition represented by the subset. The notation, $|I|$, is the cardinality of the set I.

A function $m\colon 2^{\Theta} \to [0,1]$ is called a *basic probability assignment* if it satisfies $m(\phi) = 0$ and

$$\sum_{A \subseteq \Theta} m(A) = 1.$$

The quantity, $m(A)$, is called A's *basic probability number*; it represents our exact belief in the proposition represented by A. The relation between these concepts and probabilities are discussed in Sect. 2.3. If m is a basic probability assignment, then the function defined by

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B), \quad \text{for all } A \subseteq \Theta \tag{1}$$

is a belief function. Further, if Bel is a belief function, then the function defined by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \text{Bel}(B) \tag{2}$$

is a basic probability assignment. If equations (1) and (2) are composed in either order, the result is the identity-transformation. Therefore, there corresponds to each belief function one and only one basic probability assignment. Conversely, there corresponds to each basic probability assignment one and only one belief function. Hence, a belief function and a basic probability assignment convey exactly the same information.

Other measures are useful in dealing with belief functions in this theory. A function $Q\colon 2^{\Theta} \to [0,1]$ is a *commonality function* if there is a basic probability assignment, m, such that

$$Q(A) = \sum_{A \subseteq B} m(B) \tag{3}$$

for all $A \subseteq \Theta$. Further, if Q is a commonality function, then the function defined by

$$\mathrm{Bel}(A) = \sum_{B \subseteq \neg A} (-1)^{|B|} Q(B)$$

is a belief function. From this belief function, the underlying basic probability assignment can be recovered using (2); if this is substituted into (3), the original Q results. Therefore, the sets of belief functions, basic probability assignments, and commonality functions are in one-to-one correspondence and each representation conveys the same information as any of the others.

Corresponding to each belief function are two other commonly used quantities that also carry the same information. Given a belief function Bel, the function $\mathrm{Dou}(A) = \mathrm{Bel}(\neg A)$, is called the *doubt function* and the function $P^{\star}(A) = 1 - \mathrm{Dou}(A) = 1 - \mathrm{Bel}(\neg A)$, is called the *upper probability function.*

For notational convenience, it is assumed that the functions Bel, m, Q, Dou, and $P^{\star}$ are each derived from one another. If one is subscripted, then all others with the same subscript are assumed to be derived from the same underlying information.

2.2 An Interpretation

It is useful to think of the basic probability number, $m(A)$, as the measure of a probability mass constrained to stay in A but otherwise free to move. This freedom is a way of imagining the noncommittal nature of our belief, i.e., it represents our ignorance because we can not further subdivide our belief and restrict the movement. Using this allusion, it is possible to give intuitive interpretations to the other measures appearing in the theory.

The quantity $\mathrm{Bel}(A) = \sum_{[B \subseteq A]} m(B)$ is the measure of the total probability mass constrained to stay somewhere in A. On the other hand, $Q(A) = \sum_{[A \subseteq B]} m(B)$ is the measure of the total probability mass that can move freely to any point in A. It is now possible to understand the connotation intended in calling m the measure of our exact belief and Bel the measure of our total belief. If $A \subseteq B \subseteq \Theta$, then this is equivalent to the logical statement that A implies B. Since $m(A)$ is part of the measure $\mathrm{Bel}(B)$, but not conversely, it follows that the total belief in B is the sum of the exact belief in all propositions that imply B plus the exact belief in B itself.

With this interpretation of Bel, it is easy to see that $\mathrm{Dou}(A) = \mathrm{Bel}(\neg A)$ is the measure of the probability mass constrained to stay out of A. Therefore, $P^\star(A) = 1 - \mathrm{Dou}(A)$ is the measure of the total probability mass that can move into A, though it is not necessary that it can all move to a single point, hence $P^\star(A) = \sum_{[A \cap B \neq \emptyset]} m(B)$ is immediate. It follows that $P^\star(A) \geq \mathrm{Bel}(A)$ because the total mass that can move into A is a superset of the mass constrained to stay in A.

2.3 Comparison with Bayesian Statistics

It is interesting to compare this and the Bayesian model. In the latter, a function $p\colon \Theta \to [0,1]$ is a *chance density function* if $\sum_{[a \in \Theta]} p(a) = 1$; and the function $\mathrm{Ch}\colon 2^\Theta \to [0,1]$ is a *chance function* if $\mathrm{Ch}(\emptyset) = 0$, $\mathrm{Ch}(\Theta) = 1$, and $\mathrm{Ch}(A \cup B) = \mathrm{Ch}(A) + \mathrm{Ch}(B)$ when $A \cap B = \emptyset$. Chance density functions and chance functions are in one-to-one correspondence and carry the same information. If Ch is a chance function, then $p(a) = \mathrm{Ch}(\{a\})$ is a chance density function; conversely, if p is a chance density function, then $\mathrm{Ch}(A) = \sum_{[a \in A]} p(a)$ is a chance function.

If p is a chance density function and we define $m(\{a\}) = p(a)$ for all $a \in \Theta$ and make $m(A) = 0$ elsewhere, then m is a basic probability assignment and $\mathrm{Bel}(A) = \mathrm{Ch}(A)$ for all $A \in 2^\Theta$. Therefore, the class of Bayesian belief functions is a subset of the class of belief functions. Basic probability assignments are a generalization of chance density functions while belief functions assume the role of generalized chance functions.

The crucial observation is that a Bayesian belief function ties all of its probability masses to *single points* in Θ, hence there is no freedom of motion. This follows immediately from the definition of a chance density function and its correspondence to a basic probability assignment. In this case, $P^\star = \mathrm{Bel}$ because, with no freedom of motion, the total probability mass that can move into a set is the mass constrained to stay there.

What this means in practical terms is that the user of a Bayesian belief function must somehow divide his belief among the singleton propositions. In some instances, this is easy. It we believe that a fair die shows an even number, then it seems natural to divide that belief evenly into three parts. If we don't know or don't believe the die is fair, then we are stuck.

In other words, there is trouble representing what we actually know without being forced to overcommit when we are ignorant. With the theory described here there is no problem—just let $m(\text{EVEN})$ measure the belief and the knowledge that is available. This is not to say that one should not use Bayesian statistics. In fact, if one has the necessary information, I know of no other proposed methodology that works as well. Nor are there any serious philosophical arguments against the use of Bayesian statistics. However, when our knowledge is not complete, as is often the case, the theory of Dempster and Shafer is an alternative to be considered.

