A REASONING MODEL BASED ON AN EXTENDED DEMPSTER-SHAFER THEORY *

John Yen

Computer Science Division
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720

ABSTRACT

The Dempster-Shafer (D-S) theory of evidence suggests a coherent approach to aggregate evidence bearing on groups of mutually exclusive hypotheses; however, the uncertain relationships between evidence and hypotheses are difficult to represent in applications of the theory. In this paper, we extend the multivalued mapping in the D-S theory to a probabilistic one that uses conditional probabilities to express the uncertain associations. In addition, Dempster's rule is used to combine belief update rather than absolute belief to obtain results consistent with Bayes' theorem. The combined belief intervals form probability bounds under two conditional independence assumptions. Our model can be applied to expert systems that contain sets of mutually exclusive and exhaustive hypotheses, which may or may not form hierarchies.

I INTRODUCTION

Evidence in an expert system is sometimes associated with a group of mutually exclusive hypotheses but says nothing about its constituents. For example, a symptom in CADIAG-2/RHEUMA (Adlassnig, 1985a)(Adlassnig, 1985b) may be a supportive evidence for rheumatoid arthritis, which consists of two mutually exclusive subclasses: seropositive rheumatoid arthritis and seronegative rheumatoid arthritis. The symptom, however, carries no information in differentiating between the two subclasses. Therefore, the representation of ignorance is important for the aggregation of evidence bearing on hypothesis groups.

Two previous approaches to the problem were based on Bayesian probability theory (Pearl, 1985) and the Dempster-Shafer (D-S) theory of evidence (Gordon and Shortliffe, 1985). While the Bayesian approach failed to express the impreciseness of its probability judgements, the D-S approach was not fully justified because of the difficulty to represent uncertain relationships between evidence and hypotheses in the D-S theory. As a result, the belief functions of the D-S approach are no longer probability bounds.

In this paper, we propose a reasoning model in which degrees of belief not only express ignorance but also forms interval probabilities. The multi-valued mapping in the D-S theory is first extended to a probabilistic one, so the uncertain relationships between evidence and hypothesis groups are described by conditional probabilities. The probability mass distribution induced from the mapping are then transformed to the *basic certainty assignment*, which measures belief update. Applying Dempster's rule to combine basic certainty assignments, we obtain the belief function that forms probability bounds under two conditional independence assumptions.

* This research was supported by National Science Foundation Grant ECS-8209679.

II TWO PREVIOUS APPROACHES

A. The Bayesian Approach

In a Bayesian approach presented by Judea Pearl (Pearl, 1986), the belief committed to a hypothesis group is always distributed to its constituents according to their prior probabilities. A point probability distribution of the hypothesis space thus is obtained. However, the distribution is much too precise than what is really known, and the ranges that the estimated probability judgements may vary are lost.

B. The Dempster—Shafer Approach

Jean Gordon and Edward Shortliffe have applied the D-S theory to manage evidence in a hierarchical hypothesis space (Gordon and Shortliffe, 1985), but several problems still exist. In order to define the terminology for our discussions, we describe the basics of the D-S theory before we discuss Gordon and Shortliffe's work.

1. Basics of the Dempster—Shafer Theory of Evidence

The Dempster-Shafer theory originated from the concept of the lower probabilities and the upper probabilities induced by a multivalued mapping (Dempster, 1967). Consider two spaces E and Θ and a multivalued mapping $\Gamma: E \rightarrow 2^{\Theta}$ that associate each element in E with a set of of elements in Θ. If the probability distribution in E is known, a basic probability assignment (bpa), denoted by $m: 2^{\Theta} \rightarrow [0, 1]$, is induced by the mapping. The basic probability value of a subset A of space Θ is**

$$m(A) = \sum_{\Gamma e_i = A} p(e_i) \tag{2.1}$$

The subset A is also called *focal element*. The space Θ is *the frame of discernment*. A legal bpa thus has the following properties.

$$\sum_{A \subset \Theta} m(A) = 1 \qquad m(\emptyset) = 0 \tag{2.2}$$

** For simplicity, we assume that Γ does not map any element of the space E to the empty set.

In general, the probability distribution of the space Θ is constrained by the bpa. The probability of a subset B of the frame of discernment is thus bounded below by the belief of B, denoted by Bel(B), and above by the plausibility of B, denoted by Pls(B). These two quantities are obtained from the bpa as follows:

$$Bel(B) = \sum_{A \subset B} m(A), \quad Pls(B) = \sum_{A \cap B \neq \emptyset} m(A). \tag{2.3}$$

Hence, the belief interval [Bel(B), Pls(B)] is the range of B's probability.

An important advantage of the D-S theory is its ability to express degree of ignorance. In the theory, the commitment of belief to a subset does not force the remaining belief to be committed to its complement, i.e., $Bel(B) + Bel(B^c) \leq 1$. The amount of belief committed to neither B nor B's complement is the degree of ignorance.

