A SET-THEORETIC FRAMEWORK FOR THE PROCESSING OF UNCERTAIN KNOWLEDGE

S. Y. Lu and H. E. Stephanou

Long Range Research Division
Exxon Production Research Co.
P. O. Box 2189
Houston, Texas 77001

ABSTRACT

In this paper, a knowledge base is represented by an input space, an output space, and a set of mappings that associate subsets of the two spaces. Under this representation, knowledge propocessing has three major parts: (1) The user enters observations of evidence in the input space and assigns a degree of certainty to each observation. (2) A piece of evidence that receives a non-zero certainty activates a mapping. This certainty is multiplied by the certainty associated with the mapping, and is thus propagated to a proposition in the output space. (3) The consensus among all the propositions that have non-zero certainties is computed, and a final set of conclusions is drawn. A degree of support is associated with each conclusion.

The underlying model of certainty in this processing scheme is based on the Dempster-Shafer mathematical theory of evidence. The computation of the consensus among the propositions uses Dempster's rule of combination. The inverse of the rule of combination, which we call the rule of decomposition, is derived in this paper. Given an expected consensus, the inverse rule can generate the certainty required for each proposition. Thus, the certainties in the mappings can be inferred iteratively through alternating use of the rule of combination and the rule of decomposition.

1. INTRODUCTION

In this paper, we propose a new representation of knowledge based on set theory. A knowledge base consists of three parts: an input space from which evidence is drawn, an output space that consists of propositions to be proved, and a set of mappings that associate subsets of the input space with subsets of the output space. In this representation, two types of certainties are defined. The certainty assigned to a piece of evidence expresses the degree of confidence that a user has in his observation of the evidence. The certainty assigned to a mapping expresses the degree of confidence that an expert has in his definition of the mapping. These two sources of certainty are compounded in proving a proposition.

The theoretical foundation for handling partial certainty under this representation is based on the Dempster-Shafer "theory of evidence" [1]. Shafer defines certainty to be a function that maps subsets in a space on a scale from zero to one, where the total certainty over the space is one. The definition also allows one to assign a non-zero certainty to the entire space as an indication of ignorance. This provision for expressing ignorance is one way in which Shafer's theory differs from conventional probability theory, and is a significant advantage, since in most applications the available knowledge is incomplete and involves a large degree of uncertainty.

A mapping is activated when the input part of the mapping, the user's observation of evidence, receives a non-zero certainty. The product of this certainty with the certainty in the mapping is the certainty in the proposition. Dempster's rule of combination provides a mechanism to combine the certainty of several propositions, which can be concordant or contradictory. When this mechanism is used, reasoning becomes a process of seeking consensus among all the propositions that are supported by the user's observations. This approach is attractive, since such problems as conflicting observations from multiple experts, knowledge updating, and ruling-out are resolved automatically by the rule of combination. The conventional approaches to knowledge processing, which use tightly coupled chains or nets such as deductive rules or semantic nets, do not have this advantage [2,3].

The use of the Dempster-Shafer theory of evidence to handle uncertainty in knowledge processing was first discussed at the Seventh International Conference on Artificial Intelligence, 1981. Two papers related to the subject were presented [4,5]. Barnett discussed the computational aspects of applying the theory to knowledge processing [4]. Garvey et al. applied the method to model the response from a collection of disparate sensors [5]. The signals from an emitter are parameterized, and the likelihood of a range of parameter values is expressed by Shafer's definition of certainty. The integration of parameters is computed by using Dempster's rule of combination.

In this paper, we use the theory of evidence as an underlying model for partial certainty in a general knowledge-processing scheme.

2. A SET-THEORETIC REPRESENTATION

A knowledge base can be represented by two spaces and a set of mappings between the two spaces. Let the two spaces be called an input space, labeled I, and an output space, labeled O. A proposition in I is represented by a subset of elements in I. Its relation to a subset of O, which represents a proposition in O, defines a mapping. Let us denote the collection of mappings defined in this way by R. Then $R : I \rightarrow O$. The input space consists of evidence that can be observed by users. The ouput space consists of conclusions that can be deduced from the observations.