2.4 The Combination of Evidence

The previous sections describe belief functions, the technique for representing evidence. Here, the theory's method of combining evidence is introduced. Let m_1 and m_2 be basic probability assignments on the same frame, Θ, and define $m = m_1 \oplus m_2$, their *orthogonal sum*, to be $m(\emptyset) = 0$ and

$$\begin{aligned} m(A) &= K \sum_{X \cap Y = A} m_1(X) \cdot m_2(Y) \\ K^{-1} &= 1 - \sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y) = \sum_{X \cap Y \neq \emptyset} m_1(X) \cdot m_2(Y), \end{aligned}$$

when $A \neq \emptyset$. The function m is a basic probability assignment if $K^{-1} \neq 0$; if $K^{-1} = 0$, then $m_1 \oplus m_2$ does not exist and m_1 and m_2 are said to be *totally* or *flatly contradictory.* The quantity $\log K = \mathrm{Con}(\mathrm{Bel}_1, \mathrm{Bel}_2)$ is called the *weight of conflict* between Bel_1 and Bel_2. This formulation is called *Dempster's rule of combination.*

It is easy to show that if m_1, m_2, and m_3 are combinable, then $m_1 \oplus m_2 = m_2 \oplus m_1$ and $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$. If v is the basic probability assignment such that $v(\Theta) = 1$ and $v(A) = 0$ when $A \neq \Theta$, then v is called the *vacuous belief function* and is the representation of total ignorance. The function, v, is the identity element for $\oplus$, i.e., $v \oplus m_1 = m_1$.

Figure 1 is a graphical interpretation of Dempster's rule of combination. Assume $m_1(A), m_1(B) \neq 0$ and $m_2(X), m_2(Y), m_2(Z) \neq 0$ and that m_1 and m_2 are 0 elsewhere. Then $m_1(A) + m_1(B) = 1$ and $m_2(X) + m_2(Y) + m_2(Z) = 1$. Therefore, the square in the figure has unit area since each side has unit length. The shaded rectangle has area $m_1(B) \cdot m_2(Y)$ and belief proportional to this measure is committed to $B \cap Y$. Thus, the probability number $m(B \cap Y)$ is proportional to the sum of the areas of all such rectangles committed to $B \cap Y$. The constant of proportionality, K, normalizes the result to compensate for the measure of belief committed to $\emptyset$. Thus, $K^{-1} = 0$ if and only if the combined belief functions invest no belief in intersecting sets; this is what is meant when we say belief functions are totally contradictory.

Using the graphical interpretation, it is straightforward to write down the formula for the orthogonal sum of more than two belief functions. Let $m = m_1 \oplus \cdots \oplus m_n$, then $m(\emptyset) = 0$ and

$$\begin{aligned} m(A) &= K \sum_{\cap A_i = A} \prod_{1 \leq i \leq n} m_i(A_i) \qquad (4) \\ K^{-1} &= 1 - \sum_{\cap A_i = \emptyset} \prod_{1 \leq i \leq n} m_i(A_i) = \sum_{\cap A_i \neq \emptyset} \prod_{1 \leq i \leq n} m_i(A_i) \end{aligned}$$

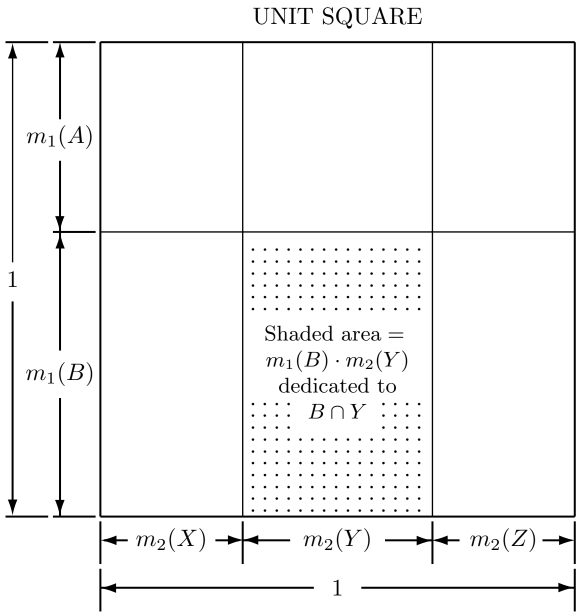

Fig. 1. Graphical representation of an orthogonal sum

when $A \neq \emptyset$. As above, the orthogonal sum is defined only if $K^{-1} \neq 0$ and the weight of conflict is $\log K$.

Since Bel, m, Q, Dou, and $P^\star$ are in one-to-one correspondence, the notation $\mathrm{Bel} = \mathrm{Bel}_1 \oplus \mathrm{Bel}_2$, etc., is used in the obvious way. It is interesting to note that if $Q = Q_1 \oplus Q_2$, then $Q(A) = KQ_1(A)Q_2(A)$ for all $A \subseteq \Theta$ where $A \neq \emptyset$.

2.5 Comparison with Conditional Probabilities

In the Bayesian theory, the function $\mathrm{Ch}(\cdot|B)$ is the *conditional chance* function, i.e., $\mathrm{Ch}(A|B) = \mathrm{Ch}(A \cap B)/\mathrm{Ch}(B)$, is the chance that A is true given that B is true. $\mathrm{Ch}(\cdot|B)$ is a chance function. A similar measure is available using Dempster's rule of combination.

Let $m_B(B) = 1$ and let m_B be 0 elsewhere. Then Bel_B, is a belief function that focuses all of our belief on B. Define $\mathrm{Bel}(\cdot|B) = \mathrm{Bel} \oplus \mathrm{Bel}_B$. Then [16] shows that $P^\star(A|B) = P^\star(A \cap B)/P^\star(B)$; this has the same form as the Bayesians' rule of conditioning, but in general, $\mathrm{Bel}(A|B) = (\mathrm{Bel}(A \cup \neg B) - \mathrm{Bel}(\neg B))/(1 - \mathrm{Bel}(\neg B))$. On the other hand, if Bel is a Bayesian belief function, then $\mathrm{Bel}(A|B) = \mathrm{Bel}(A \cap B)/\mathrm{Bel}(B)$.

Thus, Dempster's rule of combination mimics the Bayesians' rule of conditioning when applied to Bayesian belief functions. It should be noted, however, that the function Bel_B is *not* a Bayesian belief function unless $|B| = 1$.

2.6 Simple and Separable Support Functions

Certain kinds of belief functions are particularly well suited for the representation of actual evidence, among them are the classes of simple and separable support functions. If there exists an $F \subseteq \Theta$ such that $\mathrm{Bel}(A) = s \neq 0$ when $F \subseteq A$ and $A \neq \Theta$, $\mathrm{Bel}(\Theta) = 1$, and $\mathrm{Bel}(A) = 0$ when $F \nsubseteq A$, then Bel is a *simple support function*, F is called the *focus* of Bel, and s is called Bel's *degree of support*.