If m_1 and m_2 are two bpa's induced by two independent evidential sources, the combined bpa is calculated according to Dempster's rule of combination:

$$m_1 \oplus m_2(C) = \frac{\sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)} \tag{2.5}$$

2. Gordon and Shortliffe's Work

Gordon and Shortliffe (G-S) applied the D-S theory to combine evidence in a hierarchical hypothesis space, but they viewed MYCIN's CF as bpa without formal justification (Gordon and Shortliffe, 1985). As a result, the belief and plausibility in their approach were not probability bounds. Moreover, the applicability of Dempster's rule became questionable because it was not clear how one could check the independence assumption of Dempster's rule in the G-S approach.

The G-S approach also proposed an efficient approximation technique to reduce the complexity of Dempster's rule, but Shafer and Logan has shown that Dempster's rule can be implemented efficiently in a hierarchical hypothesis space (Shafer and Logan, 1985). Hence, the G-S's approximation technique is not necessary.

III A NEW APPROACH

A. An Extension to the Dempster–Shafer Theory

One way to apply the D-S theory to reasoning in expert systems is to consider the space E as an *evidence space* and the space Θ as a *hypothesis space*. An evidence space is a set of mutually exclusive outcomes (possible values) of an evidential source. For example, all possible results of a laboratory test form an evidence space because they are mutually exclusive. The elements of an evidence space are called *the evidential elements*. A hypothesis space is a set of mutually exclusive and exhaustive hypotheses. These hypotheses may or may not form a strict hierarchy.

The multivalued mapping in the D-S theory is a collection of conditional probabilities whose values are either one or zero. Suppose that an evidential element e_1 is mapped to a hypothesis group S_1. This implies that if e_1 is known with certainty, the probability of A_1 is one and the probability of A_1^c is zero, i.e., $P(A_1 \mid e_1) = 1$ and $P(A_1^c \mid e_1) = 0$. However, the mapping fails to express uncertain relationships such as "the probability of the hypothesis A is 0.8 given the evidence e". In order to represent this kind of uncertain knowledge, we extend the multivalued mapping to a probabilistic multi-set mapping.

A probabilistic multi-set mapping from an evidence space to a hypothesis space is a function that associates each evidential element to a collection of non-empty disjoint hypothesis groups accompanied by their conditional probabilities. A formal definition is given below.

Definition 1: A probabilistic multi-set mapping from a space E to a space Θ is a function $\Gamma^*: E \rightarrow 2^{2^\Theta \times [0,1]}$. The image of an element in E, denoted by $\Gamma^*(e_i)$, is a collection of subset-probability pairs, i.e.,

$\Gamma^*(e_i) = \{(A_{i_1}, P(A_{i_1} \mid e_i)) \cdots (A_{i_m}, P(A_{i_m} \mid e_i))\}$,

that satisfies the following conditions:

(1) $A_{i_j} \neq \emptyset \quad j = 1, \ldots, m$

(2) $A_{i_j} \cap A_{i_k} = \emptyset, \ j \neq k$

(3) $P(A_{i_j} \mid e_i) > 0, \quad j = 1, \ldots, m$

(4) $\sum_{j=1}^{m} P(A_{i_j} \mid e_i) = 1$

where e_i is an element of E, $A_{i_1}, \ldots, A_{i_m}$ are subsets of Θ,

For the convenience of our discussion, we introduce the following terminology. A *granule* is a subset of the hypothesis space Θ that is in the image of some evidential elements under the mapping. The *granule set* of an evidential element, denoted by G, is a set of all the granules associated with that element. For example, the granule set of e_i in the definition is the set of $A_{i_1}, \ldots, A_{i_m}$, i.e., $G(e_i) = \{A_{i_1}, \ldots, A_{i_m}\}$. The focal element in the D-S theory is the union of the granules in a granule set; moreover, because these granules are mutually exclusive, they form a partition of the focal element.

Since the mapping in the D-S theory has been extended to a probabilistic one, the probability mass of an evidential element e_i is now distributed among its granules. More precisely, the portion of e_i's probability mass assigned to its granule A is the product of the conditional probability $P(A \mid e_i)$ and the mass $P(e_i \mid E')$. Thus, the basic probability value of the granule A is the total mass assigned to it by all the evidential elements whose granule sets contain A.

Definition 2: Given a probabilistic multi-set mapping from an evidence space E to a hypothesis space Θ and a probability distribution of the space E, a mass function m is induced:

$$m(A \mid E') = \sum_{\substack{i \\ A \in G(e_i)}} P(A \mid e_i) P(e_i \mid E') \tag{3.1}$$

where E' denotes the background evidential source.

