IN PROCEEDINGS, INTERNATIONAL SYMPOSIUM ON ARTIFICIAL INTELLIGENCE, DEC. 7-11, 1992, CUNAC, MEXICO.

ON THE DECISION MAKING PROBLEM IN DEMPSTER-SHAFER THEORY

Yong Ma  
Department of Computer Science  
University of Illinois  
405 North Mathews Avenue  
Urbana, IL 61801, USA  
yongma@cs.uiuc.edu

John S. Chandler  
Office for Info. Management  
University of Illinois  
1407 West Gregory Drive  
Urbana, IL 61801, USA  
chandlerj@vmd.csso.uiuc.edu

David C. Wilkins  
Department of Computer Science  
University of Illinois  
405 North Mathews Avenue  
Urbana, IL 61801, USA  
wilkins@cs.uiuc.edu

ABSTRACT

A decision making procedure is an essential component of any evidential reasoning system under uncertainty. That is, after the system gathers and combines all of the available information, a decision has to be made as the final output (e.g., diagnosis.) In probability theory, for example, a final decision is the hypothesis with the highest probability. Unfortunately, in Dempster-Shafer theory, one has more difficulties deciding the final output since more uncertainty is retained, even in the combination process of evidence. As the matter of fact, the effective decision making procedure remains an open problem for Dempster-Shafer theory. It is this problem for Dempster-Shafer theory that constitutes the topic of this paper.

Dempster-Shafer theory is one of the major paradigms for reasoning under uncertainty and it has been actively and extensively explored in many perspectives [7,8,10,18]. There are also some work on moving the theory into practical uses [4,14,15]. Therefore, the framework to be proposed for the decision making procedure will be an important and useful ingredient of the use and development of the theory.

The remainder of the paper is organized as follows. After Dempster-Shafer theory is introduced in Section 2, the decision making problem is identified and discussed in Section 3. In Section 4, a few related work are reviewed and discussed. In Section 5, some formal relationships among hypotheses are presented and proved. In Section 6, the sufficient conditions for sound decision making are derived from the formal relationships among hypotheses and some intuitive observations about the decision making processes. When the sufficient conditions are not satisfied for some cases, a heuristic combination function is proposed to make the final choice. The properties; or forms; of these functions are discussed. From the theoretical derivations and the heuristic proposal, an effective decision making algorithm is given, along with some comments. Finally, in Section 7, some concluding remarks are made and possible future studies are identified.

INTRODUCTION

Reasoning under uncertainty has been widely and extensively investigated in artificial intelligence and other related fields. Researchers in these fields have developed many methods to represent uncertain knowledge and draw inferences from them. Among these methods are the certainty factor model in MYCIN [3], Bayesian probability theory (as in PROSPECTOR) [6], belief networks [12,13], Dempster-Shafer's evidence theory [9,16], and Zadeh's possibility theory (fuzzy logic) [20]. One of the key issues for all of these methods is determining a decision making procedure; that is, after an expert system gathers and combines all of the available information, a decision has to be made as the final output (e.g., diagnosis.) In probability theory, for example, a final decision is the hypothesis with the highest probability (given the unit utility function, e.g., in diagnosis.) Unfortunately, in Dempster-Shafer theory, one has more difficulties deciding the final output since more uncertainty is retained, even in the combination process of evidence. As the matter of fact, the

DEMPSTER-SHAFER THEORY

Let \( \Theta \) be a set of mutually exclusive and exhaustive propositions about a domain, e.g., a set of hypotheses for a diagnostic system. \( \Theta \) is called the frame of discernment. Let \( 2^\Theta \) denote the set of all subsets of \( \Theta \). Elements of \( 2^\Theta \), i.e., subsets of \( \Theta \), are the general propositions in the domain with which the theory is concerned. Three basic functions are to be defined: the basic probability assignment \( m \), the belief function \( Bel \), and the plausibility function \( Pl \). They all define a numerical quantity between 0 and 1 to indicate the degree of support provided to a proposition by the piece of available evidence; their domain is \( 2^\Theta \) and their range is \( [0,1] \).