The vacuous belief function is a simple support function with focus Θ. If Bel is a simple support function with focus $F \neq \Theta$, then $m(F) = s$, $m(\Theta) = 1 - s$, and m is 0 elsewhere. Thus, a simple support function invests all of our committed belief on the disjunction represented by its focus, F, and all our uncommitted belief on Θ.

A separable support function is either a simple support function or the orthogonal sum of two or more simple support functions that can be combined. If it is assumed that simple support functions are used to represent the results of experiments, then the separable support functions are the possible results when the evidence from the several experiments is pooled together.

A particular case has occurred frequently. Let Bel_1 and Bel_2 be simple support functions with respective degrees of support s_1 and s_2, and the common focus, F. Let $\mathrm{Bel} = \mathrm{Bel}_1 \oplus \mathrm{Bel}_2$. Then $m(F) = 1 - (1 - s_1)(1 - s_2) = s_1 + s_2(1 - s_1) = s_2 + s_1(1 - s_2) = s_1 + s_2 - s_1 s_2$ and $m(\Theta) = (1 - s_1)(1 - s_2)$; m is 0 elsewhere.

The point of interest is that this formula appears as the rule of combination in mycin [17] and [11] as well as many other places. In fact, the earliest known development appears in the works of Jacob [2] circa 1713. For more than two and a half centuries, this formulation has had intuitive appeal to workers in a variety of fields trying to combine bodies of evidence pointing in the same direction. Why not use ordinary statistical methods? Because the simple support functions are not Bayesian belief functions unless $|F| = 1$.

We now turn to the problem of computational complexity.

3 The Computational Problem

Assume the result of an experiment—represented as the basic probability assignment, m—is available. Then, in general, the computation of $\mathrm{Bel}(A)$, $Q(A)$, $P^\star(A)$, or $\mathrm{Dou}(A)$ requires time exponential in $|\Theta|$. The reason[1] is the need to enumerate all subsets or supersets of A. Further, given any one of the functions, Bel, m, Q, $P^\star$, or Dou, computation of values of at least two of the others requires exponential time. If something is known about the structure of the belief function, then things may not be so bad. For example, with a simple support function, the computation time is no worse than $o(|\Theta|)$.

[1] I have not proved this. However, if the formulae introduced in Sect. 2 are directly implemented, then the statement stands.

The complexity problem is exaggerated when belief functions are combined. Assume $\mathrm{Bel} = \mathrm{Bel}_1 \oplus \cdots \oplus \mathrm{Bel}_n$, and the Bel_i are represented by the basic probability assignments, m_i. Then in general, the computations of K, $\mathrm{Bel}(A)$, $m(A)$, $Q(A)$, $P^\star(A)$, and $\mathrm{Dou}(A)$ require exponential time. Once again, knowledge of the structure of the m_i may overcome the dilemma. For example, if a Bayesian belief function is combined with a simple support function, then the computation requires only linear time.

The next section describes a particularly useful structuring of the m_i. Following sections show that all the basic quantities of interest can be calculated in $o(|\Theta|)$ time when this structure is used.

4 Structuring the Problem

Tonight you expect a special guest for dinner. You know it is important to play exactly the right music for her. How shall you choose from your large record and tape collection? It is impractical to go through all the albums one by one because time is short. First you try to remember what style she likes—was it jazz, classical, or pop? Recalling past conversations you find some evidence for and against each. Did she like vocals or was it instrumentals? Also, what are her preferences among strings, reeds, horns, and percussion instruments?

4.1 The Strategy

The problem solving strategy exemplified here is the well known technique of partitioning a large problem space in several independent ways, e.g., music style, vocalization, and instrumentation. Each partitioning is considered separately, then the evidence from each partitioning is combined to constrain the final decision. The strategy is powerful because each partitioning represents a smaller, more tractable problem.

There is a natural way to apply the plausible reasoning methodology introduced in Sect. 2 to the partitioning strategy. When this is done, an efficient computation is achieved. There are two computational components necessary to the strategy: the first collects and combines evidence within each partitioned space, while the second pools the evidence from among the several independent partitions.

In [16], the necessary theory for pooling evidence from the several partitions is developed using Dempster's rule of combination and the concept of refinings of compatible frames; in [1], computational methods are being developed for this activity. Below, a formulation for the representation of evidence within a single partitioning is described, then efficient methods are developed for combining this evidence.

4.2 Simple Evidence Functions

Let Θ be a partitioning comprised of n elements, i.e., $|\Theta| = n$; for example, if Θ is the set of possibilities that the dinner guest prefers jazz, classical, or pop music, then $n = 3$. Θ is a frame of discernment and, with no loss of generality, let $\Theta = \{i | 1 \leq i \leq n\}$. For each $i \in \Theta$, there is a collection of basic probability assignments μ_{ij} that represents evidence in favor of the proposition i, and a collection, ν_{ij} that represents the evidence against i. The natural embodiment of this evidence is as simple support functions with the respective foci $\{i\}$ and $\neg\{i\}$.

Define $\mu_i(\{i\}) = 1 - \prod(1 - \mu_{ij}(\{i\}))$ and $\mu_i(\Theta) = 1 - \mu_i(\{i\})$. Then μ_i is a basic probability assignment and the orthogonal sum of the μ_{ij}. Thus, μ_i is the totality of the evidence in favor of i, and $f_i = \mu(\{i\})$ is the degree of support from this simple support function. Similarly, define $\nu_i(\neg\{i\}) = 1 - \prod(1 - \nu_{ij}(\neg\{i\}))$ and $\nu_i(\Theta) = 1 - \nu_i(\neg\{i\})$. Then $a_i = \nu_i(\neg\{i\})$ is the total weight of support against i. Note, $\neg\{i\} = \Theta - \{i\}$, i.e., set complementation is always relative to the fixed frame, Θ. Note also that j, in μ_{ij}, and ν_{ij}, runs through respectively the sets of experiments that confirm or deny the proposition i.

The combination of all the evidence directly for and against i is the separable support function, $e_i = \mu_i \oplus \nu_i$. The e_i formed in this manner are called the *simple evidence functions* and there are n of them, one for each $i \in \Theta$. The only basic probability numbers for e_i that are not identically zero are $p_i = e_i(\{i\}) = K_i \cdot f_i \cdot (1 - a_i)$, $c_i = e_i(\neg\{i\}) = K_i \cdot a_i \cdot (1 - f_i)$, and $r_i = e_i(\Theta) = K_i \cdot (1 - f_i) \cdot (1 - a_i)$, where $K_i = (1 - a_i f_i)^{-1}$. Thus, p_i is the measure of support pro i, c_i is the measure of support con i, and r_i is the measure of the residue, uncommitted belief given the body of evidence comprising μ_{ij} and ν_{ij}. Clearly, $p_i + c_i + r_i = 1$.