The mass function defined satisfies the properties of bpa described in (2.2). In fact, the bpa in the D-S theory (2.1) is a special case of our mass distribution with all conditional probabilities being either zero's or one's.

The belief and the plausibility obtained from our mass function bound the posterior probability under the conditional independence assumption that given the evidence, knowing its evidential source does not affect our belief in the hypotheses, i.e., $P(A \mid e_i, E') = P(A \mid e_i)$.

Lemma 1: If we assume that $P(A \mid e_i, E') = P(A \mid e_i)$ for any evidential element e_i and its granule A, then for an arbitrary subset B of the hypothesis space, we have $Bel(B \mid E') \leq P(B \mid E') \leq Pls(B \mid E')$.
(The proofs have been relegated to Appendix)

If all the granule sets of all evidential element are identical for a mapping, the basic probability value of a granule is not only its belief but also its plausibility. In particular, If all the granules are singletons, then the mass function determines a Bayesian Belief Function (Shafer, 1976).

Lemma 2: If $G(e_i) = G(e_j)$ for all $e_i, e_j \in E$, then for any granule A, we have $m(A \mid E') = Bel(A \mid E') = Pls(A \mid E') = P(A \mid E')$.

B. Combination of Evidence

In the Dempster-Shafer Theory, bpa's are combined using Dempster's rule; nevertheless, using the rule to combine our mass distributions will overweigh the prior probability as shown in the following example.

Example 1: E1 and e2 are two pieces of independent evidence bearing on the same hypothesis group A. If both e1 and e2 are known with certainty, each of them will induce a mass distribution from Definition 2:

$$m(A \mid e1) = P(A \mid e1),\ m(A^c \mid e1) = P(A^c \mid e1) \quad \text{and}$$
$$m(A \mid e2) = P(A \mid e2),\ m(A^c \mid e2) = P(A^c \mid e2).$$

The combined belief in A using Dempster's rule is

$$Bel(A \mid e1, e2) = \frac{P(A \mid e1) \times P(A \mid e2)}{P(A \mid e1)P(A \mid e2) + P(A^c \mid e1)P(A^c \mid e2)}$$
$$= \frac{P(e1 \mid A)P(e2 \mid A)P(A)^2}{P(e1 \mid A)P(e2 \mid A)P(A)^2 + P(e1 \mid A^c)P(e2 \mid A^c)P(A^c)^2}$$

Because both P(A|e1) and P(A|e2) are affected by the prior probability of A, the effect of the prior is doubled in the combined belief. In fact, the more evidential sources are combined, the bigger is the weight of the prior in the combined belief. Even if e1 and e2 are assumed to be conditionally independent on A, the combined belief could not be interpreted as lower probability. The D-S theory does not have such problem because its bpa does not count prior belief.

In order to combine our mass distributions, we define a quantity called *basic certainty value*, denoted by C, to discount the prior belief from the mass distribution. The basic certainty value of a hypothesis subset is the normalized ratio of the subset's mass to its prior probability*, i.e.,

$$C(A \mid E') = \frac{\frac{m(A \mid E')}{P(A)}}{\sum_{A \subset \Theta} \frac{m(A \mid E')}{P(A)}} \qquad (3.2)$$

Hence, any basic probability assignment can be transformed to a basic certainty assignment (bca) using the equation above. Intuitively, the basic certainty value measures the belief update, while both the bpa and the belief function measure absolute belief. Since both the CF in MYCIN (Shortliffe and Buchanan, 1975) and the likelihood ratio in PROSPECTOR (Duda, 1976) measure belief update, we may expect a relationship among them. In fact, as shown in the section IV-B, the probabilistic interpretations of CF given by Heckerman (Heckerman, 1985) are functions of basic certainty values.

* A special case of the basic certainty value is the belief measure B(h,e) in (Grosof, 1985).

Theorem 1: Consider two evidential spaces E_1 and E_2 that bear on a hypothesis space Θ. E_{1i} and $e2_j$ denote elements in E_1 and E_2. A_k and B_l denote granules of $e1_i$ and $e2_j$ respectively. Assuming that

$$P(e1_i \mid A_k)P(e2_j \mid B_l) = P(e1_i, e2_j \mid A_k \cap B_l) \quad A_k \cap B_l \neq \emptyset \qquad (3.3)$$

and

$$P(E_1' \mid e1_i)\, P(E_2' \mid e2_j) = P(E_1', E_2' \mid e1_i, e2_j) \qquad (3.4)$$

then

$$\frac{\sum_{A_k \cap B_l = D} C(A_k \mid E_1')\, C(B_l \mid E_2')}{\sum_{A_k \cap B_l \neq \emptyset} C(A_k \mid E_1')C(B_l \mid E_2')} = C(D \mid E_1', E_2') \qquad (3.5)$$

where E_1' and E_2' denote the evidential sources of the space E_1 and the space E_2 respectively.