A function \( m : 2^\Theta \rightarrow [0,1] \) is a basic probabil-

293
ity assignment if it satisfies
\[ m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \subseteq \Theta} m(A) = 1. \] (1)

The quantity, \( m(A) \), represents the exact belief committed in the proposition represented by \( A \). Therefore, if \( m(G) = 1 \) and for \( A \neq G, m(A) = 0, A \subseteq \Theta \) (2) then a piece of evidence is certain, in that the decision making output must be one of the elements in set \( G \). However, it does not necessarily say which element of \( G \), unless \( G \) is a singleton set. On the other hand, the case admitted in the

where \( \Theta \) represents a situation of total ignorance or total lack of knowledge.

The belief function \( Bel : 2^\Theta \rightarrow [0, 1] \) can be defined in terms of \( m \)
\[ Bel(A) = \sum_{B \subseteq A} m(B) \quad \text{for} \quad A \subseteq \Theta \] (3)

In fact, Shafer has shown that the basic probability assignment that produces a given belief function is unique and can be recovered from the belief function by the following formulae.
\[ m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B) \quad \text{for} \quad A \subseteq \Theta \] (4)

where \( |X| \) is the cardinality of the set \( X \). Therefore, there is a one-to-one correspondence between basic probability assignments and belief functions.

Finally the plausibility function \( Pl : 2^\Theta \rightarrow [0, 1] \) is defined to be
\[ Pl(A) = 1 - Bel(-A) \quad \text{for} \quad A \subseteq \Theta \] (5)

It can be shown that the plausibility function \( Pl \) carries exactly the same information as \( m \) and \( Bel \) do.

Now let us see what those quantities intuitively mean although they are all equivalent in terms of information. If we think of the elements as points, \( m(A), A \subseteq \Theta \), measures the total probability mass constrained to stay in \( A \) but not confined to any proper subset of \( A \). It represents our ignorance of not being able to subdivide our belief to any subsets of \( A \). Then the quantity, \( Bel(A) \), is the measure of the total probability mass constrained to stay somewhere in \( A \). Since \( Bel(-A) \) is the measure of the probability mass constrained to stay out of \( A \), \( Pl(A) = 1 - Bel(-A) \) is the measure of the total probability mass that can move into \( A \) (some of which is already in \( A \) if \( Bel(A) > 0 \).) To put it other way, \( Bel(A) \) is the measure of the lower probability of \( A \) and \( Pl(A) \) is the measure of the upper probability of \( A \). Hence, an interval notation, \([Bel(A), Pl(A)]\), is often used to represent the range of the probability of \( A \).

Next we look at how the various distinct pieces of evidence are combined, i.e., how the fundamental operation of uncertainty reasoning is accomplished in Dempster-Shafer theory. Let \( m_1 \) and \( m_2 \) be two basic probability assignments representing two uncertain pieces of evidence for the same frame of discernment, \( \Theta \). Dempster's combination rule computes a new basic probability assignment, denoted \( m_1 \oplus m_2 \), that represents the combined effect of \( m_1 \) and \( m_2 \) as follows.
\[ m(\emptyset) = 0 \]
\[ m(A) = K \sum_{X \cap Y = A} m_1(X) \cdot m_2(Y), A \subseteq \Theta, A \neq \emptyset \]
\[ K = \frac{1}{1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y)} \] (6)

The purpose of \( K \) is to redistribute the belief committed to \( \emptyset \) by the intersection operation to other non-empty non-zero belief subsets; namely; the normalization step; in order to make \( m_1 \oplus m_2 \) an eligible basic probability assignment. Therefore, if \( 1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y) = 0 \), then \( m_1 \oplus m_2 \) does not exist. This occurs when the combined belief functions invest no belief in intersecting sets, and \( m_1 \) and \( m_2 \) are said to be totally or flatly contradictory. The quantity \( \log(K) \) is called the weight of conflict between \( Bel_1 \) and \( Bel_2 \).

If \( m_1 \oplus m_2 \) exists for given \( m_1 \) and \( m_2 \), the corresponding belief function, \( Bel_1 \oplus Bel_2 \), and the plausibility function, \( Pl_1 \oplus Pl_2 \), are then easily computed from \( m_1 \oplus m_2 \) by the definitions.