In this representation, we consider two types of certainty. One is associated with the user's observation of evidence in the input space. The second type is the certainty that an expert assigns to the mappings. By combining these two certainties, a knowledge processing scheme deduces the most likely hypotheses in the output space. Among the problems that can be represented in this way are those of classification, translation, and diagnosis.

CLASSIFICATION. A typical pattern recognition problem can be represented by an input space that is a space of independent features, an output space that is a group of disjoint classes, and a set of mappings that are described by a classifier.

TRANSLATION. Language translation is a typical example of a translation process. In a translation scheme, the two spaces are the source language and the target language. Within each space, elements are genetically related and well structured. These relations are characterized by syntax, and the mappings can be represented by a transformational grammar.

DIAGNOSIS. Medical diagnosis can involve more than two spaces. First, there is a symptom space, which is composed of features of visible symptoms or laboratory measurements. The second space may consist of possible diseases, and the third space of treatments to be administered. The ruling-out capability is important in this case, since some treatments can be fatal to a patient with certain symptoms or diseases.

3. THE DEMPSTER-SHAFER THEORY OF EVIDENCE

Shafer defines certainty to be a function that maps subsets in a space on a scale of 0 to 1, where the total certainty over the space is 1. If a certainty function assigns 0.4 to a subset, it means that there is 0.4 certainty that the truth is somewhere in this subset. The definition also allows one to assign a non-zero certainty to the entire space. This is called the degree of "ignorance." It means that any given subset is no closer to containing the truth than any other subset in the space.

Some definitions that are used throughout the paper are given in this section.

Definition 1

Let Θ be a space; then a function $m:2^{\Theta} \rightarrow [0,1]$ is called a certainty function whenever

(1) $m(\varphi) = 0$, where φ is an empty set,

(2) $0 < m(A) < 1$, and

(3)
$$\sum_{A \subset \Theta} m(A) = 1.$$

The space "Θ", and the certainty function "m", are called the "frame of discernment", and the "basic probability assignment", respectively, in [1].

A subset A of Θ is called a focal element if $m(A) > 0$. The simplest certainty function is one that has only one focal element.

Definition 2

A certainty function is called a simple certainty function when

(1) $m(A) > 0$,

(2) $m(\Theta) = 1 - m(A)$, and

(3) $m(B) = 0$, for all other $B \subset \Theta$.

The focus of the simple certainty function is A.

Here, a simple certianty function is called a "simple support function" in [1].

The quantity $m(A)$ measures the certainty that one commits specifically to A as a whole, i.e. to no smaller subset of A. However, this quantity is not the total belief that one commits to A. Shafer defines the total belief committed to A to be the sum of certainties that are committed to A and all the subsets of A.

Definition 3

A function Bel:$2^{\Theta} \rightarrow [0,1]$ is called a belief function over Θ if it is given by

$$Bel(A) = \sum_{B \subset A} m(B). \tag{1}$$

Dempster defines an operation on certainty functions that is called "orthogonal sum," and is denoted by $\oplus$.

Definition 4

Let m_1 and m_2 be two certainty functions over the same space Θ, with focal elements $A_1, \ldots, A_k$ and $B_1, \ldots, B_l$, respectively. Suppose that

$$\sum_{A_i \cap B_j = \varphi} m_1(A_i) m_2(B_j) < 1.$$

Then the function $m: 2^{\Theta} \rightarrow [0,1]$ is defined by $m(\varphi) = 0$, and

$$m(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \varphi} m_1(A_i) m_2(B_j)} \tag{2}$$

for all non-empty subsets $A \subset \Theta$, is a certainty function, and $m = m_1 \oplus m_2$. Equation (2) is called Dempster's rule of combination.

