ABSTRACT

The main purpose of the present work is to develop a theory for multiple knowledge systems. A knowledge system could be a sensor or an expert system, but it must specialize in one feature. The problem is that we have an exhaustive list of possible answers to some query (such as "What object is it?"). By collecting different feature values, we should, in principle, be able to give an answer to the query, or at least narrow down the list.

Since a sensor, or for that matter an expert system, do not in most cases yield a precise value for the feature, uncertainty must be built into the model. Also, we must have a formal mechanism to be able to put the information together. We chose to use the Dempster-Shafer approach to handle the problems mentioned above.

We introduce the concept of a state of recognition and point out that there is a relation between receiving updates and defining a set valued Markov Chain. Also, deciding what the value of the next set valued variable is can be phrased in terms of classical decision making theory such as minimizing the maximum regret. Other related problems are looked at.
INTRODUCTION

The purpose of the present work is to show how taking independent and very diverse evidence, we can piece things together to arrive at an answer to the question: "What object is it?". We will take the Dempster - Shafer approach to put the evidence together. Such an approach has recently been taken in expert systems, see [2], [3], and [4]. However, to the best of this writer's knowledge, the results shown here are original. We start out with a simple example.

Consider the following data which assigns masses to subsets of \{Bird, Plane, & Superman\} according to the velocity observed:

<table>
<thead>
<tr>
<th>VELOCITY</th>
<th>B</th>
<th>P</th>
<th>S</th>
<th>{BP}</th>
<th>{BS}</th>
<th>{PS}</th>
<th>{BPS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 100</td>
<td>.5</td>
<td>.1</td>
<td>.1</td>
<td>2</td>
<td>.04</td>
<td>.04</td>
<td>.02</td>
</tr>
<tr>
<td>101 - 200</td>
<td>0</td>
<td>.4</td>
<td>.1</td>
<td>0</td>
<td>0</td>
<td>.5</td>
<td>0</td>
</tr>
<tr>
<td>201 - 500</td>
<td>0</td>
<td>.5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>.4</td>
<td>0</td>
</tr>
<tr>
<td>&gt; 500</td>
<td>0</td>
<td>.1</td>
<td>.7</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

NOTE:
- Birds don't fly with velocity > 100.
- Superman likes to fly at over 500 but he can fly at any speed he wants to.

It should be noted that the sum across each row is 1. The interpretation of the results says, for example, that at velocities exceeding 500 mph, the expert believes that the object is Superman. That expert doesn't totally rule out the possibility of plane as he assigns a mass of .1 to that event, and also that expert is somewhat unsure if the object is plane or Superman and therefore, he assigns a mass of .2 to that aggregate. Note that we do not have Probability of \{PS\} be the sum of the Probability of P and of S. Masses assigned to sets that are not singletons denote the uncertainty of the expert. For example, .02 assigned to \{BPS\} reflects the degree of total ignorance that the expert has regarding what the object is when the object travels at less than 100 mph. Such a mass assigned is typical of the Dempster - Shafer approach to handle uncertainty in expert systems. See [10].

We now write down the data relative to observed color:

<table>
<thead>
<tr>
<th>COLOR</th>
<th>B</th>
<th>P</th>
<th>S</th>
<th>{BP}</th>
<th>{BS}</th>
<th>{PS}</th>
<th>{BPS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>SILVER</td>
<td>.05</td>
<td>.6</td>
<td>.05</td>
<td>.1</td>
<td>0</td>
<td>.15</td>
<td>.05</td>
</tr>
<tr>
<td>WHITE</td>
<td>.1</td>
<td>.1</td>
<td>.05</td>
<td>.5</td>
<td>.05</td>
<td>.15</td>
<td>.05</td>
</tr>
<tr>
<td>RED</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.1</td>
</tr>
<tr>
<td>BLUE</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.2</td>
<td>.2</td>
<td>.2</td>
<td>.1</td>
</tr>
<tr>
<td>RED-BLUE</td>
<td>.04</td>
<td>.04</td>
<td>.8</td>
<td>0</td>
<td>.05</td>
<td>.05</td>
<td>.02</td>
</tr>
<tr>
<td>OTHER</td>
<td>.6</td>
<td>.1</td>
<td>0</td>
<td>.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