The goal of the rest of this paper is to find efficient methods to compute the quantities associated with the orthogonal sum of the n simple evidence functions. Though the simple evidence functions arise in a natural way when dealing with partitions, the results are not limited to this usage—whenever the evidence in our domain consists of simple support functions focused on singleton propositions and their negations, the methodology is applicable.

4.3 Some Simple Observations

In the development of computational methods below, several simple observations are used repeatedly and the quantity $d_i = 1 - p_i = c_i + r_i$ appears. The first thing to note is $K_i^{-1} = 0$ iff $a_i = f_i = 1$. Further, if $K^{-1} \neq 0$ and v is the vacuous belief function, then

$p_i = 1$ iff $f_i = 1$	$c_i = 1$ iff $a_i = 1$
$p_i = 1 \Rightarrow c_i = r_i = 0$	$c_i = 1 \Rightarrow p_i = r_i = 0$
$f_i = 1$ iff $\exists j\, \mu_{ij}(\{i\}) = 1$	$a_1 = 1$ iff $\exists j\, \nu_{ij}(\neg\{i\}) = 1$

$$p_i = 0 \text{ iff } f_i = 0 \vee a_i = 1 \qquad c_i = 0 \text{ iff } a_i = 0 \vee f_i = 1$$
$$f_i = 0 \text{ iff } \forall j\, \mu_{ij} = v \qquad a_i = 0 \text{ iff } \forall j\, \nu_{ij} = v$$
$$r_i = 1 \text{ iff } p_i = c_i = 0 \qquad r_i = 0 \text{ iff } f_i = 1 \vee a_i = 1$$

5 Algorithms and Computations

The goal is to calculate quantities associated with $m = e_1 \oplus \cdots \oplus e_n$, where $n = |\Theta|$ and the e_i are the simple evidence functions defined in the previous section. All computations are achieved in $o(n)$ time measured in arithmetic operations.

Figure 2 is a schematic of information flow in a mythical system. The μ_{ij} and ν_{ij} may be viewed as sensors, where a sensor is an instance of a knowledge source that transforms observations into internally represented evidence, i.e., belief functions. Each is initially v, the vacuous belief function. As time passes and events occur in the observed world, these sensors can update their state by increasing or decreasing their degree of support. The simple evidence function, e_i, recomputes its state, a_i and f_i, and changes the stored values of p_i, d_i, c_i, and r_i each time one of its sensors reports a change. From the definitions of μ_{ij}, ν_{ij}, and e_i it is evident that the effect of an update can be recorded in constant time. That is to say, the time is independent of both the ranges of j in μ_{ij} and ν_{ij} and of n.

A user asks questions about the current state of the evidence. One set of questions concerns the values of various measures associated with arbitrary

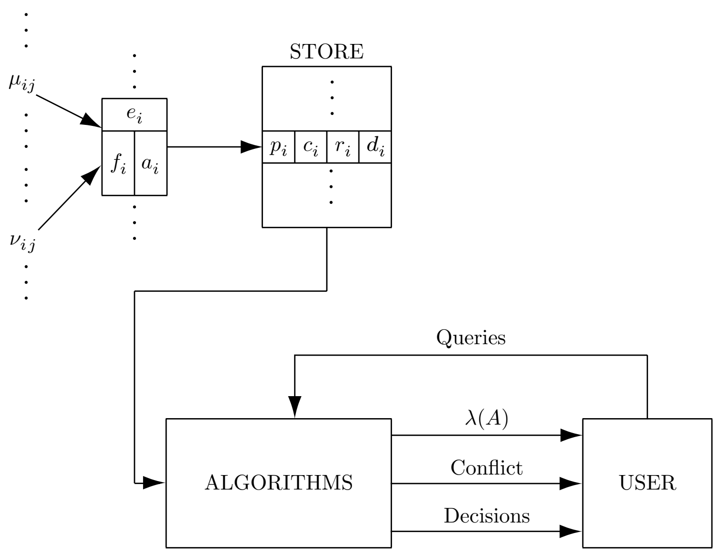

Fig. 2. Data flow model

$A \subseteq \Theta$. These questions take the form 'what is the value of $\lambda(A)$?', where λ is one of the functions Bel, m, Q, $P^\star$, or Dou. The other possible queries concern the general state of the inference process. Two examples are 'what is the weight of conflict in the evidence?' and 'is there an A such that $m(A) = 1$; if so, what is A?'. The o(n) time computations described in this section and in Sect. 6 answer all these questions.

One more tiny detour is necessary before getting on with the business at hand: it is assumed that subsets of Θ are represented by a form with the computational nicety of bit-vectors as opposed to, say, unordered lists of elements. The computational aspects of this assumption are: (1) the set membership test takes constant time independent of n and the cardinality of the set; (2) the operators $\subseteq$, $\cap$, $\cup$, $=$, complementation with respect to Θ, null, and cardinality compute in o(n) time.

5.1 The Computation of K

From equation (4), $K^{-1} = \sum_{[\cap A_i \neq \emptyset]} \prod_{[1 \leq i \leq n]} e_i(A_i)$ and the weight of internal conflict among the e_i is $\log K$ by definition. Note that there may be conflict between the pairs of μ_i and ν_i that is not expressed because K is calculated from the point of view of the given e_i. Fortunately, the total weight of conflict is simply $\log[K \cdot \prod K_i]$; this quantity can be computed in o(n) time if K can be.

In order to calculate K, it is necessary to find the collections of A_i that satisfy $\cap A_i \neq \emptyset$ and $e_i(A_i) \neq 0$, i.e., those collections that contribute to the summation. If A_i is not $\{i\}$, $\neg\{i\}$, or Θ, then $e_i = 0$ identically from the definition of the simple evidence functions. Therefore, assume throughout that $A_i \in \{\{i\}\, \neg\{i\}\, \Theta\}$.

There are exactly two ways to select the A_i such that $\cap A_i \neq \emptyset$.

1. If $A_j = \{j\}$ for some j, and $A_i = \neg\{i\}$ or $A_i = \Theta$ for $i \neq j$, then $\cap A_i = \{j\} \neq \emptyset$. However, if two or more A_i are singletons, then the intersection is empty.
2. If none of the A_i are singletons, then the situation is as follows. Select any $S \subseteq \Theta$ and let $A_i = \Theta$ when $i \in S$ and $A_i = \neg\{i\}$ when $i \notin S$. Then $\cap A_i = S$. Therefore, when no A_i is a singleton, $\cap A_i \neq \emptyset$ unless $A_i = \neg\{i\}$ for all i.