4. KNOWLEDGE PROCESSING UNDER PARTIAL CERTAINTY

Given this definition of certainty, we can quantify our belief in a mapping. We assume that a mapping defines a simple certainty function over the output space. This certainty indicates the degree of association between elements in I and elements in O. Therefore, a mapping in R is expressed as

$$e \dashrightarrow h, \nu, \tag{3}$$

where $e \subset I$, $h \subset O$, and $0 \leq \nu \leq 1$. This mapping defines a simple certainty function, where the focus in O is h, and ν is the degree of association, in the expert's opinion, between e and h. That is, $\nu = 1$ means complete confidence, and $1 - \nu$ is the degree to which the expert chooses to be noncommittal, or the degree of ignorance. Furthermore, a mapping is assumed to be an indenpendent piece of knowledge.

The user of a knowledge processor gives his observation of evidence in the input space, and also the certainty associated with that observation. Each observation defines a certainty function over the input space. Let the certainty function be denoted by q, $q: 2^I \dashrightarrow [0,1]$. The user is allowed to make multiple observations. Assuming that each observation is independent, we can derive a combined observation by using the orthogonal sum of these observations. That is, $q = q_1 \oplus q_2 \oplus \cdots \oplus q_n$, where $q_1, \ldots, q_n$ are n independent observations. Then the belief function defined by the combination of the observations is denoted by Bel_q.

We say that a mapping is activated when the evidence for that mapping is assigned a non-zero belief in a combined observation. That is, the mapping $e \dashrightarrow h$ is activated if $Bel_q(e) > 0$. When a mapping is activated, the certainty in the evidence is propagated by the mapping to a decision in the output space. As a result, an activated mapping defines a certainty function $\zeta : 2^\Theta \dashrightarrow [0,1]$ over the output space, where $\zeta(h) = \nu \times \mu$, and $\zeta(\Theta) = 1 - \nu \times \mu$.

In the case where there is more than one mapping activated in a run, several certainty functions will be defined over the output space. The final certainty function is the combined certainty obtained by taking the orthogonal sum of all the activated mappings. Finally, the total belief for the output space is computed by using Eq. (1) in Definition 2.

In summary, this processing procedure has five steps:

(1) Query for Observations. An observation of evidence, and the certainty associated with the observation are entered by the user. They define a certainty function over the input space.

(2) Normalize the Certainty in the Input Space. The user is allowed to make multiple independent observations. The certainty functions defined by these observations are combined by using the rule of combination.

(3) Activate the Mappings. A mapping is activated when the evidence for the mapping receives a non-zero belief in the combined observation.

(4) Propagate the Certainty to the Output Space. The certainty of the evidence in an activated mapping is multiplied by the certainty in the mapping. The result is a certainty function defined over the output space for each activated mapping.

(5) Normalize the Certainty in the Output Space. By means of the rule of combination, all the activated certainty functions in the output space are combined into a single certainty function. From this certainty function, the total belief for the output space is computed.

This process is illustrated by the following example.

Example 1

A knowledge base is schematized in Figure 1: an input space that contains subsets A, B, and C, and an output space that contains subsets X, Y, and Z. Three mappings are given by

$A \dashrightarrow X$, 0.8

$B \dashrightarrow Y$, 0.7

$C \dashrightarrow Z$, 1.

Suppose that the user makes two independent observations defined by the certainty functions

$q_1(A) = 0.8$,

$q_1(I) = 0.2$,

and

$q_2(B) = 0.4$,

$q_2(I) = 0.6$.

Then the combined observation is given by the certainty function

$q(A) = 0.48$,

$q(B) = 0.08$,

$q(A \cap B) = 0.32$,

$q(I) = 0.12$.

From the combined observation, we find that the belief in A, B, and C of the input space, I, is given by the belief function

$Bel_q(A) = q(A) + q(A \cap B) = 0.8$,

$Bel_q(B) = q(B) + q(A \cap B) = 0.4$, and

$Bel_q(C) = q(A \cap B) = 0.32$.

Therefore, for the given observations all three mappings are activated. The three activated certainty functions in the output space, O, are ζ_1, ζ_2, and ζ_3:

$\zeta_1(X) = 0.8 \times 0.8 = 0.64$, and $\zeta_1(O) = 0.36$;

$\zeta_2(Y) = 0.4 \times 0.7 = 0.28$, and $\zeta_2(O) = 0.72$;

$\zeta_3(Z) = 0.32 \times 1 = 0.32$, and $\zeta_3(O) = 0.68$.