NOTE:
- Red and blue generate the same (conditional) mass
- A gray bird may appear silver
- Superman wears red and blue but from some angles he may appear all red or all blue
- When flying at certain speeds, Superman may appear as a white or silver streak
- Color other than red, blue, white, silver rules out Superman

These two tables sum up the information collected from the experts. What we would like to do, of course, is to put these two pieces of evidence together.
CONCEPTS AND NOTATIONS

We now will formally define the concept of a mass function. A mass function is a function from subsets of the frame of discernment \( \Theta \) into \([0, 1]\) satisfying the following conditions:

(i) \( m(\emptyset) = 0 \)

(ii) \( \sum m(A) = 1 \) where the sum is over all subsets of \( \Theta \).

If \( m_1 \) & \( m_2 \) are two mass functions, we define

\[
(m_1 \oplus m_2)(C) = \sum_{A \land B = C} m_1(A)m_2(B) (1 - k)
\]

Where \( k \) is the conflict

\[
k = \sum_{A \land B = \emptyset} m_1(A)m_2(B)
\]

The operation defined above defines how to put information together. If two knowledge systems generate \( m_1 \) & \( m_2 \), \( m_1 \oplus m_2 \) is the mass generated by combining the two knowledge systems, see [10]. For a very readable interpretation of the combination rule in the setting of databases, see [12].

The belief generated by \( m \) is defined by

\[
Bel(A) = \sum m(B) \text{ over all sets } B \text{ such that } B \subset A.
\]

Also we define the plausibility by

\[
Pls(A) = \sum m(B) \text{ over all sets } B \text{ such that } B \land A \neq \emptyset.
\]

Now if the \( l \) th sighting takes place at time \( t_l \), set \( dt_l = t_l - t_{l-1} \)

Obviously, \( dt_l \) denotes the elapsed time between sightings. Assume that we have a weight function \( \lambda(*) \) satisfying

\[
(i) 0 \leq \lambda(dt_l) \leq 1, \ dt_l > 0
\]

\[
(ii) \lambda(dt_l) \text{ non decreasing as a function of } dt_l
\]

We use weight to adjust masses. There are two ways to adjust

\[
m_l(*) = m_1(*)
\]

a) \( m_l(*) = \lambda(dt_l) m_l(*) + (1 - \lambda(dt_l)) m_{l-1}(*) \)

b) \( m_l(*) = \lambda(dt_l) m_l(*) + (1 - \lambda(dt_l)) m_{l-1}(*) \)

If \( \lambda \) is high i.e., \( dt_l \) high, go with the current observation

If \( \lambda \) is low i.e., \( dt_l \) low, go with the accumulated data.

Note that the first update is Markov - like as it only uses the mass collected on the previous sighting.
For our example, we could define the weight function by:

$$\lambda (dt_i) = \begin{cases} \frac{dt_i}{300} & \text{if } dt_i \leq 300 \\ 1 & \text{otherwise} \end{cases}$$

That is, after 5 minutes, forget the previous observations and assume that a new object is being observed. The rationale for this is that the data has become too old to be reliable.