Proof of Theorem 1 can be found in (Yen, 1985).

Based on Theorem 1, we apply Dempster's rule to combine basic certainty assignments. The aggregated bca can be further combined with other independent bca's. To obtain the updated belief function, the aggregated bca is transformed to the aggregated bpa through the following equation:

$$m(A \mid E') = \frac{C(A \mid E')P(A)}{\sum_{A \subset \Theta} C(A \mid E')P(A)} \qquad (3.6)$$

From Lemma 1, the belief and the plausibility of a hypothesis subset obtained from the updated bpa are lower probability and upper probability of the subset given the aggregated evidence.

In summary, combination of evidence is performed by first transforming bpa's from independent sources of evidence into bca's which are then combined using Dempster's rule. The final combined bca is transformed to a combined bpa, from which we obtain the updated belief function that forms interval probabilities.

C. Independence Assumptions of the Combining Rule

The two conditions assumed in Theorem 1 correspond to conditional independence of evidence and the independence assumption of Dempster's rule. In fact, the first assumption (3.3) is weaker than the strong conditional independence assumption employed in MYCIN and PROSPECTOR. The second assumption (3.4) is implicitly made in these systems.

1. The First Assumption

The first assumption (3.3) describes the conditional independence regarding the two evidence spaces and the hypothesis space. Sufficient conditions of the assumption are

$$P(e1_i \mid A_k) = P(e1_i \mid A_k \cap B_l),\ P(e2_j \mid B_l) = P(e2_j \mid A_k \cap B_l) \qquad (3.7)$$

and

$$P(e1_i \mid A_k \cap B_l) P(e2_j \mid A_k \cap B_l) = P(e1_i, e2_j \mid A_k \cap B_l). \quad (3.8)$$

The condition (3.7) is the conditional independence assumption

$$P(e \mid A, A_s) = P(e \mid A) \quad A_s \subset A$$

stating that if A is known with certainty, knowing its subset does not change the likelihood of e. A similar assumption is made in the Bayesian approach of (Pearl, 1986). The Bayesian approach applies the assumption to distribute the subset's mass to each of the subset's constituents. In our approach, however, the assumption is applied only when two bodies of evidence are aggregated to give support to a more specific hypothesis group. The assumption (3.7) is a consequence of the aggregation of evidence, not a deliberate effort to obtain a point distribution.

The equation (3.8) states that elements of different evidential sources are conditionally independent on their granules' non-empty intersections. Since the granules of an evidential element are disjoint, the intersections of two granule sets are also disjoint. Hence, two evidential elements of different sources are conditionally independent on a set of **mutually disjoint** hypothesis groups. In particular, pieces of evidence are not assumed to be conditionally independent on single hypotheses and their negations (complements) because generally they are not mutually disjoint. Therefore, the equation (3.7) is weaker than PROSPECTOR and MYCIN's assumption that pieces of evidence bearing on the same hypothesis are conditionally independent on the hypothesis and its negation. As a result, we solve their inconsistency problems dealing with more than two mutually exclusive and exhaustive hypotheses (Heckerman, 1985)(Konolige, 1979).

2. The Second Assumption

The second assumption (3.4) describes the conditional independence regarding the two evidence spaces and their background evidential sources. Sufficient conditions of the assumption (3.4) are

1. The probability distribution of the space E_2 conditioned on the evidence in E_1 is not affected by knowing the evidential source of E_1.

$$P(e2_j \mid e1_i) = P(e2_j \mid e1_i, E_1') \quad (3.10)$$

2. Similarly, the distribution of the space E_1 conditioned on the evidence in E_2 is not affected by knowing E_2'.

$$P(e1_i \mid e2_j) = P(e1_i \mid e2_j, E_2'). \quad (3.11)$$

3. The evidential sources E_1 and E_2 are conditionally independent on the joint probability distribution of $E_1 \times E_2$.

$$P(E_1' \mid e1_i, e2_j, E_2') = P(E_1' \mid e1_i, e2_j) \quad (3.12)$$

The assumption (3.4) corresponds to the independence assumption of Dempster's rule (Dempster, 1967),

$$P(e1_i \mid E_1') P(e2_j \mid E_2') = P(e1_i, e2_j \mid E_1', E_2'), \quad (3.13)$$

because (3.4) can be reformulated as

$$\frac{P(e1_i \mid E_1') P(e2_j \mid E_2') P(E_1') P(E_2')}{P(e1_i) P(e2_j)} = \frac{P(e1_i, e2_j \mid E_1', E_2') P(E_1', E_2')}{P(e1_i, e2_j)}. \quad (3.14)$$

The Dempster's independence assumption differs from (3.14) in that it does not contain prior probabilities. This difference is understood because in the D-S theory there is no notion of posterior versus prior probability in the evidence space. Therefore (3.4) intuitively replaces the independence of evidential sources assumed in Dempster's rule of combination.