Finally, by using the rule of combination we can compute the certainty function ζ, $\zeta = \zeta_1 \oplus \zeta_2 \oplus \zeta_3$:

$\zeta(X) = 0.3133$,

$\zeta(Y) = 0.0675$,

$\zeta(Z) = 0.0829$,

$\zeta(X \cap Y) = 0.1218$,

$\zeta(X \cap Z) = 0.1474$,

$\zeta(Y \cap Z) = 0.0322$,

$\zeta(X \cap Y \cap Z) = 0.0573$, and

$\zeta(O) = 0.1662$.

The final belief function over the output space is

$Bel_\zeta(X) = 0.64$,

$Bel_\zeta(Y) = 0.28$,

$Bel_\zeta(Z) = 0.32$,

$Bel_\zeta(X \cap Y) = 0.1792$,

$Bel_\zeta(X \cap Z) = 0.2048$,

$Bel_\zeta(Y \cap Z) = 0.0896$,

$Bel_\zeta(X \cap Y \cap Z) = 0.0573$.

As has been said earlier, the set-theoretic representation and the rule of combination are attractive for knowledge processing because such problems as multiple experts, knowledge updating, and ruling-out can be automatically resolved in the processing scheme.

A mapping represents an opinion. When the rule of combination is applied to combine several mappings, the result can be interpreted as the consensus among several opinions. When some opinions are contradictory, they erode each other. On the other hand, concurring opinions reinforce each other. The problem of multiple experts can be handled by treating them as a knowledge base with several sets of mappings, each contributed by a different expert. If all experts' opinions are weighted equally, then it makes no difference whom the mappings come from. The problem of knowledge updating can be handled by simply adding a new set of mappings to an existing knowledge base. The following examples illustrate the handling of conflicting opinions and "ruling-out" in this scheme.

Example 2

The input and output spaces are the same as in Example 1, but the knowledge base now contains the following mappings:

$A \dashrightarrow X,\ 0.8,$

$B \dashrightarrow \bar{X},\ 0.7,$

$C \dashrightarrow Z,\ 1.$

That is, B supports the opposite of what A supports. Assume that the user's observations are the same as in Example 1. Then the final belief function in the output space is

$Bel_\zeta(X) = 0.5614,$

$Bel_\zeta(\bar{X}) = 0.1228,$

$Bel_\zeta(Z) = 0.32,$

$Bel_\zeta(X \cap Z) = 0.1796,$

$Bel_\zeta(\bar{X} \cap Z) = 0.0393.$

In comparison with the results in Example 1, the belief in X is eroded to some extent, but the belief in Z remains the same.

Ruling-out means that if evidence x is observed, then proposition y is false with total certainty (i.e. it is ruled out). This is represented as

$$x \dashrightarrow \bar{y},\ 1.$$

Example 3

The second mapping in Example 2 is changed to be a "ruling-out" mapping for proposition X if B is observed; that is,

$$B \dashrightarrow \bar{X},\ 1.$$

If the same observations as in the previous examples are used, the final belief function is

$Bel_\zeta(X) = 0.4161,$

$Bel_\zeta(\bar{X}) = 0.1935,$

$Bel_\zeta(Z) = 0.32,$

$Bel_\zeta(X \cap Z) = 0.1651,$

$Bel_\zeta(\bar{X} \cap Z) = 0.0619.$

Because the belief in B is not fully supported by the observations, proposition X is not completely suppressed by the ruling-out mapping.

Example 4

Assume that the user makes the following observations:

$q_1(A) = 0.8,$

$q_1(I) = 0.2,$

and

$q_2(B) = 1.$

If the knowledge base in Example 3 is used, the final belief function is

$$Bel_\zeta(\bar{X}) = 1,$$

and the belief in all the other propositions is 0.