Going back to our example of bird, plane, and Superman, assume we have three sightings:

<table>
<thead>
<tr>
<th>SIGHTING</th>
<th>TIME</th>
<th>( dt_i )</th>
<th>VELOCITY</th>
<th>COLOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1:00 p.m.</td>
<td>0</td>
<td>101-200</td>
<td>WHITE</td>
</tr>
<tr>
<td>2</td>
<td>1:01 p.m.</td>
<td>60</td>
<td>201-500</td>
<td>WHITE</td>
</tr>
<tr>
<td>3</td>
<td>1:29 p.m.</td>
<td>1740</td>
<td>0-100</td>
<td>OTHER</td>
</tr>
</tbody>
</table>

The combined masses, not time adjusted are given below:

<table>
<thead>
<tr>
<th>SIGHTING</th>
<th>B</th>
<th>P</th>
<th>S</th>
<th>{BP}</th>
<th>{BS}</th>
<th>{PS}</th>
<th>{BPS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( m_1(*) )</td>
<td>0</td>
<td>.775</td>
<td>.1</td>
<td>0</td>
<td>0</td>
<td>.125</td>
<td>0</td>
</tr>
<tr>
<td>2 ( m_2(*) )</td>
<td>0</td>
<td>.810</td>
<td>.086</td>
<td>0</td>
<td>0</td>
<td>.1013</td>
<td>0</td>
</tr>
<tr>
<td>3 ( m_3(*) )</td>
<td>.811</td>
<td>.102</td>
<td>0</td>
<td>.0866</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The combined masses, time adjusted are given below:

<table>
<thead>
<tr>
<th>SIGHTING</th>
<th>B</th>
<th>P</th>
<th>S</th>
<th>{BP}</th>
<th>{BS}</th>
<th>{PS}</th>
<th>{BPS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( m_1(*) )</td>
<td>0</td>
<td>.775</td>
<td>.1</td>
<td>0</td>
<td>0</td>
<td>.125</td>
<td>0</td>
</tr>
<tr>
<td>2 ( m_2(*) )</td>
<td>0</td>
<td>.782</td>
<td>.097</td>
<td>0</td>
<td>0</td>
<td>.12026</td>
<td>0</td>
</tr>
<tr>
<td>3 ( m_3(*) )</td>
<td>.811</td>
<td>.102</td>
<td>0</td>
<td>.0866</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

NOTE: \( m_3 \) is not really time adjusted as sightings are more than 5 minutes apart.
Computing the belief, at each sighting, with respect to the time adjusted mass we have:

<table>
<thead>
<tr>
<th>SIGHTING</th>
<th>OBJECT</th>
<th>Bel (•)</th>
<th>Bel (¬•)</th>
<th>Bel (•) - Bel (¬•)</th>
<th>CLASSIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>.775</td>
<td>.1</td>
<td>.675</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>.1</td>
<td>.775</td>
<td>-.675</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>.78202</td>
<td>.09772</td>
<td>.6843</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>.09772</td>
<td>.78202</td>
<td>-.6843</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>.811</td>
<td>.1024</td>
<td>.7086</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>.1024</td>
<td>.811</td>
<td>-.7086</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td></td>
</tr>
</tbody>
</table>

The rationale for the table above is that Bel (•) - Bel (¬•) measures how much a specific object exceeds, belief-wise, its competition. This criterion was already used in [8]. Thus the conclusion is that a plane was observed on the first and second sighting and a bird was observed on the last sighting. This example shows that there will be a payoff in studying a multi-knowledge systems setting. We also remark that a similar, but somewhat more complex approach could be used to obtain classification sequences.

MESHING THE INFORMATION COLLECTED FROM MULTIPLE KNOWLEDGE SYSTEMS

In this section, we consider the composition rule to be defined by the numerator only of \((m_1 \Theta m_2) (C)\). (Thus the empty set may pick up mass).

We now shift somewhat our perspective. Consider Knowledge Systems \(KS_1, KS_2, \ldots, KS_n\). \(KS_j\) reads the \(J^{th}\) feature and interprets its value to be \(f^j_i\) with a probability \(a_{j,i}\).

It is important to keep in mind that in this setting, each knowledge system specializes in recognizing a specific feature.

Thus, \(KS_i\) defines the mass \(m_j\) on \(\Theta\) by

\[ m_j (A_{j,i}) = a_{j,i} \]

where \(A_{j,i}\) denotes all objects of \(\Theta\) whose \(j^{th}\) feature has value \(f^j_i\).