Let J, K, L be predicates respectively asserting that exactly one A_i is a singleton, no A_i is a singleton, i.e., all $A_i \in \{\neg\{i\}\, \Theta\}$, and all $A_i = \neg\{i\}$. Then equation (4) can be written as

$$\begin{aligned} K^{-1} &= \sum_{\cap A_i \neq \emptyset} \prod_{1 \leq i \leq n} e_i(A_i) \\ &= \sum_{J} \prod_{1 \leq i \leq n} e_i(A_i) + \sum_{K} \prod_{1 \leq i \leq n} e_i(A_i) - \sum_{L} \prod_{1 \leq i \leq n} e_i(A_i). \end{aligned}$$

Now the transformation, below called transformation T,

$$\sum_{x_j \in S_j} \prod_{1 \le i \le n} f_i(x_i) = \prod_{1 \le i \le n} \sum_{x \in S_i} f_i(x) \tag{T}$$

can be applied to each of the three terms on the right; after some algebra, it follows that

$$K^{-1} = \sum_{1 \le q \le n} p_q \prod_{i \ne q} d_i + \prod_{1 \le i \le n} d_i - \prod_{1 \le i \le n} c_i, \tag{5}$$

where $p_i = e_i(\{i\})$, $c_i = e_i(\neg\{i\})$, and $d_i = e_i(\neg\{i\}) + e_i(\Theta)$ have been substituted. If $p_q = 1$ for some q, then $d_q = c_q = 0$ and $K^{-1} = \prod_{[i \ne q]} d_i$. On the other hand, if $p_i \ne 1$ for all i, then $d_i \ne 0$ for all i and equation (5) can be rewritten as

$$K^{-1} = \Big[\prod_{1 \le i \le n} d_i\Big]\Big[1 + \sum_{1 \le i \le n} p_i/d_i\Big] - \prod_{1 \le i \le n} c_i. \tag{6}$$

In either case, it is easy to see that the computation is achieved in o(n) time, as is the check for $p_i = 1$.

5.2 The Computation of m(A)

From equation (4), the basic probability numbers, $m(A)$ for the orthogonal sum of the simple evidence functions are

$$m(A) = K \sum_{\cap A_i = A} \prod_{1 \le i \le n} e_i(A_i),$$

for $A \ne \emptyset$ and by definition, $m(\emptyset) = 0$. Also, m can be expressed by

$$\begin{aligned} m(\emptyset) &= 0 \\ m(\{q\}) &= K\Big[p_q \prod_{i \ne q} d_i + r_q \prod_{i \ne q} c_i\Big] \\ M(A) &= K\Big[\prod_{i \in A} r_i\Big]\Big[\prod_{i \notin A} c_i\Big], \qquad \text{when } |A| \ge 2. \end{aligned} \tag{7}$$

It is easy to see that the calculation is achieved in o(n) time since $|A| + |\neg A| = n$.

Derivation of these formulae is straightforward. If $A = \cap A_i$, then $A \subseteq A_i$ for $1 \le i \le n$ and for all $j \notin A$, there is an A_i such that $j \notin A_i$. Consider the case in which A is a nonsingleton nonempty set; If $i \in A$, then $A_i = \Theta$— the only other possibilities are $\{i\}$ or $\neg\{i\}$, but neither contains A. If $i \notin A$, then both $A_i = \neg\{i\}$ and $A_i = \Theta$ are consistent with $A \subseteq A_i$. However, if $A_i = \Theta$ for some $i \notin A$, then $\cap A_i \supseteq A \cup \{i\} \ne A$. Therefore, the only choice is $A_i = \neg\{i\}$ when $i \notin A$ and $A_i = \Theta$ when $i \in A$. When it is noted that

$e_i(\Theta) = r_i$ and $e_i(\neg\{i\}) = c_i$ and, transformation T is applied, the formula for the nonsingleton case in equation (7) follows.

When $A = \{q\}$, there are two possibilities: $A_q = \Theta$ or $A_q = \{q\}$. If $A_q = \Theta$, then the previous argument for nonsingletons can be applied to justify the appearance of the term $r_q \prod_{[i \neq q]} c_i$. If $A_q = \{q\}$, then for each $i \neq q$ it is proper to select either $A_i = \Theta$ or $A_i = \neg\{i\}$ because, for both choices, $A \subseteq A_i$; actually, $\cap A_i = \{q\} = A$ because $A_q = A$. Using transformation T and noting that $e_q(\{q\}) = p_q$ and $d_i = c_i + r_i$ gives the term $p_q \prod_{[i \neq q]} d_i$ in the above and completes the derivation of equation (7).

5.3 The Computations of Bel(A), $P^\star$(A), and Dou(A)

Since $\mathrm{Dou}(A) = \mathrm{Bel}(\neg A)$ and $P^\star(A) = 1 - \mathrm{Dou}(A)$, the computation of $P^\star$ and Dou is $\mathrm{o}(n)$ if Bel can be computed in $\mathrm{o}(n)$ because complementation is an $\mathrm{o}(n)$ operation. Let Bel be the orthogonal sum of the n simple evidence functions. Then $\mathrm{Bel}(\emptyset) = 0$ by definition and for $A \neq \emptyset$,

$$\begin{aligned}\mathrm{Bel}(A) &= \sum_{B \subseteq A} m(B) = \sum_{\emptyset \neq B \subseteq A} K \sum_{\cap B_i = B} \prod_{1 \leq i \leq n} e_i(B_i) \\ &= K \sum_{\emptyset \neq \cap A_i \subseteq A} \prod_{1 \leq i \leq n} e_i(A_i).\end{aligned}$$

Bel is also expressed by

$$\mathrm{Bel}(A) = K\Big[\Big[\prod_{1 \leq i \leq n} d_i\Big]\Big[\sum_{i \in A} p_i/d_i\Big] + \Big[\prod_{i \notin A} c_i\Big]\Big[\prod_{i \in A} d_i\Big] - \prod_{1 \leq i \leq n} c_i\Big] \qquad (8)$$

when $d_i \neq 0$ for all i. If $d_q = 0$, then $p_q = 1$. Therefore, $m(\{q\}) = \mathrm{Bel}(\{q\}) = 1$. In all variations, $\mathrm{Bel}(A)$ can be calculated in $\mathrm{o}(n)$ time. Since the formula evaluates $\mathrm{Bel}(\emptyset)$ to 0, only the case of nonempty A needs to be argued.

The tactic is to find the collections of A_i satisfying $\emptyset \neq \cap A_i \subseteq A$ then apply transformation T. Recall that the only collections of A_i that satisfy $\emptyset \neq \cap A_i$ are those in which (1) exactly one A_i is a singleton or (2) no A_i is a singleton and at least one $A_i = \Theta$. To satisfy the current constraint, we must find the subcollections of these two that also satisfy $\cap A_i \subseteq A$.