5. THE DECOMPOSITION OF CERTAINTY

One difficulty in this knowledge processing scheme is the assignment of certainty to a mapping. Even the domain expert can provide only a crude approximation, since the degree of belief is a relative matter. It is difficult for one to be consistent in assigning certainty to a mapping on a scale of 0 to 1 when a large number of such mappings are involved. Motivated by this difficulty, we have derived an inverse to the rule of combination. We call it "the rule of decomposition." With initial certainty assignments to the mappings, the expert can use the knowledge base by entering evidence and observing the final deduced belief function over the output space. If the final belief function is inconsistent with the expert's expectation, he can use the rule of decomposition to modify the certainty assignment for individual mappings. If the expert is consistent, the knowledge base will approach consistency after a number of iterations.

The rule of decomposition decomposes a certainty function into a number of simple certainty functions. However, not all certainty functions can be decomposed into simple certainty functions. Shafer defines the class of certainty functions that can be decomposed as "separable" certainty functions. Shafer also proves that the decomposition of a separable certainty function is unique.

Before deriving the general rule of decomposition, we consider four special cases of combining n simple certainty functions. The rule of decomposition is then derived for each of the four cases. Finally, we give a procedure for decomposing any separable certainty function.

Lemma 1

n Simple Certainty Functions with Identical Focus

Let $m_1, m_2, \ldots, m_n$ be n simple certainty functions, where $A \subset \Theta$ is the only focus, and

$m_i(A) = \alpha_i$, for $1 \le i \le n$,

and $0 \le \alpha_i \le 1$. Then the combined certainty function, $m = m_1 \oplus m_2 \oplus \cdots \oplus m_n$, is

$$m(A) = 1 - \prod_{i=1}^{n}(1 - \alpha_i). \tag{4}$$

Lemma 2

n Simple Certainty Functions with n Disjoint Focuses

Let $m_1, m_2, \ldots, m_n$ be n simple certainty functions, and $A_1, A_2, \ldots, A_n$ be their n focuses, respectively. $A_i \cap A_j = \varphi$, for all $i, j, i \neq j$. Assume that $m_i(A_i) = \alpha_i$, and $0 \leq \alpha_i \leq 1$, for all i. Then the combined certainty function is

$$m(A_i) = \frac{\alpha_i \prod_{\substack{j=1 \\ j \neq i}}^{n} (1 - \alpha_j)}{\sum_{i=1}^{n} \alpha_i \prod_{\substack{j=1 \\ j \neq i}}^{n} (1 - \alpha_j) + \prod_{j=1}^{n} (1 - \alpha_j)} \tag{5}$$

for $i = 1, \ldots, n$.

Lemma 3

n Simple Certainty Functions with n Focuses Where the Intersection of Any Number of Focuses Is Non-Empty

Let $m_1, m_2, \ldots, m_n$ be n simple certainty functions, and $A_1, A_2, \ldots, A_n$ be their n focuses, respectively. Also, let κ be a subset of the index set, $\{1, 2, \ldots, n\}$, and $\bigcap_{i \varepsilon \kappa} A_i$ be the intersection of the subsets for which the indexes are in κ. Assume that $\bigcap_{i \varepsilon \kappa} A_i \neq \varphi$, for all possible κ, and $\bigcap_{i \varepsilon \kappa} A_i \neq \bigcap_{i \varepsilon \iota} A_j$ for all κ and ι, where both κ and ι are subsets of the index set $\{1, 2, \ldots, n\}$, and $\kappa \neq \iota$. Then the combined certainty function is

$$m(\bigcap_{i \varepsilon \kappa} A_i) = \prod_{i \varepsilon \kappa} \alpha_i \prod_{j \varepsilon \bar{\kappa}} (1 - \alpha_j), \tag{6}$$

where $\bar{\kappa} = \{1, 2, \ldots, n\} - \kappa$.