After we have interrogated \(KS_1, KS_2, \ldots, KS_q\) possible answers are in sets \(A_{1t_1} \land A_{2t_2} \land \ldots \land A_{qt_q}\). Let \(X_q\) be set-valued variables whose values are \(A_{1t_1} \land \ldots \land A_{qt_q}\) \((t_1, \ldots, t_q\) range over all possible values of the corresponding features). \(X_q\) indicates the current state of recognition. \(X_q\) may be viewed as a random set [9].

We have shown that \(X_q\) forms a (non-stationary) Markov chain. In fact, the transition probabilities are given by

\[ Pr(X_{q+1} = A_{1t_1} \land \ldots \land A_{qt_q} \land A_{q+1t_{q+1}} \mid X_q = A_{1t_1} \land \ldots \land A_{qt_q}) = \sum m_{q+1} (B_{q+1,i}) \]
Where the sum is over sets $B_{q+1,i}$ such that

$$A_{1t_1} \land \ldots \land A_{qt_q} \land B_{q+1,i} = A_{1t_1} \land \ldots \land A_{qt_q} \land A_{q+1t_{q+1}}$$

If

$$Pr\left(X_q = A_{1t_1} \land \ldots \land A_{qt_q}\right) = \mu_q\left(A_{1t_1} \land \ldots \land A_{qt_q}\right)$$

then

$$\mu_{q+1}(A) = \sum_{B} Pr\left(X_{q+1} = A \mid X_q = B\right) \mu_q(B)$$

where

$A$ is of the form $A_{1t_1} \land \ldots \land A_{qt_q} \land A_{q+1t_{q+1}}$ and $B$ is of the form $A_{1u_1} \land \ldots \land A_{qu_q}$ with $B \land A_{q+1t_{q+1}} = A$

Since $X_q$ forms a Markov Chain, a study of absorbing sets as well as entry and exit times could be made. We choose not to deal with these rather general questions but rather to pause some specific problems such as: what is the probability of realizing for the first time, as we interrogate $KS_{q+1}$, that the answer is not in the frame of discernment. What is the probability of getting no information from $KS_{q+1}$? (Of course we assume that $KS_1, KS_2, \ldots, KS_q$ were already interrogated). The answer to such questions has been derived and is given below.

$$Pr\left(\text{Realizing for the 1st time at time } q+1 \text{ that answer not in frame of discernment}\right)$$

where the sum is over all local elements $B_{q+1,i}$ of $m_{q+1}$ such that

$$A_{1t_1} \land \ldots \land A_{qt_q} B_{q+1,i} = \emptyset \text{ yet } A_{1t_1} \land \ldots \land A_{qt_q} \neq \emptyset$$

That is, the averaged $Bel_{q+1}$ of being outside the range of $X_q$ when $KS_{q+1}$ is interrogated.

$$Pr\left(\text{No Info. from } KS_{q+1}\right) = \sum_{B} m_{q+1}(B_{q+1,i}) \mu_q(A)$$

where the sum is over $B_{q+1,i}$ superset of $A$, and $A$ is of the form $A_{1t_1} \land \ldots \land A_{qt_q}$

Using the transition probabilities we have

$$Pr\left(X_{q+1} = A_{q+1t_{q+1}} \land \ldots \land A_{1t_1} \ldots X_1 = A_{1t_1}\right)$$

$$= Pr\left(X_1 = A_{1t_1}\right) \left(\sum m_2(B)\right) \left(\sum m_{q+1}(B)\right)$$

where the first sum is over all $B \land A_{1t_1} = A_{2t_2} \land A_{1t_1}$, and the last sum is over all $B$ such that $B \land A_{qt_q} \land \ldots \land A_{1t_1} = A_{q+1t_{q+1}} \land \ldots \land A_{1t_1}$

All of this points out that it is very important to carefully evaluate $X_q$
TRUNCATING THE INFORMATION

Let \( \{A_1, A_2, \ldots \} \) be distinct focal elements of \( m_1, m_2, \ldots \). We can view \( KS_t \) as confirming \( A_i \) to the degree \( \alpha_{ti} \). Let \( \alpha_{ti*} \) be the largest \( \alpha_{ti} \) for \( t \) fixed \( (i* \text{ depends on } t) \).