The assumption is always satisfied when evidence is known with certainty. For example, if $e1_1$ and $e2_3$ are known with certainty, the equation (3.4) then becomes

$$P(e1_1 \mid e1_i) P(e2_3 \mid e2_j) = P(e1_1, e2_3 \mid e1_i, e2_j)$$

Both the left hand side and the right hand side of the equation above are zeros for all values of i and j except when $i=1$ and $j=3$ in which case both sides are one. Therefore, the equality holds. It is also straightforward to prove Theorem 1 without (3.4) assuming that evidence is known with certainty.

PROSPECTOR and Heckerman's CF model made similar assumptions in the combining formula:

$$\frac{P(E_1', E_2' \mid h)}{P(E_1', E_2' \mid \bar{h})} = \frac{P(E_1' \mid h)}{P(E_1' \mid \bar{h})} \frac{P(E_2' \mid h)}{P(E_2' \mid \bar{h})}.$$

Hence, we are not adding any assumptions to those of PROSPECTOR and MYCIN.

D. An Example

Suppose h_1, h_2, h_3, and h_4 are mutually exclusive and exhaustive hypotheses. Thus, they constitute a hypothesis space Θ. The prior probabilities of the hypotheses are $P(h_1) = 0.1$, $P(h_2) = 0.4$,$P(h_3) = 0.25$, and $P(h_4) = 0.25$. Two pieces of evidence collected are e_1 and e_2. E_1 strongly supports the hypothesis group $\{h_1, h_2\}$, with the following probability values:

$$P(\{h_1, h_2\} \mid e_1) = 0.9, \quad P(\{h_3, h_4\} \mid e_1) = 0.1.$$

E_2 supports h_1 with the following probability values while its negation gives no information:

$$P(h_1 \mid e_2) = 0.67, P(\{h_2, h_3, h_4\} \mid e_2) = 0.33, \text{ and } P(\Theta \mid \bar{e_2}) = 1$$

Suppose that e_1 is known with certainty, and e_2 is likely to be present with probability 0.3 (i.e., $P(e_1 \mid E_1') = 1$, $P(e_2 \mid E_2') = 0.3$, where E_1' and E_2' denote background evidential sources for e_1 and e_2 respectively). Although we have not assumed the prior probability of e_2, it is easy to check that a consistent prior for e_2 must be less then 0.14925. Therefore, e_2 with a posterior probability of 0.3 is still a piece of supportive evidence for h_1. The effect of e_1 on the belief in the hypotheses is represented by the following mass distribution:

$$m(\{h_1, h_2\} \mid E_1') = 0.9 \quad m(\{h_3, h_4\} \mid E_1') = 0.1$$

and m is zero for all other subsets of Θ. The corresponding basic certainty assignment (bca) is

$$C(\{h_1, h_2\} \mid E_1') = 0.9, \quad C(\{h_3, h_4\} \mid E_1') = 0.1.$$

Similarly, the effect of e_2 on the belief in the hypotheses is represented by the following mass distribution:

$$m(\{h_1\} \mid E_2') = 0.201 \quad m(\{h_1\}^c \mid E_2') = 0.099$$

$$m(\Theta \mid E_2') = 0.7$$

and m is zero for all other subsets of Θ. The mass distribution is transformed to the following bca:

$$C(\{h_1\} \mid E_2') = 0.7128 \quad C(\{h_1\}^c \mid E_2') = 0.039$$

$$C(\Theta \mid E_2') = 0.2482$$

Using Dempster's rule to combine the two bca's, we get the following combined bca:

$C(\{h_1\} \mid E_1'E_2') = 0.6907 \quad C(\{h_2\} \mid E_1'E_2') = 0.0378$

$C(\{h_1, h_2\} \mid E_1'E_2') = 0.2406 \quad C(\{h_3, h_4\} \mid E_1'E_2') = 0.0309.$

From the combined bca, we obtain the following combined bpa:

$m(\{h_1\} \mid E_1'E_2') = 0.314 \quad m(\{h_2\} \mid E_1'E_2') = 0.069$

$m(\{h_1, h_2\} \mid E_1'E_2') = 0.547 \quad m(\{h_3, h_4\} \mid E_1'E_2') = 0.07$

and m is zero for all other subsets of Θ. The belief intervals of the hypotheses are h_1: [0.314, 0.861], h_2: [0.069, 0.616], h_3: [0, 0.07], and h_4: [0, 0.07]. These intervals determine the following partial ordering: h_1 is more likely than h_3 and h_4, and h_2 is incomparable with h_1, h_3, and h_4. However, the Bayesian approach (Pearl, 1986) yields a different result: Bel(h_1) = 0.424, Bel(h_2) = 0.506, Bel(h_3) = 0.035, and Bel(h_4) = 0.035. The posterior probability of h_2 is higher than that of h_1 because majority of the mass assigned to the hypothesis group $\{h_1, h_2\}$ is allocated to h_2 for its relatively high prior probability.