An important property of Dempster's rule is that it is commutative and associative. This is desirable because evidence aggregation should be independent of the order of its gathering.

**THE DECISION MAKING PROBLEM FOR THE THEORY**

The problem of an effective decision making procedure for Dempster-Shafer theory has been recognized by several researchers:

"Dempster and Shafer's theory of plausible inference provides a methodology for the representation and combination of evidence. However, several questions need to be addressed before it can be used effectively for AI applications. Perhaps, the most pressing problem is that no effective decision making procedure is available." [2, p. 24].

"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." [1, p. 874].

The concern of the problem is that, when it comes to make a decision, which of the estimates provided by \( m, Bel, Pl \) should be used. If we have \( h_1 : [0.5, 0.8] \) and \( h_2 : [0.4, 0.9] \), which one of \( h_1 \) and \( h_2 \) should
be chosen as the final output? The difficulty of the decision stems from the fact that although $Bel(h_1)$ is greater than $Bel(h_2)$, $Pl(h_1)$ is less than $Pl(h_2)$. It is indeed a very important problem because an expert system has to make a decision at the end of information gathering and aggregation.

Why is it a problem for Dempster-Shafer theory? It is because Dempster-Shafer theory represents ignorance and uncertainty explicitly (that is, the interval representation) and retains them through the combination process. This may suggest one advantageous use of the Dempster-Shafer approach; that is, we initially retain ignorance. As the evidence narrows down the possibilities we may or may not have a decision. If not, we then apply some ad hoc selective methods to make our final choice. In this paper, we attempt to attack the problem by deriving the theoretical conditions under which a sound decision can be made and by proposing heuristic combination functions to accommodate those cases where the derived conditions are not satisfied.

RELATED WORK

Strat [19] studied the problem of decision analysis using belief functions for a specific type of the frame of discernment. The special requirements are that (1) the elements of the frame of discernment are scalar; (2) thus, the utility function is the magnitude of an element; (3) the concern is the total expected utility of some course of action; and (4) the probability that all residual ambiguity will turn out as favorably as possible ($p$) must be estimated. He defined an expected value interval (EVI) as:

$$E(x) = [E_+(x), E^*(x)]$$

where

$$E_+(x) = \sum_{A_i \subseteq \Theta} \min(A_i) m_\Theta(A_i)$$

$$E^*(x) = \sum_{A_i \subseteq \Theta} \max(A_i) m_\Theta(A_i)$$

And then the expected utility is given by

$$E(x) = E_+(x) + p[E^*(x) - E_+(x)]$$

Finally, the choice should be the one that maximizes the above expression.

Note, however, although he speaks of the interval representation, the interval in EVI does not come from a belief function itself. Instead it is from the particular utility function used, i.e., $\min(A_i)$ and $\max(A_i)$. For example, if the utility function were a unity (as in some diagnostic cases), then $E_+(x)$ would be equal to $E^*(x)$, and thus no interval would be produced regardless of the underlying belief function. Clearly, we address different problems with ours aiming at a more general solution.

Another approach is so-called the generalized insufficient reason principle advocated by [17] and [5]. By this principle, for each $A \subseteq \Theta$, $m(A)$ is equally distributed among all its singleton elements. Then, the decision making problem is considered according to probability theory and its associated utility theory. Because only the singleton hypotheses are considered, it is a special case of our general solution (to be shown later.)

A more closely related approach would be Lesh’s proposal [11]. In his method, an “ignorance preference coefficient” $\tau$ is empirically derived to compute the distinguished point called “expected evidential belief (EEB)”:  

$$EEB(A) = \frac{Bel(A) + Pl(A)}{2} + \tau \left[Pl(A) - Bel(A)\right]$$

(11)

Then a choice is made by selecting the action that maximizes the “expected evidential value (EEV)”:  

$$EEV = \sum_{A_i \subseteq \Theta} A_i EEB(A_i)$$

(12)

Once again, the total expected utility of some action is concerned; while we worry about how to select a single best hypothesis (a subset of $\Theta$) to be the output.

FORMAL RELATIONSHIPS AMONG HYPOTHESES

Before we can devise a decision making algorithm for Dempster-Shafer theory we need to define some terms and prove some preliminary formal relationships among hypotheses which are ordered by their current belief values.