Lemma 4

n Simple Certainty Functions with Nested Focuses

Let $A_1, A_2, \ldots, A_n$ be the n focuses of n simple certainty functions, $m_1, m_2, \ldots, m_n$, respectively. Assume that $A_1 \subset A_2 \subset \cdots \subset A_n$. Then the combined certainty function is

$$m(A_i) = \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j). \tag{7}$$

Equation (4) to (7) can be proved by induction.

The inverse of the three special cases described in Lemma 2 to 4 are given by equations (8) to (10), respectively.

$$\alpha_i = \frac{m(A_i)}{1 - \sum_{\substack{j=1 \\ j \neq i}}^{n} m(A_j)} \tag{8}$$

$$\alpha_i = \sum_{i \varepsilon \kappa}^{\kappa} m(A_\kappa) \tag{9}$$

and

$$\alpha_i = \frac{m(A_i)}{1 - \sum_{j=1}^{i-1} m(A_j)} \tag{10}$$

For the case where the focuses of n simple certainty functions are identical, the combined certainty function is also a simple certainty function, as shown in Eq. (4). For this case, although a simple certainty function can be decomposed into several simple certainty functions on the same focus, the decomposition is not unique.

In Eqs. (8) to (10) we have derived the decomposition of three special types of separable certainty functions into simple certainty functions. We now show a procedure for decomposing any separable certainty function into two certainty functions: one simple certainty function and one separable certainty function. By repeatedly applying this procedure, one can decompose a given separable certainty function into a number of simple certainty functions.

Lemma 5

The Decomposition of a Separable Certainty Function

Let m be a separable certainty function with focal elements $A_1, A_2, \ldots, A_n$. Then m can be decomposed into m_1 and m_2. That is, $m = m_1 \oplus m_2$, where m_1 is a simple certainty function focused on A_i, and m_2 is a separable certainty function.

Choose A_i such that $A_i \subset A_k \cap A_l$ is not true for all $1 \leq k, l \leq n$. Let A_i be the focus of m_1. Let κ be a subset of the index set $\{1, 2, \ldots, i-1, i+1, \ldots, n\}$ such that j is in κ if and only if $A_j \neq A_l \cap A_i$ for some $l, l = 1, \ldots, n$. Assume that A_j's, for all $j \varepsilon \kappa$, are focal elements of m_2. Then using the rule of combination we have

$$m(A_i) = \frac{m_1(A_i) \times m_2(\Theta)}{1 - \sum_{A_i \cap A_j = \varphi} m_1(A_i) \times m_2(A_j)} \tag{11}$$

and

$$m(A_j) = \frac{m_1(\Theta) \times m_2(A_j)}{1 - \sum_{A_i \cap A_j = \varphi} m_1(A_i) \times m_2(A_j)} \tag{12}$$

for all $j \varepsilon \kappa$.

From the definitions of a certainty function, we have

$$m_1(A_i) + m_1(\Theta) = 1, \tag{13}$$

and

$$\sum_{j \varepsilon \kappa} m_2(A_j) + m_2(\Theta) = 1. \tag{14}$$

From Eqs. (11) through (14), we can derive

$$\frac{m_1(\Theta)}{m_2(\Theta)} = \frac{m(A_i) + m(\Theta)}{\sum_{j \varepsilon \kappa} m(A_j) + m(\Theta)} \tag{15}$$

Now, from Eqs. (11), (12), and (15) we have

$$\frac{m_1(A_i)}{m_2(A_j)} = \frac{m_2(\Theta)}{m_1(\Theta)} \times \frac{m(A_i)}{m(A_j)} \tag{16}$$

for $j \varepsilon \kappa$.

Therefore, substituting Eq. (16) for m_2 in Eq. (14), we have

$$m_1(A_i) = \frac{m_1(\Theta)}{m_2(\Theta)} \times \frac{m(A_i)}{\sum_{j\epsilon\kappa} m(A_j) + m(\Theta)} \tag{17}$$

and

$$m_1(\Theta) = \frac{m_1(\Theta)}{m_2(\Theta)} \times \frac{m(\Theta)}{\sum_{j\epsilon\kappa} m(A_j) + m(\Theta)}$$

Similarly, substituting Eq. (16) for m_1 in Eq. (13), we have

$$m_2(A_j) = \frac{m_2(\Theta)}{m_1(\Theta)} \times \frac{m(A_j)}{m(A_i) + m(\Theta)} \tag{18}$$

and

$$m_2(\Theta) = \frac{m_2(\Theta)}{m_1(\Theta)} \times \frac{m(\Theta)}{m(A_i) + m(\Theta)}$$

In the case where a certainty function is not separable, the two certainty functions can still be derived with Eqs. (17) and (18). However, their orthogonal sum will not be equal to the original certainty function.