IV COMPARISONS

A. Relationship to Bayes' Theorem

The result of our model is consistent with Bayes' theorem under conditional independence assumption. To show this, we consider n evidential sources $E_1, E_2, \ldots, E_n$ bearing on a hypothesis space $\Theta = \{h_1, h_2, \ldots, h_m\}$. The values of each evidential sources are known with certainty to be $e_1, e_2 \cdots e_n$ respectively. Also, the granules for every evidential sources are all singletons. It then follows from Lemma 2 that

$$m(\{h_i\} \mid e_1,e_2, \cdots e_n) = Bel(\{h_i\} \mid e_1,e_2, \cdots e_n) = P(h_i \mid e_1,e_2, \cdots e_n) \quad (4.1)$$

The basic probability assignment due to the evidential source E_j is

$$m(\{h_i\} \mid e_j) = P(h_i \mid e_j), \; i = 1, \ldots, m.$$

The corresponding basic certainty assignment is

$$C(\{h_i\} \mid e_j) = \frac{\frac{P(h_i \mid e_j)}{P(h_i)}}{\sum_k \frac{P(h_k \mid e_j)}{P(h_k)}} = \frac{P(e_j \mid h_i)}{\sum_k P(e_j \mid h_k)}, \; i = 1, \ldots, m.$$

Combining the basic certainty assignments from n evidential sources, we get

$$C(\{h_i\} \mid e_1,e_2, \cdots e_n) = \frac{P(e_1 \mid h_i)P(e_2 \mid h_i) \cdots P(e_n \mid h_i)}{\sum_i P(e_1 \mid h_i)P(e_2 \mid h_i) \cdots P(e_n \mid h_i)}$$

Through the transformation (3.6), we obtain the combined basic probability assignment:

$$m(\{h_i\} \mid e_1,e_2, \cdots e_n) = \frac{P(e_1 \mid h_i)P(e_2 \mid h_i) \cdots P(e_n \mid h_i)P(h_i)}{\sum_i P(e_1 \mid h_i)P(e_2 \mid h_i) \cdots P(e_n \mid h_i)P(h_i)} \quad (4.2)$$

From the equations (4.1) and (4.2), we get Bayes' theorem under the assumption that $e_1, e_2, \ldots, e_n$ are conditionally independent on each hypothesis in Θ.

B. Mapping Basic Certainty Assignment to CF

The probabilistic interpretations of certainty factor (CF) given by Heckerman (Heckerman, 1985) are functions of basic certainty values. One of Heckerman's formulations for the CF of a hypothesis h given a piece of evidence e is

$$CF(h,e) = \frac{\lambda - 1}{\lambda + 1}$$

where λ is the likelihood ratio defined to be

$$\lambda = \frac{P(e \mid h)}{P(e \mid \bar{h})}$$

In our model, when the frame of discernment contains only two hypotheses (i.e., $\Theta = \{h, \bar{h}\}$), and a piece of evidence e is known with certainty, the basic certainty assignment of Θ is:

$$C(\{h\} \mid e) = \frac{\lambda}{\lambda + 1} \quad C(\{\bar{h}\} \mid e) = \frac{1}{\lambda + 1}.$$

Therefore, one of Heckerman's probabilistic interpretations of CF is the difference of $C(\{h\} \mid e)$ and $C(\{\bar{h}\} \mid e)$ in this case. Moreover, the relationship can be comprehended as follows:

(1) If the basic certainty values of the hypothesis h and its negation are both 0.5, no belief update occurs. Hence, the certainty factor CF(h,e) is zero.

(2) On the other hand, if basic certainty value of the hypothesis is greater than that of its negation, degree of belief in h is increased upon the observation of the evidence. Thus the certainty factor CF(h,e) is positive.

In general, the probabilistic interpretations of CF are functions of $C(\{h\} \mid e)$.

A similar mapping between Heckerman's CF and a "belief measure" B to which Dempster's rule applies was found by Grosof (Grosof, 1985). In fact, Grosof's belief measure B(h,e) is equivalent to basic certinty value $C(\{h\} \mid e)$ in this special case. However, the distinction between belief update and absolute belief was not made in Grosof's paper. Thus, our approach not only generalized Grosof's work but also distinguishes basic certainty assignments from mass distributions in a clear way.

V CONCLUSIONS

By extending the D-S theory, we have developed a reasoning model that is consistent with Bayes' theorem with conditional independence assumptions. The D-S theory is extended to handle the uncertainty associated with rules. In addition, the Dempster rule is used to combine belief update rather than absolute belief, and the combined belief and plausibility are lower probability and upper probability respectively under two conditional independence assumptions.