Definition 1 A subset $A$ is said to be the largest belief subset if the following conditions are satisfied by $A$.

- $A$ is a proper subset of $\Theta$, that is, $A \subset \Theta$,
- $Bel(A) \geq Bel(X)$ for all $X \subset \Theta$.

Clearly, $\Theta$ would be the largest belief subset if $A$ were not defined to be a proper subset of $\Theta$. However, $\Theta$ is not an interesting candidate for decision since its selection is vacuous. Note that there may be more than one such subsets, in which case, an arbitrary one can be chosen.

Definition 2 A subset $A$ of $\Theta$ is said to be the largest belief with minimum element subset if

- $A$ is the largest belief subset;
- No proper subset of $A$ has the same belief value as $A$, that is, $Bel(A) > Bel(X)$ for all $X \subset A$.

Definition 3 A subset $A$ of $\Theta$ is said to be the largest belief with maximum element subset if

- $A$ is the largest belief subset;
- If there exists $A'$ such that $Bel(A') = Bel(A)$, then $|A| \geq |A'|$; where $|X|$ is the cardinality of set $X$.
From these definitions, the existence can be shown of certain relationships among these subsets.

Most Specific Hypotheses

Theorem 1. If A is the largest belief with minimum element subset of \( \Theta \), B the second largest belief with minimum element subset of \( \Theta \), and \( A \neq B \), then exactly one of the following relations holds.

1. \( B \subseteq A \).
2. \( A \cup B = \emptyset \).

Proof: By similar arguments, we would have

\[
Bel(A \cup B) \geq Bel(A).
\]  

If \( B \subseteq A \), then \( A \cup B \supseteq A \). But if \( A \cup B \neq \emptyset \), then \( A \cup B \) would be the largest belief with maximum element subset; a contradiction.

The claim that exactly one relation holds can also be analogously proved.

Most General Hypotheses

Theorem 2. If \( A \) is the largest belief with maximum element subset of \( \Theta \), B the second largest belief with maximum element subset of \( \Theta \), and \( A \neq B \), then exactly one of the following relations holds.

1. \( B \subseteq A \).
2. \( A \cup B = \emptyset \).

Proof: We will prove it by contradiction. Assume that the claim were not true, then we would have \( B \nsubseteq A \) and \( A \cup B \neq \emptyset \). From \( A \neq B \) and \( B \nsubseteq A \), we have \( A \cap B \subseteq B \). By the definition of \( Bel(X) \), \( Bel(B) \geq Bel(A \cap B) \). But \( Bel(B) \neq Bel(A \cap B) \), for, if they were equal, then \( A \cap B \) would be the second largest belief with minimum element subset by the definition, contrary to the assumption that \( B \) was. Therefore, \( Bel(B) > Bel(A \cap B) \).

On the other hand,

\[
Bel(A \cup B) = Bel(A) + Bel(B) - Bel(A \cap B) + \sum_{X \subseteq A \cup B, X \neq A, X \neq B} m(X)
\]

Since \( Bel(B) > Bel(A \cap B) \) and \( m(X) \geq 0 \), for all \( X \subseteq \Theta \), \( Bel(A \cup B) > Bel(A) \).

From \( A \neq B \) and \( B \nsubseteq A \), we also have \( A \cup B \nsubseteq A \). Together with \( A \cup B \neq \emptyset \), \( A \cup B \) should be the largest belief with minimum element subset, a contradiction.

Now what is left to be shown is that one and only one of the two relations holds. If \( B \subseteq A \), then \( A \cup B = A \). By definition of \( A \), \( A \neq \emptyset \), thus \( A \cup B \neq \emptyset \). If \( A \cup B = \emptyset \), then it must be the case that \( B \nsubseteq A \); for if not, then \( B \subseteq A \) and \( A \cup B = A \neq \emptyset \) by definition of \( A \), contrary to the assumption that \( A \cup B = \emptyset \). □

For the convenience later in the paper, the theorem is rewritten as the following corollary.