If exactly one A_i is a singleton, say $A_q = \{q\}$, then $\cap A_i = \{q\}$. In order that $\cap A_i \subseteq A$ it is necessary and sufficient that $q \in A$. Thus, the contribution to $\mathrm{Bel}(A)$, when exactly one singleton A_i is permitted, is the sum of the contributions for all $i \in A$. A brief computation shows this to be $[\prod_{[1 \leq i \leq n]} d_i][\sum_{[i \in A]} p_i/d_i]$.

When no A_i is a singleton, it is clear that $A_i = \neg\{i\}$ for $i \notin A$; otherwise, $i \in A$ and $\cap A_i \not\subseteq A$. For $i \in A$, either $A_i = \neg\{i\}$ or $A_i = \Theta$ is permissible. The value of the contribution to Bel from this case is given by the term $[\prod_{[i \notin A]} c_i][\prod_{[i \in A]} d_i]$. Since at least one of the $A_i = \Theta$ is required, we must deduct for the case in which $A_i = \neg\{i\}$ for all i, and this explains the appearance of the term $-\prod_{[1 \leq i \leq n]} c_i$.

5.4 The Computation of $Q(A)$

The definition of the commonality function shows that $Q(\emptyset) = 1$ identically. For $A \neq \emptyset$

$$Q(A) = \sum_{A \subseteq B} m(B) = \sum_{A \subseteq B} K \sum_{\cap A_i = B} \prod_{1 \le i \le n} e_i(A_i) = K \sum_{A \subseteq \cap A_i} \prod_{1 \le i \le n} e_i(A_i).$$

Q can be expressed also by

$$\begin{aligned} Q(\emptyset) &= 1 \\ Q(\{q\}) &= K(p_q + r_q) \prod_{i \neq q} d_i \\ Q(A) &= K\Big[\prod_{i \in A} r_i\Big]\Big[\prod_{i \notin A} d_i\Big], \qquad \text{when } |A| \ge 2. \end{aligned}$$

In order that a collection, A_i, satisfy $A \subseteq \cap A_i$, it is necessary and sufficient that $A \subseteq A_i$ for all i. If $i \notin A$, then both $A_i = \neg\{i\}$ and $A_i = \Theta$ fill this requirement but $A_i = \{i\}$ fails. If $i \in A$, then clearly $A_i = \neg\{i\}$ fails and $A_i = \Theta$ works. Further, $A_i = \{i\}$ works iff $A = \{i\}$. It is now a simple matter to apply transformation T and generate the above result. It is evident that $Q(A)$ can be calculated in o(n) time.

6 Conflict and Decisiveness

In the previous section, a mythical system was introduced that gathered and pooled evidence from a collection of sensors. It was shown how queries such as 'what is the value of $\lambda(A)$?' could be answered efficiently, where A is an arbitrary subset of Θ and λ is one of Bel, m, Q, $P^\star$, or Dou. It is interesting to note that a sensor may change its value over time. The queries report values for the current state of the evidence. Thus, it is easy to imagine an implementation performing a monitoring task, for which better and more decisive data become available, as time passes, and decisions are reevaluated and updated on the bases of the most current evidence.

In this section, we examine more general queries about the combined evidence. These queries seek the subsets of Θ that optimize one of the measures. The sharpest question seeks the $A \subseteq \Theta$, if any, such that $m(A) = 1$. If such an A exists, it is said to be the *decision*. Vaguer notions of decision in terms of the other measures are examined too.

The first result is the necessary and sufficient conditions that the evidence be totally contradictory. Since the orthogonal sum of the evidence does not exist in this case, it is necessary to factor this out before the analysis of decisiveness can be realized. All queries discussed in this section can be answered in o(n) time.

6.1 Totally Contradictory Evidence

Assume there are two or more $p_i = 1$, say $p_a = p_b = 1$, where $a \neq b$. Then $d_j = c_j = r_j = 0$, for both $j = a$ and $j = b$. The formula for K is

$$K^{-1} = \sum_{1 \leq q \leq n} p_q \prod_{i \neq q} d_i + \prod_{1 \leq i \leq n} d_i - \prod_{1 \leq i \leq n} c_i,$$

and it is easy to see that $K^{-1} = 0$ under this assumption. Therefore, the evidence is in total conflict by definition.

Let $p_a = 1$ and $p_i \neq 1$ for $i \neq a$. Then $d_a = c_a = 0$, and $d_i \neq 0$ for $i \neq a$. Therefore. the above formula reduces to $K^{-1} = \prod_{[i \neq a]} d_i \neq 0$ and the evidence is not totally contradictory.

Now assume $p_i \neq 1$, hence $d_i \neq 0$, for all i. Can $K^{-1} = 0$? Since $d_i = c_i + r_i$, it follows that $\prod d_i - \prod c_i \geq 0$. If $K^{-1} = 0$, this difference must vanish. This can happen only if $r_i = 0$ for all i. Since $p_i \neq 0$, this entails $c_i = 1$ for all i. In this event the $p_i = 0$ and $K^{-1} = 0$.

Summary: The evidence is in total conflict iff either (1) there exists an $a \neq b$ such that both $p_a = p_b = 1$ or (2) $c_i = 1$ for all $i \in \Theta$.

6.2 Decisiveness in *m*

The evidence is *decisive* when $m(A) = 1$ for some $A \subseteq \Theta$ and A is called the *decision.* If the evidence is decisive and A is the decision, then $m(B) = 0$ when $B \neq A$ because the measure of m is 1. The evidence cannot be decisive if it is totally contradictory because the orthogonal sum does not exist, hence m is not defined. The determination of necessary and sufficient conditions that the evidence is decisive and the search for the decision is argued by cases.

If $p_q = 1$ for some $q \in \Theta$, then the evidence is totally contradictory if $p_i = 1$ for some $i \neq q$. Therefore, assume that $p_i \neq 1$ for $i \neq q$. From equation (7) it is easy to see $m(\{q\}) = K \prod_{[i \neq q]} d_i$ because $r_q = 0$. Further, it was shown directly above that $K^{-1} = \prod_{[i \neq q]} d_i$ under the same set of assumptions. Thus, $m(\{q\}) = 1$.

