SOLVING PROBLEMS BY INTERROGATING SETS OF KNOWLEDGE SYSTEMS: TOWARD A THEORY OF MULTIPLE KNOWLEDGE SYSTEMS

FINAL REPORT

NASA/ASEE Summer Faculty Fellowship Program -- 1989

Johnson Space Center

Prepared By:	Andre de Korvin, Ph.D.
Academic Rank:	Professor
University & Department:	University of Houston - Downtown
	Applied Mathematical Sciences
	Houston, Tx. 77002

NASA / JSC

Directorate:	Mission Support
Division:	Mission Planning and Analysis
Branch:	Technology Development and Applications
JSC Colleague:	Robert Lea, Ph.D.
Date Submitted:	August 11, 1989
Contract Number:	NGT 44-001-800

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:

VELOCITY	B	P	S	{BP}	{BS}	{PS}	{BPS}
0 - 100	.5	.1	.1	.2	.04	.04	.02
101 - 200	0	.4	.1	0	0	.5	0
201 - 500	0	.5	.1	0	0	.4	0
> 500	0	.1	.7	0	0	.2	0

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:

COLOR	B	P	S	{BP}	{BS}	{PS}	{BPS}
SILVER	.05	.6	.05	.1	0	.15	.05
WHITE	.1	.1	.05	.5	.05	.15	.05
RED	.1	.1	.1	.2	.2	.2	.1
BLUE	.1	.1	.1	.2	.2	.2	.1
RED-BLUE	.04	.04	.8	0	.05	.05	.02
OTHER	.6	.1	0	.3	0	0	0

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 Θ into $[0, 1]$ satisfying the following conditions:

(i) $\quad m(\varnothing) = 0$

(ii) $\quad \sum m(A) = 1 \quad$ where the sum is over all subsets of Θ

If m_1 & m_2 are two mass functions, we define

$$(m_1 \oplus m_2)(C) = \sum_{A \wedge B = C} m_1(A) m_2(B) \bigg| (1 - k)$$

Where k is the conflict

$$k = \sum_{A \wedge B = \varnothing} m_1(A) m_2(B)$$

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

$Bel(A) = \sum m(B) \quad$ over all *sets B such that* $B \subset A$. Also we define the plausability by

$$Pls(A) = \sum m(B) \quad \text{over all sets } B \text{ such that } B \wedge A \neq \varnothing$$

Now if the lth sighting takes place at time t_l, set $\quad dt_l = t_l - t_{l-1}$

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

$$(i)\, 0 \leq \lambda(dt_l) \leq 1, \quad dt_l > 0$$

$$(ii)\, \lambda(dt_l) \text{ non decreasing as a function of } dt_l$$

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

$$m_1(\bullet) = m_1(\bullet)$$

a) $\quad m_l(\bullet) = \lambda(dt_l) m_l(\bullet) + (1 - \lambda(dt_l)) m_{l-1}(\bullet)$

b) $\quad m_l(\bullet) = \lambda(dt_l) m_l(\bullet) + (1 - \lambda(dt_l)) m_{l-1}(\bullet)$

If λ is high i.e., dt_l high, go with the current observation

If λ 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:

$$\lambda \ (dt_l) = \begin{cases} dt_l \ / \ 300 & if \ dt_l \ \leq 300 \\ 1 & otherwise \end{cases}$$

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:

SIGHTING	TIME	dt_l	VELOCITY	COLOR
1	1:00 p.m.	0	101 - 200	WHITE
2	1:01 p.m.	60	201 - 500	WHITE
3	1:29 p.m.	1740	0 - 100	OTHER

The combined masses, not time adjusted are given below:

SIGHTING	B	P	S	{BP}	{BS}	{PS}	{BPS}
1 $m_1(\cdot)$	0	.7750	.1	0	0	.125	0
2 $m_2(\cdot)$	0	.8101	.0886	0	0	.1013	0
3 $m_3(\cdot)$	.811	.1024	0	.0866	0	0	0

The combined masses, time adjusted are given below:

SIGHTING	B	P	S	{BP}	{BS}	{PS}	{BPS}
1 $m_1(\cdot)$	0	.775	.1	0	0	.125	0
2 $m_2(\cdot)$	0	.78202	.09772	0	0	.12026	0
3 $m_3(\cdot)$	.811	.1024	0	.0866	0	0	0