Corollary 1. If \( A \) is the largest belief with minimum element subset of \( \Theta \), \( B \) the second largest belief with minimum element subset of \( \Theta \), and \( A \neq B \), then exactly one of the following relations holds.

1. \( B \subseteq A \).
2. \( A \cap B = \emptyset \), and \( A \cup B = \emptyset \).
3. \( A \cap B \neq \emptyset \), and \( A \cup B = \emptyset \).

Proof: It is the immediate result of Theorem 1. □

Generalization

Theorem 3. Let \( A_k \) be the kth largest belief with minimum (maximum) element subset of \( \Theta \). Then \( \{A_0, A_1, A_2, \ldots, A_k\} \) is a sequence of hypotheses in descending order by their belief values; where \( A_0 = \emptyset \). Then, for any \( k \), \( 1 \leq k < l \), exactly one of the following relations holds.

1. \( A_{k+1} \subseteq A_k \).
2. \( A_k \cap A_{k+1} = \emptyset \), and \( A_k \cup A_{k+1} = A_j \), for some \( j \), \( 0 \leq j < k \).
3. \( A_k \cap A_{k+1} \neq \emptyset \), and \( A_k \cup A_{k+1} = A_j \), for some \( j \), \( 0 \leq j < k \).

Proof: The proof procedure is similar to the above. If none of the relations holds, \( A_k \cup A_{k+1} \), being different from \( A_k \), would be in the sequence before \( A_k \). □

An example will help to understand the essence of Theorem 3. Suppose that \( \Theta = \{a, b, c\} \) and that the combined basic probability assignment is as follows: \( m(\{a\}) = .3, m(\{c\}) = .2, m(\{a, b\}) = .1, m(\Theta) = .4, m(X) = .0 \), for all other \( X \subseteq \Theta \). Then, it is straightforward to calculate \( Bel(A) \) with \( A \subseteq \Theta \).
The results are shown in Fig. 1. Therefore, \([\mathcal{O}, \{a, c\}, \{a, b\}, \{a\}, \{c\}]\) is the sequence of the ordered hypotheses, according to the \(k\)th largest belief with minimum element subsets. Now, it is easy to check that, for any \(k (1 \leq k < 4)\), either \(A_{k+1} \subseteq A_k\), or there exists \(j\) such that \(0 \leq j < k\), and \(A_{k+1} \cup A_k = A_j\).

### THE DECISION MAKING PROCEDURE

In this section, we begin with setting up the problem and speculate some observations and intuitions behind decision making processes. Then, we move on to derive the sufficient conditions under which a sound decision can be made. For those cases where the conditions are not met, a heuristic combination function is proposed to make the final choice. The coefficients in the heuristic function can be learned to best fit the domain under investigation. Based on the theoretical derivations and the heuristic combination function, an algorithm for decision making is given. Finally, some comments concerning the algorithm and empirical experiments are made.

#### Problem Setting

Suppose that a utility function \(u(A)\) is defined over \(\mathcal{O}\). Then, the utility value interval is denoted by \([Bel_u(A), Pl_u(A)]\), where

\[
Bel_u(A) = Bel(A) \times u(A) \quad (17) \\
Pl_u(A) = Pl(A) \times u(A) \quad (18)
\]

However, for simple cases of diagnosis or classification tasks, a unity utility is often used although more complicated schemes are possible. For example, the pure task of classifying someone as American or English would give an equal weight to both status. Thus, we assume that \(u(A) = 1\) (or any constant), for the rest of the paper. Given this assumption, the notations \(m(A), Bel(A), Pl(A)\) are still used, but should be understood that the utility function is already incorporated.

Secondly, a problem control structure can be added for the problem at hand. It defines which hypotheses are of interests and which are not: e.g., only a tree (not a graph) \([9]\). Given this structure information, the \(m(A)\) of any hypothesis \(A\) not to be considered will be distributed according to the generalized insufficient reason principle. The remaining hypotheses are competing against each other. (Note that if the problem control structure restricts to the singleton hypotheses only, then the pure generalized insufficient reason principle is achieved.)