The major advantage of our model over the Bayesian approach (Pearl, 1986) is the representation of ignorance. In our model, the amount of belief directly committed to a set of hypotheses is not distributed among its constituents until further evidence is gathered to narrow the hypothesis set. Therefore, degree of ignorance can be expressed and updated coherently as the degree of belief does. In the Bayesian approach, the amount of belief committed to a hypothesis group is always distributed among its constituents.

Directions for future research are mechanism to perform chains of reasoning, computational complexity of the model, and decision making using belief intervals. In chaining, Definition 2 is no longer valid because the probability distribution of an evidence space may not be known exactly. Although the extension of the Definition can be straight forward, a justification similar to Theorem 1 is difficult to establish. The computational complexity of our model is dominated by that of Dempster's rule, so any efficient implementations of the rule greatly reduce the complexity of our model. Interval-based decision making has been discussed in (Loui, 1985) and elsewhere, yet the problem is not completely solved and needs further research.

The proposed reasoning model is ideal for the expert system applications that (1) contain mutually exclusive and exhaustive hypotheses, (2) provide the required probability judgements, and (3) satisfy the two conditional independence assumptions. The model is currently implemented in a medical expert system that diagnoses rheumatic diseases.

ACKNOWLEDGEMENTS

The author is indebted to Professor Zadeh for his encouragement and support. The author would also like to thank Dr. Peter Adlassnig for valuable discussions and his comments on the paper.

Appendix

Lemma 1: If we assume that $P(A \mid e_i, E') = P(A \mid e_i)$ for any evidential element e_i and its granule A, then for an arbitrary subset B of the hypothesis space, we have $Bel(B \mid E') \leq P(B \mid E') \leq Pls(B \mid E')$.

Proof: Let us consider the conditional probability of B, an arbitrary subset of the hypothesis space, given the evidence e_i. The conditional probabilities of e_i's granules contribute to $P(B \mid e_i)$ depending on their set relationships with B:

(1) If the granule is included in B, all its conditional probability must be assigned to $P(B \mid e_i)$.

(2) If the granule has non-empty intersection with B, but is not included in B, its conditional probability may or may not be assigned to $P(B \mid e_i)$.

(3) If the granule has no intersection with B, its conditional probability can not be assigned to $P(B \mid e_i)$.

Since e_i's granules are disjoint, the sum of the conditional probabilities of the first type granules is the lower bound of $P(B \mid e_i)$. Similarly, the sum of the conditional probabilities of the first type granules and the second type granules is the upper bound of $P(B \mid e_i)$. Thus, we get

$$\sum_{\substack{j \\ A_j \subset B \\ A_j \in G(e_i)}} P(A_j \mid e_i) \leq P(B \mid e_i) \leq \sum_{\substack{k \\ A_k \cap B \neq \emptyset \\ A_k \in G(e_i)}} P(A_k \mid e_i).$$

Since $P(e_i \mid E')$ is positive and the equation above holds for any evidential element e_i, we have

$$\sum_i \sum_{\substack{j \\ A_j \subset B \\ A_j \in G(e_i)}} P(A_j \mid e_i) P(e_i \mid E') \leq \sum_i P(B \mid e_i) P(e_i \mid E') \tag{A.2}$$

$$\leq \sum_i \sum_{\substack{k \\ A_k \cap B \neq \emptyset \\ A_k \in G(e_i)}} P(A_k \mid e_i) P(e_i \mid E') .$$

From the definition of belief function and Definition 2, we have

$$Bel(B \mid E') = \sum_{\substack{j \\ A_j \subset B}} m(A_j \mid E')$$

$$= \sum_{\substack{j \\ A_j \subset B}} \sum_{\substack{i \\ A_j \in G(e_i)}} P(A_j \mid e_i) P(e_i \mid E') \tag{A.3}$$

Similarly, from the definition of plausibility and Definition 2, we get

$$Pls(B \mid E') = \sum_{\substack{j \\ A_j \cap B \neq \emptyset}} m(A_j \mid E')$$

$$= \sum_{\substack{j \\ A_j \cap B \neq \emptyset}} \sum_{\substack{i \\ A_j \in G(e_i)}} P(A_j \mid e_i) P(e_i \mid E') \tag{A.4}$$

Also, from the assumption that $P(A \mid e_i, E') = P(A \mid e_i)$, we get

$$P(B \mid E') = \sum_i P(B \mid e_i) P(e_i \mid E') \tag{A.5}$$

It thus follows from (A.2), (A.3), (A.4), and (A.5) that

$$Bel(B \mid E') \leq P(B \mid E') \leq Pls(B \mid E') . \quad \blacksquare$$

Lemma 2: If $G(e_i) = G(e_j)$ for all $e_i, e_j \in E$, then for any granule A, we have $m(A \mid E') = Bel(A \mid E') = Pls(A \mid E') = P(A \mid E')$.