The other possibility is that $p_i \neq 1$, hence $d_i \neq 0$, for all $i \in \Theta$. Define $C = \{i | c_i = 1\}$, and note that if $|C| = n$, the evidence is totally contradictory. For $i \in C$, $p_i = r_i = 0$ and $d_i = 1$. If $|C| = n - 1$, then there is a w such that $\{w\} = \Theta - C$. Now $p_w \neq 1$ and $c_w \neq 1$ entails $r_w \neq 0$; therefore, from equation (7)

$$m(\{w\}) \;=\; K\Big[p_w \prod_{i \neq w} d_i + r_w \prod_{i \neq w} c_i\Big] \;=\; K[p_w + r_w] \;\neq\; 0.$$

If there is a decision in this case, it must be $\{w\}$. Direct substitution into equation (5) shows that, in this case, $K^{-1} = p_w + r_w$ and therefore, $m(\{w\}) = 1$.

Next, we consider the cases where $0 \leq |C| \leq n-2$ and therefore, $|\neg C| \geq 2$. Then, from equation (7)

$$m(\neg C) \;=\; K\Big[\prod_{i\notin C} r_i\Big]\Big[\prod_{i\in C} c_i\Big] \;=\; K\prod_{i\notin C} r_i \;\neq\; 0 \tag{9}$$

because $i \notin C$ iff $c_i \neq 1$ (and $p_i \neq 1$ for all $i \in \Theta$) has been assumed: hence, $r_i \neq 0$ for all $i \in \neg C$. Therefore, if the evidence is decisive, $m(\neg C) = 1$ is the only nonzero basic probability number. Can there be a $p_q \neq 0$? Obviously, $q \notin C$. The answer is no since $d_i \neq 0$, hence, $m(\{q\}) = K[p_q \prod_{[i\neq q]} d_i + r_q \prod_{[i\neq q]} c_i] \neq 0$, a contradiction. Thus, $p_i = 0$ for all $i \in \Theta$. From equation (5) it now follows that $K^{-1} = \prod_{[1\leq i\leq n]} d_i - \prod_{[1\leq i\leq n]} c_i$. Therefore, from (9), $\prod_{[i\notin C]} r_i = \prod_{[1\leq i\leq n]} d_i - \prod_{[1\leq i\leq n]} c_i$ if $m(\neg C) = 1$. Since $d_i = c_i = 1$ when $i \in C$, this can be rewritten as $\prod_{[i\notin C]} r_i = \prod_{[i\notin C]} d_i - \prod_{[i\notin C]} c_i$. But $d_i = c_i + r_i$. Therefore, this is possible exactly where $c_i = 0$ when $i \notin C$.

Summary: Assuming the evidence is not in total conflict, it is decisive iff either (1) exactly one $p_i = 1$; the decision is $\{i\}$. (2) There exists a w such that $c_w \neq 1$ and $c_i = 1$ when $i \neq w$; the decision is $\{w\}$. Or (3) there exists a $W \neq \emptyset$ such that $r_i = 1$ when $i \in W$ and $c_i = 1$ when $i \notin W$; the decision is W.

6.3 Decisiveness in Bel, $P^\star$, and Dou

If $\mathrm{Bel}(A) = \mathrm{Bel}(B) = 1$, then $\mathrm{Bel}(A \cap B) = 1$ and it is always true that $\mathrm{Bel}(\Theta) = 1$. The minimal A such that $\mathrm{Bel}(A) = 1$ is called the *core* of Bel. If the evidence is decisive, i.e., $m(A) = 1$ for some $A \subseteq \Theta$, then clearly A is the core of Bel. Assume the evidence is not decisive, not totally contradictory, and $\mathrm{Bel}(A) = 1$, then equations (8) and (6) can be smashed together and rearranged to show that

$$\sum_{q\notin A} p_q \prod_{i\neq q} d_i + \prod_{i\in A} d_i\Big[\prod_{i\notin A} d_i - \prod_{i\notin A} c_i\Big] = 0.$$

Since the evidence is not decisive, $d_i \neq 0$. Further, $d_i = c_i + r_i$ so that $r_i = 0$ when $i \notin A$; otherwise, the expression $\prod d_i - \prod c_i$ makes a nonzero contribution to the above. Similarly, $p_i = 0$ when $i \notin A$; hence $c_i = 1$ is necessary. Let $A = \{i | c_i \neq 1\}$, then substitution shows $\mathrm{Bel}(A) = 1$ and A is clearly minimal.

Summary: The decision is the core when the evidence is decisive, otherwise $\{i | c_i \neq 1\}$ is the core.

$P^\star$ and Dou do not give us interesting concepts of decisiveness because $\mathrm{Dou}(A) = \mathrm{Bel}(\neg A) = 0$ would be the natural criterion. However this test is passed by any set in the complement of the core as well as others. Therefore, in general, no unique decision is found. A similar difficulty occurs in an attempt to form a concept of decisiveness in $P^\star$ because $P^\star(A) = 1 - \mathrm{Dou}(A)$.

6.4 Decisiveness in Q

Since $Q(\emptyset) = 1$ and $Q(A) \leq Q(B)$ when $B \subseteq A$, it is reasonable to ask for the maximal N such that $Q(N) = 1$. This set, N, is called the *nucleus* of Bel. If

$m(A) = 1$, then the decision, A, is clearly the nucleus. If $i \in N$, then $i \in A$ for all $m(A) \neq 0$. Further, $Q(\{i\}) = 1$ iff i is an element of the nucleus.

Assume that the simple evidence functions are not totally contradictory and there is no decision. Then $d_i \neq 0$ and there is no w such that $c_i = 1$ whenever $i \neq w$. The necessary and sufficient conditions, then, that $Q(\{z\}) = 1$, and hence $z \in N$ are (1) $p_i = 0$ if $i \neq z$ and (2) $c_z = 0$. To wit,

$$\begin{aligned}
Q(\{z\}) &= 1 \\
K(p_z + r_z)\prod_{i \neq z} d_i &= 1 \\
(p_z + r_z)\prod_{i \neq z} d_i &= K^{-1} \\
(p_z + r_z)\prod_{i \neq z} d_i &= \sum_{1 \leq q \leq n} p_q \prod_{i \neq q} d_i + \prod_{1 \leq i \leq n} d_i - \prod_{1 \leq i \leq n} c_i \\
\sum_{q \neq z} p_q \prod_{i \neq q} d_i + (d_z - r_z)\prod_{i \neq z} d_i - \prod_{1 \leq i \leq n} c_i &= 0 \\
\sum_{q \neq z} p_q \prod_{i \neq q} d_i + c_z \prod_{i \neq z} d_i - \prod_{1 \leq i \leq n} c_i &= 0 \\
\sum_{q \neq z} p_q \prod_{i \neq q} d_i + c_z \Bigl(\prod_{i \neq z} d_i - \prod_{i \neq z} c_i\Bigr) &= 0
\end{aligned}$$

Since $d_i \neq 0$, it follows that $p_q = 0$ for $q \neq z$, else the first term makes a nonzero contribution. Since $d_i = c_i + r_i$, the quantity, $\prod d_i - \prod c_i$, can vanish only if $r_i = 0$ when $i \neq z$. However, this and $p_i \neq 1$ because there is no decision, entails $c_i = 1$ when $i \neq z$. Therefore, either $\{z\}$ is the decision or the evidence is contradictory. Thus, $c_z = 0$ so that the second term of the last equation vanishes. Since the steps above are reversible, these are sufficient conditions too.