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:

SIGHTING	OBJECT	Bel ($\cdot$)	Bel ($\neg\,\cdot$)	Bel ($\cdot$) - Bel ($\neg\,\cdot$)	CLASSIFICATION
1	B	0	1	-1	
	P	.775	.1	.675	P
	S	.1	.775	-.675	
2	B	0	1	-1	
	P	.78202	.09772	.6843	P
	S	.09772	.78202	-.6843	
3	B	.811	.1024	.7086	
	P	.1024	.811	-.7086	B
	S	0	1	-1	

The rationale for the table above is that Bel ($\bullet$) - Bel ($\neg\,\bullet$) 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 $(m_1 \ominus m_2)$ (C). (Thus the empty set may pick up mass).

We now shift somewhat our perspective. Consider Knowledge Systems $KS_1, KS_2, \dots KS_n$

KS_j reads the Jth feature and interprets its value to be f_j^i with a probability a_{ji}
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 Θ by
$$m_j\,(A_{ji}) = a_{ji} \text{ where}$$

A_{ji} denotes all objects of Θ whose j^{th} feature has value f_j^i

After we have interrogated $KS_1, KS_2, \dots, KS_q$ possible answers are in sets
$A_{1t_1} \wedge A_{2t_2} \wedge \dots \wedge A_{qt_q}$. Let X_q be set-valued variables whose values are $A_{1t_1} \wedge \dots \wedge A_{qt_q}$ ($t_1, \dots, t_q$ 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

$$Pr\left(X_{q+1} = A_{1t_1} \wedge \dots \wedge A_{qt_q} \wedge A_{q+1\,t_{q+1}} \,\Big|\, X_q = A_{1t_1} \wedge \dots \wedge A_{qt_q}\right) = \sum m_{q+1}\left(B_{q+1,i}\right)$$

Where the sum is over sets $B_{q+1,i}$ such that

$$A_{1t_1} \wedge ... \wedge A_{qt_q} \wedge B_{q+1,i} = A_{1t_1} \wedge ... \wedge A_{qt_q} \wedge A_{q+1t_{q+1}}$$

If

$$Pr\left(X_q = A_{1t_1} \wedge ... \wedge A_{qt_q}\right) = \mu_q\left(A_{1t_1} \wedge ... \wedge A_{qt_q}\right) \quad \text{then}$$

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

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

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 $KS_1, KS_2, ..., KS_q$ were already interrogated). The answer to such questions has been derived and is given below.

Pr (Realizing for the 1st time at time $q+1$ that answer not in frame of discernment)

$$= \sum m_{q+1}\left(B_{q+1,i}\right)\mu_q\left(A_{1t_1} \wedge ... \wedge A_{qt_q}\right)$$

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

$$A_{1t_1} \wedge ... \wedge A_{qt_q} B_{q+1,i} = \varnothing \quad \text{yet} \quad A_{1t_1} \wedge ... \wedge A_{qt_q} \neq \varnothing$$

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

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

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

Using the transition probabilities we have

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

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

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

All of this points out that it is very important to carefully evaluate X_q

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 α_{ti} Let α_{ti*} be the largest α_{ti} for t fixed $(i^*$ depends on $t)$.

Now view KS_t as confirming A_{i*} to the degree α_{ti*} and ignore the rest of the information yielded by KS_t (i.e., take <u>only</u> 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

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

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

$$Bel(A_j) = Kp_j \prod_{i \neq j} r_i \quad \text{Where}$$

$$K = 1 \left| \left[\prod_j r_j \left(1 + \sum_i p_i \middle| r_i \right) \right] \right.$$

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 ealier, assigns α_{ti*} to A_{i*} and if the second highest α_{ti}, call it β_{tj*}, is assigned to

$$\{A_i \ / \ i \neq i^*\} \quad \text{(spread around the 2nd highest)}$$

The rest $1 - \alpha_{ti*} - \beta_{tj*}$ is assigned to $\{A_1, \ldots A_n\}$

It can be shown, see [1], that

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

$$K = \prod_j d_j \left[1 + \sum_i p_i \middle| r_i \right] - \prod_j c_j$$

Where $p_i + c_i + d_i = 1$.