#### Intuitions And Observations

Suppose that, after gathering and aggregating available evidence, an expert system has the following quantities: \(m(h), Bel(h), Pl(h)\), for all \(h \subseteq \mathcal{O}\) if \(h_0\) is chosen as the final output, what can be said about the property or characteristics of \(h_0\) in terms of \(m, Bel, Pl\)? Since \(m, Bel, Pl\) are numerical values, it amounts to asking what relations should hold between the chosen hypothesis \(h_0\) and all other hypotheses expressed in \(m, Bel, Pl\). The following are some intuitive observations. (Note: since the interval notation is often, and easily used, the following relations do not use \(m(h)\).)

- If \(Bel(h_1) \geq Bel(h_2)\), then \(h_1\) is more probable than \(h_2\) as the final output.
- If \(Pl(h_1) \geq Pl(h_2)\), then \(h_1\) has more potential to become more probable than \(h_2\) as the final output.
- Thus, if both \(Bel(h_1) \geq Bel(h_2)\) and \(Pl(h_1) \geq Pl(h_2)\), then \(h_1\) should be preferred to \(h_2\) as the final output given the currently available evidence. Such a decision is termed to be a sound decision.

Unfortunately, though, it is not always possible to have both \(Bel(h_1) \geq Bel(h_2)\) and \(Pl(h_1) \geq Pl(h_2)\) satisfied for some \(h_1\) and \(h_2\). Thus, we will next derive the conditions that guarantee the satisfaction of both inequality relations. As it is known, every interval converges to a point, i.e., \(Bel(h) = Pl(h)\) if given complete information. (Note that the point case can be outside of the current interval, see comments later.) In this case, decision making reverts to probability theory.

Another intuition is that a more specific hypothesis \(A\) is always preferred to a more general hypothesis \(A'\) if they have the same probability interval, i.e., if \(Bel(A) = Bel(A')\), \(Pl(A) = Pl(A')\), and \(A \subseteq A'\). The reason is simple: the more specific the hypothesis, the more informative it is.

#### Sufficient Conditions For Sound Decision Making

Given the formal relationships among ordered hypotheses derived in the last section, we can show that certain numerical relations exist among them as well, under certain conditions. These conditions are then considered to be the sufficient conditions to make a sound decision, according to the above intuitions and observations.

**Theorem 4** If \(A\) is the largest belief with minimum (maximum) element subset of \(\mathcal{O}\), \(B\) the second largest belief with minimum (maximum) element subset of \(\mathcal{O}\), and \(A \neq B\), then

\[
Pl(A) \geq Pl(B) \quad (19)
\]

holds for the following two cases:

1. \(B \subseteq A\).
2. \(A \cap B = \emptyset\), \(A \cup B = \mathcal{O}\).

Proof: Each case is dealt with separately.
Case 1: \( B \subset A \).
From \( B \subset A \), we have \( \neg A \subset \neg B \). Thus, \( \text{Bel}(\neg A) \leq \text{Bel}(\neg B) \) from the definition of \( \text{Bel}(X) \). By the definition of \( \text{Pl}(X) \),
\[
\text{Pl}(A) = 1 - \text{Bel}(\neg A) \quad (20)
\]
\[
\text{Pl}(B) = 1 - \text{Bel}(\neg B) \quad (21)
\]
Thus, \( \text{Pl}(A) \geq \text{Pl}(B) \) follows from \( \text{Bel}(\neg A) \leq \text{Bel}(\neg B) \).

Case 2: \( A \cap B = \emptyset, A \cup B = \Theta \).
Then, \( \neg A = B \) and \( \neg B = A \). Thus,
\[
\text{Pl}(A) = 1 - \text{Bel}(\neg A) = 1 - \text{Bel}(B) \quad (22)
\]
\[
\text{Pl}(B) = 1 - \text{Bel}(\neg B) = 1 - \text{Bel}(A) \quad (23)
\]
Thus, \( \text{Bel}(A) \geq \text{Bel}(B) \) implies \( \text{Pl}(A) \geq \text{Pl}(B) \).

Heuristic Combination Functions For Decision Making

When the derived sufficient conditions among hypotheses are not met, it is not guaranteed that both \( \text{Bel}(h_1) \geq \text{Bel}(h_2) \) and \( \text{Pl}(h_1) \geq \text{Pl}(h_2) \) are satisfied at the same time (resulting in overlapping choices); and, thus, the final choice must be selected based on some heuristic combination function of \( \text{Bel}(h) \) and \( \text{Pl}(h) \).