Now view \( KS_t \) as confirming \( A_i* \) to the degree \( \alpha_{ti*} \) and ignore the rest of the information yielded by \( KS_t \) (i.e., take only the highest confirmation of \( KS_t \)). If \( s_1, s_2, \ldots s_k \) supported \( A_i* \), \( A_i* \) is supported to the degree \( 1 - (1-s_1) (1-s_2) \ldots (1-s_k) \).

If the resulting mass on \( A_i \) is

\[
Evi_i(A_i) = p_i, \quad \text{we set} \quad Evi_i \{A_1, \ldots A_n\} = r_i, \quad \text{where} \quad p_i + r_i = 1
\]

The rationale for doing this is to trust our estimate of the mass on each \( A_i \), which came from the highest degrees of confirmation, and to ignore the rest, i.e., spread the rest of the mass on all the possibilities.

It can be shown, see [1] that

\[
Bel(A_i) = Kp_j \prod_{i \neq j} r_i \quad \text{Where}
\]

\[
K = 1 \left( \prod_j r_j \left( 1 + \sum_i p_i \right) \right)
\]

At stage \( q+1 \), we then pick \( A_j \) maximizing \( Bel \) coming from \( m_1, m_2, \ldots m_q+1 \). In this way, prior information given by \( m_1, \ldots m_q \), as well as the current information yielded by \( m_q + 1 \), is taken into account.

We can extend this to keeping the two highest confirmations \( KS_t \), as mentioned earlier, assigns \( \alpha_{ti*} \) to \( A_i* \) and if the second highest \( \alpha_{ti} \), call it \( \beta_{i*} \), is assigned to \( \{A_i \mid i \neq i* \} \) (spread around the 2nd highest).

The rest \( 1 - \alpha_{ti*} - \beta_{i*} \) is assigned to \( \{A_1, \ldots A_n\} \).

It can be shown, see [1], that

\[
Bel(A_j) = K \left[ p_j \prod_{i \neq j} d_i + r_j \prod_{i \neq j} c_i \right]
\]

\[
K = \prod_j d_j \left( 1 + \sum_i p_i \right) - \prod_j c_j
\]

Where \( p_i + c_i + d_i = 1 \).
APPLYING CLASSICAL DECISION THEORY TO SELECT VALUES FOR $X_{q+1}$

If $KS_1, \ldots, KS_q$ yield enough information so that

$$P_r\left(X_q = A_{1t_1} \cap \ldots \cap A_{qt_q}\right) = \mu_q\left(A_{1t_1} \cap \ldots \cap A_{qt_q}\right)$$

can be trusted then, maximize over $B_{q+1,i}$ the following expression:

$$\sum Bel\left(A_{1t_1} \cap \ldots \cap A_{qt_q} \cap B_{q+1,i}\right) \mu_q\left(A_{1t_1} \cap \ldots \cap A_{qt_q}\right)$$

$Bel\left(q+1\right)$ is generated by $m_1 \oplus \ldots \oplus m_{q+1}$

We view this as making a decision in the environment $A_{1t_1} \cap \ldots \cap A_{qt_q}$ assuming the probabilities are known. Picking the alternative $B_{q+1,i}$ yields a payoff of $Bel\left(A_{1t_1} \cap \ldots \cap A_{qt_q} \cap B_{q+1,i}\right)$ and we maximize the averaged payoff. What if $KS_1, \ldots, KS_q$ are not too reliable? We know the patterns but we are not sure about their probabilities. Pick the alternative $B_{q+1,i}$ so as to maximize the minimum of

$$Bel\left(A_{1t_1} \cap \ldots \cap A_{qt_q} \cap B_{q+1,i}\right) - Bel\left(A_{1t_1} \cap \ldots \cap A_{qt_q} \cap \bigvee_{j \neq i} B_{q+1,i}\right)$$