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

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

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

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

$Belief_{q+1}$ is generated by $\quad m_1 \oplus \ldots \oplus m_{q+1}$

We view this as making a decision in the environment $A_{1t_1} \wedge \ldots \wedge A_{qt_q}$ assuming the probabilities are known. Picking the alternative $B_{q+1,i}$ yields a payoff of $Belief_{q+1}\left(A_{1t_1} \wedge \ldots \wedge A_{qt_q} \wedge B_{q+1,i}\right)$ 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

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

Here the minimum is taken over all environments $A_{1t_1} \wedge \ldots \wedge A_{qt_q}$. Bel_{q+1} 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

$$T\left(A_{1t_1} \wedge \ldots \wedge A_{qt_q}\right) = D_i\left(A_{1t_1} \wedge \ldots \wedge A_{qt_q}\right) - \overset{Max}{i} D_i\left(A_{1t_1} \wedge \ldots \wedge A_{qt_q}\right)$$

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 $KS_1, \ldots KS_q$. Going back to the previous section, we see that the four previous algorithms are well defined if we restrict the environments $A_{1t_1} \wedge \ldots \wedge A_{qt_q}$ to a fixed set P. We now refine this by allowing P to be a fuzzy set. Thus

$$P = \sum a_{q(t_1 \ldots t_k)} \Big| A_{1t_1} \wedge \ldots \wedge A_{qt_q}$$

Here $a_{q(t_1 - t_k)}$ denotes the degree of membership of

$A_{1t_1} \wedge \ldots \wedge A_{qt_q}$ in P.

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

$$C_i\left(P\right) = \sum a_{q(t_1 \ldots t_k)} \Big| \left[Bel_{q+1}\left(A_{1t_1} \wedge \ldots \wedge A_{qt_q} \wedge B_{q+1,i}\right) - Bel_{q+1}\left(A_{1t_1} \wedge \ldots \wedge A_{qt_q} \wedge \bigvee_{j \neq i} B_{q+1,j}\right)\right]$$

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

If $\quad A = \sum a_i \Big| a_i \quad$ Set

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

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 $KS_1, \ldots KS_q$ 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 $A_{1t_1} \wedge \ldots \wedge A_{qt_q}$ highly likely?). We set

$$P_j = \sum a_{q(t_1 \ldots t_k),j} \Big| A_{1t_1} \wedge \ldots \wedge A_{qt_q}$$

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

$$m\left(P_j\right) = a_j$$

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

$$\sum_k l_i\left(P_k\right) m\left(P_k\right) \quad \text{where} \quad l_i\left(P_k\right) \quad \text{is the minimum of the fuzzy set}$$

$$\sum a_{q(t_1 \dots t_k), k} \left| \left[Bel_{q+1}\left(A_{1t_1} \wedge \dots \wedge A_{qt_q} \wedge B_{q+1,i}\right) - Bel_{q+1}\left(A_{1t_1} \wedge \dots \wedge A_{qt_q} \wedge \underset{i \neq j}{\vee} B_{q+1,j}\right)\right]\right|$$

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 lth control strategy, let Θ_l denote all <u>available</u> KS's $\Theta_1 \supset \Theta_2 \supset ... \supset \Theta_l ...$ 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 Θ_l to consider when the value of performance objective P_j p_j^k

i.e., If $CKS_j^1, ... CKS_j^{rj}$ specialize on performance j

Component A computes all

$$u_{jt}^{(k)} : 2^{\Theta_l} \rightarrow \left[0, 1 \right] \quad \text{which}$$

represent

$$u_{jt}\left(\bullet \middle| p_j^k \right) \quad 1 \le t \le r_j$$

We define $u_{jt}^{(k)}\left(B \right) = 0$ 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,

$$\beta_{jk}^{(t)} = Pr\left[P_j = p_j^k \middle| CKS_j^t \right]$$

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

$$n_j\left(B \middle| CKS_j^t \right) = \sum_k \beta_{jk}^{(t)} u_{jt}^{(k)}\left(B \right)$$

Here $u_{jt}^{(k)}\left(B \right)$ is given by component A of CKS_j^t for each value p_j^k and $\beta_{jk}^{(t)}$ is given by component $- B$ of CKS_j^t 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_j^t

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

$$n_j\left(B \middle| CKS_j^1, ... CKS_j^{rj} \right) = \overset{r_j}{\underset{t=1}{\oplus}} n_j\left(B \middle| CKS_j^t \right)$$