There are several concerns about the properties or forms of heuristic combination functions.

1. **Rationality.** If both \( \text{Bel}(h_1) \geq \text{Bel}(h_2) \) and \( \text{Pl}(h_1) \geq \text{Pl}(h_2) \), then \( h_1 \) must be preferred to \( h_2 \) by the heuristic evaluation; that is, if \( f \) is a heuristic function, then it must be the case that \( f(h_1) \geq f(h_2) \).

2. **Support evidence.** Because that \( \text{Bel}(h) \) represents the supportive evidence whereas \( \text{Pl}(h) \) only represents the potentials (it also represents doubt in \( h \) since \( \text{Pl}(h) = 1 - \text{Bel}(\neg h) \)), the heuristic combination functions should give more weights to \( \text{Bel}(h) \) than to \( \text{Pl}(h) \), unless otherwise required by the domain and specified by the user.

3. **Simplicity.** A simple function \( f(h) \), e.g., a linear combination of \( \text{Bel}(h) \) and \( \text{Pl}(h) \), is preferred to a complex one, unless other information about the domain indicates otherwise.

Among these concerns, only the rationality requirement is strongly suggested and others are dependent upon the circumstances in the question.

The following is an example of a generic heuristic combination function which meets the above concerns.

\[
f(h) = c\text{Bel}(h) + (1 - c)\text{Pl}(h), \quad 0.5 < c \leq 1 \quad (26)
\]

Clearly, if both \( \text{Bel}(h_1) \geq \text{Bel}(h_2) \) and \( \text{Pl}(h_1) \geq \text{Pl}(h_2) \), \( f(h_1) \geq f(h_2) \). Thus, the rationality requirement is met. Since \( c > 0.5 \), \( \text{Bel}(h) \) gets more weight than \( \text{Pl}(h) \). And finally, it is a simple linear function. The coefficient \( c \) in this simple heuristic combination function may vary drastically for different domains; e.g., in medical diagnosis, one may want to assign a high value for \( c \).

Learning The Coefficients Of Heuristics

To improve diagnostic accuracy, the coefficients in the heuristic combination function can be adjusted through learning to best fit the domain under investigation. To do so, a set of solved cases is used to train the heuristic combination function so that the optimal values for the coefficients in the function are found. Constraints on the coefficients can be imposed before the training. For example, for the linear heuristic combination function above, a set of solved cases may help to find that the best value of \( c \) is 0.65.

The Decision Making Algorithm

The combination of the theoretical derivations and the generic heuristic proposal results in the following ef-
fective decision making algorithm for Dempster-Shafer theory. Note that the algorithm, by default, works on the most specific hypotheses, unless requested otherwise.

Algorithm:
1. If the problem control structure is given, distribute $m(A)$, for any $A$ not to be considered, to its singletons according to the generalized insufficient reason principle.
2. Sort out all the hypotheses according to their current belief values: $Bel(h)$.
3. If most general hypotheses preferred, then arrange the hypotheses according to their generality; and denote the final sorted list by $L = [A_1, A_2, ..., A_l]$.
4. Otherwise, (i.e., most specific hypothesis preferred), then eliminate all $A'$ if there exists $A$ such that $Bel(A) = Bel(A')$ and $A \subseteq A'$; and denote the final sorted list by $L = [A_1, A_2, ..., A_l]$.
5. If $l = 1$, then output $A_1$ as the final choice; and stop.
6. Test the sufficient conditions; if any of the sufficient conditions is satisfied, then let $T = A_{l-1}$. If $l = 2$, then output $A_1$ (a sound decision) and stop.
7. Otherwise, let $T = Heuristics.evaluation(f, A_{l-1}, A_l)$.
8. Remove $A_{l-1}$ and $A_l$ from $L$, and append $T$ to $L$ at the end.