Here the minimum is taken over all environments $A_{1t_1} \cap \ldots \cap A_{qt_q}$. $Bel\left(q+1\right)$ is generated by $m_{q+1}$. The motivation is that picking the minimum represents the worst environment for payoff of alternative $B_{q+1,i}$ over competing alternatives. Picking the maximum represents then the maximum gain. This approach is pessimistic in nature (going to the worst environment and then making the best of a bad situation). At the other end of the spectrum, the maximum of the maximum represents the optimistic attitude.

Picking a convex combination of the two represents a compromise.

Another approach yet is to minimize the largest regret. Let

$$T(A_{1t_1} \cap \ldots \cap A_{qt_q}) = D_i \left(A_{1t_1} \cap \ldots \cap A_{qt_q}\right) - Max_i D_i \left(A_{1t_1} \cap \ldots \cap A_{qt_q}\right)$$

Where $D_i$ denotes the above difference of beliefs. $T$ measures the regret of picking the alternative $B_{q+1,i}$ over the best alternative ($T \leq 0$) Picking the minimum over the environment, produces the largest regret. Picking then the maximum, minimizes that largest regret.
DECISION MAKING AND TRUNCATING THE INFORMATION

Here, we believe that some patterns are a definite possibility. We also want to ignore the rest of the patterns, also we do not trust any probabilities functions associated with $KS_1, ... KS_q$. Going back to the previous section, we see that the four previous algorithms are well defined if we restrict the environments $A_{1t_1} \land ... \land A_{qt_q}$ to a fixed set $P$. We now refine this by allowing $P$ to be a fuzzy set. Thus

$$P = \sum a_{q(t_1 \ldots t_k)}^q A_{1t_1} \land ... \land A_{qt_q}$$

Here $a_{q(t_1 - t_k)}$ denotes the degree of membership of $A_{1t_1} \land ... \land A_{qt_q}$ in $P$.

We now must define an appropriate fuzzy set of payoffs. In the case of an optimistic or pessimistic or a somewhere in between attitude, we consider

$$C_i(P) = \sum a_{q(t_1 \ldots t_k)}^q \left[ Bel_{q+1} \left( A_{1t_1} \land ... \land A_{qt_q} \land B_{q+1,i} \right) - Bel_{q+1} \left( A_{1t_1} \land ... \land A_{qt_q} \land B_{q+1,j} \right) \right]$$

We need to be able to take the minimum or maximum element of a fuzzy set.

If $A = \sum a_i \left| a_i \right|$ Set

$$\psi(\lambda) = \text{Min} \left\{ a_i \left| a_i \geq \lambda \right. \right\}$$

The minimum of $A$ is defined as

$$\int_0^1 \psi(\lambda) d\lambda$$

This coincides with minimum in the crisp case and is defined in [11]. Again the 4 algorithms defined earlier now make sense.

THE GENERAL CASE

We interrogate $KS_1, ... KS_q$ and split the corresponding patterns into disjoint blocks $P_k$. The blocks could correspond to classifications such as highly likely, likely, somewhat likely patterns, etc. ... We also assume $P_k$ are fuzzy sets (is $A_{1t_1} \land ... \land A_{qt_q}$ highly likely?). We set

$$P_j = \sum a_{q(t_1 \ldots t_k)}^q A_{1t_1} \land ... \land A_{qt_q}$$
Also we assign masses to each block, reflecting the weight put on the blocks (this reflects the trust put on the corresponding patterns in the class). Let

\[ m(P_j) = a_j \]