Summary: If A is the decision, then A is the nucleus. If two or more $p_i \neq 0$, then the nucleus is $\emptyset$. If $p_z \neq 0$, $c_z = 0$, and $p_i = 0$ when $i \neq z$, then $\{z\}$ is the nucleus. If $p_i = 0$ for all i, then $\{i | c_i = 0\}$ is the nucleus. Clearly, this construction can be carried out in $\mathrm{o}(n)$ time.

6.5 Discussion

It has been noted that $p_i = 1$ or $c_i = 1$ if and only it there is a j such that respectively $\mu_{ij}(\{i\}) = 1$ or $\nu_{ij}(\neg\{i\}) = 1$, i.e., if and only if the result of some experiment is decisive within its scope. The above analyses show the effects occurring when $p_i = 1$ or $c_i = 1$; subsets of possibilities are irrevocably lost—most or all the nondecisive evidence is completely suppressed—or the evidence becomes totally contradictory.

Any implementation of this theory should keep careful tabs on those conditions leading to conflict and/or decisiveness. In fact, any decisive experiment (a degree of support of 1) should be viewed as based upon evidence so conclusive that no further information can change one's view. A value of 1 in this theory is indeed a strong statement.

7 Conclusion

Dempster and Shafer's theory of plausible inference provides a natural and powerful methodology for the representation and combination of evidence. I think it has a proper home in knowledge-based expert systems because of the need for a technique to represent weights of evidence and the need for a uniform method with which to reason. This theory provides both. Standard statistical methods do not perform as well in domains where prior probabilities of the necessary exactness are hard to come by, or where ignorance of the domain model itself is the case. One should not minimize these problems even with the proposed methodology. It is hoped that with the ability to directly express ignorance and uncertainty, the resulting model will not be so brittle.

However, more work needs to be done with this theory before it is on a solid foundation. Several problems remain as obvious topics for future research. Perhaps the most pressing is that no effective decision making procedure is available. The Bayesian approach masks the problem when priors are selected. Mechanical operations are employed from gathering evidence through the customary expected-value analysis. But our ignorance remains hidden in the priors.

The Dempster-Shafer theory goes about things differently—ignorance and uncertainty are directly represented in belief functions and remain through the combination process. When it is time to make a decision, should the estimate provided by Bel or the one provided by $P^\star$ be used? Perhaps something in between. But what? No one has a good answer to this question.

Thus, the difference between the theories is that the Bayesian approach suppresses ignorance up front while the other must deal with it after the evidence is in. This suggests one benefit of the Dempster-Shafer approach: surely, it must be right to let the evidence narrow down the possibilities, first, then apply some ad hoc method afterward.

Another problem, not peculiar to this theory, is the issue of independence. The mathematical model assumes that belief functions combined by Dempster's rule are based upon independent evidence, hence the name orthogonal sum. When this is not so, the method loses its feeling of inevitability. Also, the elements of the frame of discernment, Θ, are assumed to be exclusive propositions. However, this is not always an easy constraint to obey. For example, in the MYCIN application, it seems natural to make the frame the set of possible infections but the patient can have multiple infections. Enlarging the frame to handle all subsets of the set of infections increases the difficulty in obtaining rules and in their application; the cardinality of the frame grows from $|\Theta|$ to $2^{|\Theta|}$.

One more problem that deserves attention is computational efficiency. Above it is shown that, with a certain set of assumptions, it is possible to calculate efficiently. However, these assumptions are not valid in all or even most domains. A thorough investigation into more generous assumptions seems indicated so that more systems can employ a principled reasoning mechanism.

The computational theory as presented here has been implemented in SIMULA. Listings are available by writing directly to the author.

References

1. J. A. Barnett, Computational Methods for a Mathematical Theory of Evidence: Part II. Forthcoming.
2. Jacob, *Ars Conjectandi*, 1713.
3. A. P. Dempster, "On direct probabilities," *J. Roy. Statist. Soc. Ser.* B 25, 1963, 102–107.
4. A. P. Dempster, "New methods for reasoning toward posterior distributions based on sample data," *Ann. Math. Statis.* 37, 1967, 355–374.
5. A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statis.* 38, 1967, 325–339.
6. A. P. Dempster, "Upper and lower probability inferences based on a sample from a finite univariant population," *Biometrika* 54, 1967, 515–528.
7. A. P. Dempster, "Upper and lower probabilities generated by a random closed interval," *Ann. Math. Statis.* 39, (3), 1968, 957–966.
8. A. P. Dempster, "A generalization of Bayesian inference," *J. Roy. Statis. Soc. Ser. B* 30, 1968, 205–247.
9. A. P. Dempster, "Upper and lower probability inferences for families of hypotheses with monotone density ratios," *Ann. Math. Statis.* 40, 1969, 953–969.
10. R. O. Duda, P. E. Hart, and N. J. Nilsson, *Subjective Bayesian methods for rule-based inference systems.* Stanford Research Institute, Technical Report 124, January 1976.
11. L. Friedman, Extended plausible inference. These proceedings.
12. T. Garvey, J. Lowrance, M. Fischler, An inference technique for integrating knowledge from disparate sources. These proceedings.
13. F. Hayes-Roth and V. R. Lesser, "Focus of Attention in the Hearsay–II Speech-Understanding System," in *IJCAI77*, pp. 27–35, Cambridge, MA, 1977.
14. H. E. Pople. Jr., "The formation of composite hypotheses in diagnostic problem solving: an exercise in synthetic reasoning," in *Proc. Fifth International Joint Conference on Artificial Intelligence*, pp. 1030–1037, Dept. of Computer Science, Carnegie-Mellon Univ., Pittsburgh, Pa., 1977.
15. G. Shafer, "A theory of statistical evidence," in W. L. Harper and C. A. Hooker (eds.), *Foundations and Philosophy of Statistical Theories in Physical Sciences*, Reidel, 1975.
16. G. Shafer, *A Mathematical Theory Of Evidence*, Princeton University Press, Princeton, New Jersey, 1976.
17. E. H. Shortliffe, *Artificial Intelligence Series*, Volume 2: *Computer-Based Medical Consultations: MYCIN*, American Elsevier, Inc., N. Y., chapter IV, 1976.
18. L. A. Zadeh, "Fuzzy sets," *Information and Control* 8, 1965, 338–353.