Proof:

Part 1

Assume $m(A \mid E') \neq Bel(A \mid E')$. (A.6)

From the definition of belief function and the nonnegativity of the basic probability values, we have

$m(A \mid E') < Bel(A \mid E')$.

Hence, there exists a subset B such that

$B \subset A, B \neq A$, and $m(B \mid E') > 0$. (A.7)

From the definition 2, we know B is in a granule set, denoted as $G(e_b)$. Since A is also a granule, we denote its granule set as $G(e_a)$. Since $G(e_a) = G(e_b)$ according to the assumption of this Lemma, A and B are in the same granule set. From the Definition 1 it follows that A and B are disjoint, which contradicts (A.7). Therefore, the assumption (A.6) fails, and we have proved by contradiction that

$m(A \mid E') = Bel(A \mid E')$

Part 2

Similarly, we assume

$m(A \mid E') \neq Pls(A \mid E')$. (A.8)

From the definition of plausibility function, we know

$m(A \mid E') < Pls(A \mid E')$

Hence, there exists a subset C such that

$C \cap A \neq \emptyset, C \neq A$, and $m(C \mid E) > 0$. (A.9)

Using the arguments of Part 1, C and A must be in the same granule set, therefore they are disjoint, which contradicts (A.9). Therefore, the assumption (A.8) fails, and we have proved by contradiction that

$m(A \mid E') = Pls(A \mid E')$.

Part 3

From Lemma 1 and the proof of previous two parts, we have

$Bel(A \mid E') = P(A \mid E') = Pls(A \mid E')$. $\blacksquare$

REFERENCES

[1] Adlassnig, K.-P. "Present State of the Medical Expert System CADIAG-2", *Methods of Information in Medicine*, 24 (1985) 13-20.

[2] Adlassnig, K.-P. "CADIAG: Approaches to Computer-Assisted Medical Diagnosis", *Comput. Biol. Med.*, 15:5 (1985) 315-335.

[3] Dempster, A. P. "Upper and Lower Probabilities Induced By A Multivalued Mapping", *Annals of Mathematical Statistics*, 38 (1967) 325-339.

[4] Duda, R. O., P. E. Hart, and N. J. Nilsson. "Subjective Bayesian Methods for Rule-Based Inference Systems", *Proceedings 1976 National Computer Conference*, AFIPS, 45 (1976) 1075-1082.

[5] Gordon J. and E. H. Shortliffe. "A Method for Managing Evidential Reasoning in a Hierarchical Hypothesis Space", *Artificial Intelligence*, 26 (1985) 323-357.

[6] Grosof, B. N. "Evidential Confirmation as Transformed Probability", In *Proceedings of the AAAI/IEEE Workshop on Uncertainty and Probability in Artificial Intelligence*, 1985, pp. 185-192.

[7] Heckerman, D. "A Probabilistic Interpretation for MYCIN's Certainty Factors", In *Proceedings of the AAAI/IEEE Workshop on Uncertainty and Probability in Artificial Intelligence*, 1985, pp. 9-20.

[8] Heckerman, D. "A Rational Measure of Confirmation", MEMO KSL-86-16, Department of Medicine and Computer Science, Stanford University School of Medicine, February 1986.

[9] Konolige, K. "Bayesian Methods for Updating Probabilities" Appendix D of "A computer-Based Consultant for Mineral Exploration", Final Report of Project 6415, SRI International, Menlo Park, California, 1982.

[10] Loui, R., J. Feldman, H. Kyburg. "Interval-Based Decisions for Reasoning Systems", In *Proceedings of the AAAI/IEEE Workshop on Uncertainty and Probability in Artificial Intelligence*, 1985, pp. 193-200.

[11] Pearl, J. "On Evidential Reasoning in a Hierarchy of Hypothesis", *Artificial Intelligence Journal*, 28:1 (1986) 9-16.

[12] Shafer, G. "Mathematical Theory of Evidence", Princeton University Press, Princeton, N.J., 1976.

[13] Shafer, G. and R. Logan. "Implementing Dempster's Rule For Hierarchical Evidence", Working Paper of School of Business, University of Kansas, 1985.

[14] Shortliffe E. H. and B. G. Buchanan. "A Model of Inexact Reasoning in Medicine", *Mathematical Biosciences*, 23 (1975) 351-379.

[15] Yen, J. "A Model of Evidential Reasoning in a Hierarchical Hypothesis Space", Report No. UCB/CSD 86/277, Computer Science Division (EECS), University of California, Berkeley, December 1985.

[16] Zadeh, L. A. "Fuzzy Sets and Information Granularity", In *Advances in Fuzzy Set Theory and Applications*, 1979, pp. 3-18.