The first criteria, for example, would maximize over the alternatives \( B_{q+1,i} \)

\[
\sum_k l_i(P_k) m(P_k) \text{ where } l_i(P_k) \text{ is the minimum of the fuzzy set}
\]

\[
\sum a_{q,t_1...t_k} \left[ \text{Bel}_{q+1} \left( A_{1,t_1} \wedge ... \wedge A_{q,t_q} \wedge B_{q+1,i} \right) - \text{Bel}_{q+1} \left( A_{1,t_1} \wedge ... \wedge A_{q,t_q} \wedge i \neq j \right) B_{q+1,j} \right]
\]

It is clear that the other three algorithms generalize to this situation. The sets are replaced by "averages" and the minimum and maximum need to be taken over fuzzy sets, as explained earlier. For other methods available in the setting of decision making, the reader is referred to [5], [6], and [7].

**TOWARD A GENERAL THEORY OF MULTIPLE KNOWLEDGE SYSTEMS**

The previous discussion points out the importance of building a formal theory for the multiple knowledge systems setting. Our present work generalizes the situation described in [8] and constitutes the first steps toward such a theory. Our basic assumptions are:

(i) Our knowledge systems are independent

(ii) Each knowledge system specializes in one feature

(iii) Each feature may have several knowledge systems assigned to it.

We may not want to access all KS's and therefore, we have to solve the following problem:

The access problem: Which sets of KS's does one access? (some KS may run in parallel)

In what order does one access these sets?

We have performance parameters such as reliability, expense, response time, etc... Information regarding these parameters are contained in special KS's called CKS's (Control Knowledge Systems).

Each CKS specializes in one performance parameter

One performance parameter may correspond to several CKS's.

Each KS has two components:

a) The observational component which reports on the value of a specific feature. It may return a value or a probability distribution over the set of possible feature values (e.g., red or .8/red + .2/blue).

b) The judgemental component which reports on how likely it is that the true answer lies in some set of possible answers given that a specific value of a feature has been observed.
We define a control strategy to be a sequence of performance parameters specifications. This generates an access to a set of KS's. After these KS's have been used, the belief structure of the frame of discernment is updated. Then stopping rules are looked at. If stopping criterias are not met, we go to the next control strategy.

If all control strategies have been exhausted, a decision is made as to what the probable answer is.

Access Policy

Each control strategy is a list of performance objectives. On the $l^{th}$ control strategy, let $\Theta_l$ denote all available KS's $\Theta_1 \supset \Theta_2 \supset \ldots \supset \Theta_l \ldots$ as we don't want to reuse the same KS's (we want to have independent sources of information). The decision as to what KS's to use is made on information contained in the CKS's.

Each CKS has two components:

a) Component - $A$ which decides on what are the best subsets of $\Theta_l$ to consider when the value of performance objective $P_j = p_j^k$  

i.e., If $CKS^1_j, \ldots CKS^{r_j}_j$ specialize on performance $j$

Component A computes all

$$u_{jt}^{(k)} : \Theta_l \rightarrow [0,1]$$

represent

$$u_{jt}(p_j^k) \quad 1 \leq t \leq r_j$$

We define $u_{jt}^{(k)}(B) = 0$ if $B$ contains any pair of KS'S which can't run in parallel

Thus $u_{jt}^{(k)}$ is non-zero only on sets of KS's that run in parallel

b) Component - $B$ which makes a probabilistic judgement of what is the best value of performance $P_j$. i.e., for $CKS_t$,  

$$\beta_{jk}^{(t)} = Pr[P_j = p_j^k \mid CKS^t_j]$$

Let $B \subset \Theta_l$ (set of available KS's) Define

$$n_j(B \mid CKS^{t}_j) = \sum_k \beta_{jk}^{(t)} u_{jt}^{(k)}(B)$$

Here $u_{jt}^{(k)}(B)$ is given by component A of $CKS^t_j$ for each value $p_j^k$ and $\beta_{jk}^{(t)}$ is given by component - $B$ of $CKS^t_j$ for each value $p_j^k$. In other words, the above expression represents how good, on the average, the set B is as determined by $CKS^t_j$