Procedure Heuristics.evaluation($f, h_1, h_2$)
/* $f$ is a user-supplied heuristic combination function */
If $f(h_1) \geq f(h_2)$ then
return $h_1$
else return $h_2$

Now let us see how the algorithm correctly and automatically takes care of all the cases in Theorem 3. If the heuristic combination function in procedure Heuristics.evaluation() has the rationality property, then, whenever both $Bel(h_1) \geq Bel(h_2)$ and $Pl(h_1) \geq Pl(h_2)$, $h_1$ will prevail even if the sufficient conditions are not satisfied. It means that the rationality property of these functions captures all the cases which the sufficient conditions miss.

Let us apply the algorithm to the example in Fig. 1. First, assume that we use the linear combination function defined in (26) with $c = 0.6$. That is,
$$f(h) = 0.6 \times Bel(h) + 0.4 \times Pl(h)$$ (27)
Then the computation of various quantities is carried out and the results are shown in Table 1.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$m(h)$</th>
<th>$Bel(h)$</th>
<th>$Pl(h)$</th>
<th>$f(h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${a, c}$</td>
<td>0.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.70</td>
</tr>
<tr>
<td>${a, b}$</td>
<td>0.1</td>
<td>0.4</td>
<td>0.8</td>
<td>0.56</td>
</tr>
<tr>
<td>${a}$</td>
<td>0.3</td>
<td>0.3</td>
<td>0.8</td>
<td>0.50</td>
</tr>
<tr>
<td>${c}$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.6</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Going through the algorithm with these numbers in Table 1 results in $\{a, c\}$ being the final output. Note that $\{a, c\}$ and $\{a, b\}$ do not satisfy the sufficient conditions; but, since $Bel(\{a, c\}) > Bel(\{a, b\})$ and $Pl(\{a, c\}) > Pl(\{a, b\})$, $\{a, c\}$ is the preferred choice by the heuristic evaluation. This is an instance where the rationality property of the heuristic combination function captures what the sufficient conditions missed.

Comments And Experiments

One objection may be that $[Bel(A), Pl(A)]$ should be not used as probability bounds because $[Bel(A), Pl(A)]$ will be changed given more evidence. First, we do not admit that $[Bel(A), Pl(A)]$ is the bounds of the true probability of $A$. Second, the same is also true for probability of $A$ in probability theory. That is, $P(A)$ will be changed given more evidence. However, a decision can and has to be made based on all what is available now but not on what will be available in the future. Therefore, using the current interval $[Bel(A), Pl(A)]$ to make a decision is legitimate and in fact the only choice, just like the way a decision has to be made in probability theory given all what is available.

To briefly summarize, the contributions of the theoretical derivations of various results and the proposed framework to set up heuristic combination functions are several-fold. First, the sufficient conditions precisely identify the cases when a decision making is sound. If such a case arises in practice, one can have complete confidence believing the final output (Step 6). Without such theoretical results, one can never be sure. Heuristics alone can never guarantee anything. Second, the sufficient conditions directly relate to and indirectly lead to the rationality requirement for heuristic combination functions. Third, some formal relationships among hypotheses are made clear and explicit by various theorems. Those relationships among hypotheses are certainly the basis for further investigation of this problem and some other problems of the theory. Forth, the heuristic proposed provides a flexible framework for setting up heuristic combination functions to meet different needs. Fifth, the decision making algorithm is easy to implement in practical systems. Some projects are being planned to make use of this decision making algorithm for various domains.
CONCLUDING REMARKS

In this paper, we addressed an important problem for reasoning under uncertainty in Dempster-Shafer framework—the decision making procedure. It is important because an expert system has to make a final decision at the end of information gathering and combination. After Dempster-Shafer theory was introduced, the open problem of effective decision making procedure for the theory was then identified and discussed. To propose solutions to the problem, some formal relationships among ordered hypotheses were presented and proved. Based on these formal relationships among hypotheses, the sufficient conditions for sound decision making were derived. In case of the sufficient conditions not being met, heuristic combination functions were proposed to make the final choice. The properties or forms of these functions were emphasized. From the theoretical results and the heuristic proposed, an effective decision making algorithm for Dempster-Shafer theory was devised.

For future studies, we are to apply this decision making algorithm to various domains and to propose specific heuristic combination functions for them. Doing so may allow us to further characterize the heuristic combination functions in a greater detail.

References