Now to mesh all the performance parameters

$$n\left(B \middle| \text{all the CKS involved}\right) = \overset{s}{\underset{t=1}{\oplus}} \; n_{j_t}\left(B \middle| CKS_{j_t}^1, \ldots CKS_{j_t}^{r_{j_t}}\right)$$

Now look at all $B \subset \Theta_l$ 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 Θ_l to Θ_{l-1})

Now we have KS_i^t Recall it has an observational and a judgemental component

The judgemental component is represented by

$$v_{it}^{(k)} : 2^{\Theta}Q \to \left[0, 1\right]$$

This represents

$$v_{it}^k\left(A \middle| f_i^k\right)$$

the degree of belief that A is the smallest set containing the right answer, given feature value f_i^k

The observational component returns either a single value but in more complex case, a probability distribution. The notation:

$$KS_l^1\left(i_1\right), \ldots KS_l^{s_l}\left(i_{s_l}\right)$$

means that $KS's$ on the lth strategy that are in the selected set B report on features $i_1, i_2, \ldots i_{s_l}$
The observational components report

$$a_l^k\left(i_t, t\right) = Pr\left[f_{i_t}^k \middle| KS_l^t\left(i_t\right)\right]$$

(the l index refers to the lth control strategy)
Thus, we define masses (over fixed features)

$$m_{lt}\left(A \middle| KS_l^t\left(i_t\right)\right) = \sum_k a^k\left(i_t, t\right) v_{i_t t}^k\left(A\right)$$

(the averaged mass assigned by the judgemental component)
We now mesh over all features (determined by B, the selected set)

$$m_l\left(A \middle| KS_l^1, \ldots KS_l^{sl}\right) = \overset{sl}{\underset{t=1}{\oplus}} \; m_{lt}\left(A \middle| KS_l^t\left(i_t\right)\right)$$

At the end of the L control strategy, our total information is summed by

$$m_L\left(\bullet \middle| \text{all KS's involved}\right) = \overset{l}{\underset{l=1}{\oplus}} \; m_l\left(\bullet \middle| KS_l^1, \ldots KS_l^{sl}\right)$$

We now must deal with the decision rule of what object must be selected as a plausible answer.

Step 1:

Let a in Θ_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

1.   J. A. Barnett (1981), "Computational methods for a Mathematical Theory of Evidence" Proceedings IJCAI-81 Vancouver, B.C. pp 868-875.

2.   Davis, L.S., Rosenfeld, A. "Computer Vision Systems". Hanson, A.R., and Riseman, E.M. eds,.101, Academic Press, New York (1978).

3.   Garvey, T.D., (1976) "Perceptual Strategies For Purposive Vision" SRI International Tech. Note No. 117, Artificial Intelligence, Ctr, Menlo Park, California.

4.   J. Gordon and E. H. Shortliffe (1984), "The Dempster-Shafer Theory of Evidence Rule-based Expert Systems: The MYCIN experiments of the Stanford heuristic programming project" Eds B. G. Buchanan and E. H. Shortliffe. pp 272-292. Addison - Wesley.

5.   Kleyle, R.M. and de Korvin, A. (1984), "Switching Mechanisms in a Generalized Information System", Cybernetics and Systems, Vol. 15, pp. 145-167.

6.   Kleyle, R.M. and de Korvin, A. (1986), "A Two Phase Approach to Making Decisions Involving Goal Uncertainty", Journal of Information Science, Vol. 12, pp. 161-171.

7.   Kleyle, R.M., de Korvin, A. "A Belief Function Approach To Information Utilization In Decision Making". Accepted for publication in the Journal of the American Association for Information Science.

8.   A. de Korvin, R. Kleyle and R. Lea, "An Evidential Approach to Problem Solving When a Large Number of Knowledge Systems are Available", submitted.

9.   H. T. Nguyen (1978), "On random Sets and Belief Functions", J. of Math Analysis and Appl. 65, pp 531-542.

10.   Schafer, G. (1976), A Mathematical Theory of Evidence. Princeton, NJ, Princeton University Press.

11.   R. R. Yager (1986), "A General Approach to Decision Making With Evidential Knowledge", Uncertainty in Artificial Intelligence. Eds L. N. Kanac and J. F. Lemmer, North Holland, pp 317-327.

12.   Zadeh, L.A. (1986), "A Simple View of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination", The AI Magazine, pp. 85-90.