6. DISCUSSION

In this report, we propose a new knowledge representation and processing scheme. A knowledge base is represented by an input space, an output space, and a set of mappings. The input space and the output space define the domain of the knowledge. The mappings link subsets in the input space to subsets in the output space. In addition to being able to handle partial certainty, the new scheme also has the following advantages:

(1) The representation can handle incomplete or conflicting observations. Conflicting knowledge sources erode the certainty in the processing scheme and yield less meaningful results, but do not disrupt the reasoning process. Thus, in this representation the usual difficulties associated with multiple experts and conflicting opinions do not exist.

(2) It is easy to implement knowledge acquisition and updating. The conventional rule-based approach organizes a knowledge base as tightly coupled and consistent chains of events, so that the reasoning mechanism can be implemented easily [1]. However, adding new knowledge or modifying the existing knowledge base requires the restructuring of the chaining. The complexity of updating a knowledge base increases as the knowledge base grows larger. In the set-theoretic knowledge representation, updating can be done by expanding the input and output spaces and adding or removing mappings between the two spaces. In this case, the complexity is not related to the size of the knowledge base.

(3) The representation can be extended to multiple stages. Throughout the report, the partial certainties are presented in two stages. First the certainty in the observations is given, and then the certainty in the mappings. The idea of combining the certainty in two stages can easily be extended to multiple stages. The spaces in between the input and the output spaces correspond to intermediate hypotheses, or decisions. The normalization that takes place at each stage eliminates the problem of rapidly diminishing in probabilities during propagation in a Bayesian model.

The Proposition that a mapping in a knowledge base defines a simple certainty function is to make the processing scheme tractable. However, the normalization that is based on Dempster's rule of combination assumes the independence among mappings. To satisfy both requirements, the applicability of the knowledge processing scheme is limited to a small class of knowledge. Our future work is to expand the representation to larger classes of belief functions, namely, separable support functions and support functions in Shafer's definition. In the expanded scheme, dependent pieces of knowledge will be represented by one belief function.

ACKNOWLEDGMENT

The authors are indebted to Professor K. S. Fu of Purdue University for first bringing the Dempster-Shafer theory of evidence to their attention.

REFERENCES

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, New Jersey, 1976.

[2] R. D. Duda, P. E. Hart, and N. J. Nilsson, "Subjective Bayesian Methods for Rule-Based Inference Systems," *Proceedings of National Computer Conference*, AFIPS, Vol. 45, pp. 1075-1082, 1976.

[3] E. Shortliffe, *Computer Based Medical Consultations: MYCIN*, American Elsevier, New York, 1976.

[4] J. A. Barnett, "Computational Methods for a Mathematical Theory of Evidence," *Proceedings of the 7th International Conference on Artificial Intelligence*, Aug. 1981, Vancouver, pp. 868-875.

[5] T. D. Garvey, J. D. Lowrance, and M. A. Fischler, "An Inference Technique for Integrating Knowledge from Disparate Sources," *Proceedings of the 7th International Conference on Artificial Intelligence*, Aug. 1981, Vancouver, pp. 319-325.

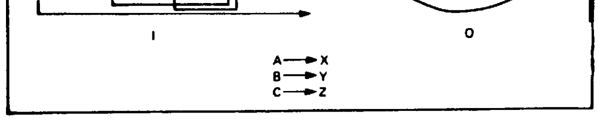

FIGURE I THE INPUT SPACE AND THE OUTPUT SPACE OF EXAMPLE 1 TO 3