We now want to mesh all the CKS for a fixed performance.

$$n_j(B \mid CKS^1_j, \ldots CKS^{r_j}_j) = \oplus_{t=1}^{r_j} n_j(B \mid CKS^t_j)$$
Now to mesh all the performance parameters
\[ n(B \mid \text{all the CKS involved}) = \bigoplus_{t=1}^{s} n_{j_t}(B \mid CKS_{j_t}^{1}, \ldots, CKS_{j_t}^{r}) \]

Now look at all \( B \subseteq \Theta_l \) made up of KS's that can be accessed in parallel. For such \( B \)'s maximize \( Bel(B) - Bel(\neg B) \)

At this point, we have picked a set of KS's to run in parallel.

We now have to interrogate the KS's that are in \( B \) and update our belief structure on the frame of discernment (and go from \( \Theta_l \) to \( \Theta_{l-1} \)).

Now we have \( KS_i \) Recall it has an observational and a judgemental component

The judgemental component is represented by
\[ \nu^{(b)}_{it} : 2^Q \rightarrow [0, 1] \]

This represents
\[ \nu^{k}_{it}(A \mid f_{i_t}^{k}) \]

the degree of belief that \( A \) is the smallest set containing the right answer, given feature value \( f_{i_t}^{k} \)

The observational component returns either a single value but in more complex case, a probability distribution. The notation:
\[ KS_{i_l}^{1}(i_1), \ldots, KS_{i_l}^{s_l}(i_{s_l}) \]

means that KS's on the \( l \)th strategy that are in the selected set \( B \) report on features \( i_1, i_2, \ldots, i_{s_l} \)

The observational components report
\[ \alpha_{i}^{k}(i_t, t) = Pr[f_{i_t}^{k} \mid KS_{i}^{t}(i_t)] \]

(the \( l \) index refers to the \( l \)th control strategy)

Thus, we define masses (over fixed features)
\[ m_{it}(A \mid KS_{i}^{t}(i_t)) = \sum_{k} \alpha_{i}^{k}(i_t, t) \nu^{k}_{i_t}(A) \]

(the averaged mass assigned by the judgemental component)

We now mesh over all features (determined by \( B \), the selected set)
\[ m_{i}(A \mid KS_{i}^{1}, \ldots, KS_{i}^{S_l}) = \bigoplus_{t=1}^{S_l} m_{it}(A \mid KS_{i}^{t}(i_t)) \]

At the end of the \( L \) control strategy, our total information is summed by
\[ m_L(\bullet \mid \text{all KS's involved}) = \bigoplus_{l=1}^{L} m_{i}(\bullet \mid KS_{i}^{1}, \ldots, KS_{i}^{S_l}) \]

We now must deal with the decision rule of what object must be selected as a plausible answer.
Step 1:
Let \( a \) in \( \Theta Q \) be the element maximizing

\[
Bel_L(a) - Bel_L(\neg a)
\]

If \( L \) denotes last control strategy, pick 'a'
else go to Step 2

Step 2:

a) If \( \delta_1 > 0 \) denotes some fixed threshold
Pick 'a' if \( Bel(a) > \delta_1 \)

b) If \( \delta_2 > 0 \) denotes some fixed number
Pick 'a' if \( Pls_L(a) - Bel(a) < \delta_2 \)

c) Combined 'a' and 'b'

If a doesn't satisfy the criterion, go to the next control strategy. The rationale for the stopping rule is 'c' is that we would like the belief in 'a' to exceed some threshold and have uncertainty relative to a drop below some predefined level.

It is clear that much research remains to be done. For example, degradation of the information contained in the KS's has not been considered in the last part of this report. This and additional problems will be addressed in future work. Finally, for applications of the Dempster-Shafer approach to artificial intelligence, the reader is referred to [3] and [4].
REFERENCES


