STATISTICAL DECISION FUSION THEORY

May 1999

Michael B. Hurley
MIT Lincoln Laboratory
244 Wood St., Lexington, MA 02420-9185

ABSTRACT

By combining information theory, statistical decision theory, and maximum entropy to address the decision fusion problems, a statistical decision fusion theory is obtained. The theory explains why decision fusion is so difficult and why the performance of decision fusion systems does not always meet expectations. The theory suggests how statistical decision systems such as the conceptual "Family of Systems" might be designed. The theory clarifies why independent subsystems are desired in data fusion systems. A decision fusion function is obtained from the theory for fusing independent decision subsystems. An examination of the characteristics of the fusion function shows that it can handle decision results from subsystems operating at different hierarchical levels in the sets of decisions and prior classes. This fusion arises naturally without the need to incorporate additional principles to convert decisions and prior classes to other hierarchical levels. In the design of decision fusion systems, the subsystems can be designed to operate at their own natural levels in the set hierarchy while the fusion can be designed to operate at the most descriptive level. The fusion function can also be applied to time evolving decision fusion systems and cast as a Bayes-Markov non-linear filtering process. The resulting process is similar to Kalman filtering and allows for the design of decision systems that de-weights the influence of previous results when new information is processed. In summary, the characteristics of the decision fusion theory have only just begun to be explored and a rich variety of decision fusion system designs await discovery.

1. Introduction

At the 1998 IRIS National Symposium on Sensor and Data Fusion, work toward a decision fusion theory was presented.[1] The theory was obtained from a melding of statistical decision theory and information theory. This paper summarizes the progress made in the last year to extend the theory. It has been discovered that the new theory readily handles the fusion of information from different levels of

abstraction in the set hierarchies, a capability highlighted as a superior characteristic of Dempster-Shafer evidential reasoning.[2] In counterpoint, the fusion of mixed hierarchy data with traditional Bayesian-based fusion techniques has been problematic. Given the Bayesian roots of the new theory, it can be stated that a Bayesian decision theory has been discovered that solves the mixed hierarchy problem.

It will also be shown how the decision fusion function can be recast as a recursive algorithm. The recursive algorithm bears some resemblance to the Kalman filter algorithm. This parallel suggests that techniques to decay or degrade old information may exist and that the recursive algorithm may be extended to account for time dependent information loss, correlated data, and Bayes-Markov processes. In addition, the recursive formulation provides new insights into the interpretation of the prior probabilities in statistical decision theory.

2. The Decision Fusion Function

The statistical decision theory problem can be defined as the selection of a decision γ from a possible set of decisions Γ, given measurements v in a feature space V containing distributions $F(v \mid s)$ of each prior class s, that together compose a set S. The prior probabilities $\sigma(s)$ for each prior class s adjust the conditional probabilities and affect the resulting decisions. Many applications of statistical decision theory have focused on the identification problem, which has a one-to-one correspondence between the members of the prior set S and the decision set Γ. The information theoretic derivations depend upon this one-to-one correspondence, although the theory can be applied to general decision problems that do not possess this correspondence.

The integration of statistical decision theory and information theory results in the cost function,

$$c(\gamma, v_1) = - \sum_{S} \frac{\sigma(s) F(v_1 \mid s)}{\sum_{S'} \sigma(s') F(v_1 \mid s')} \ln \left(\frac{\int_{V_\gamma} \sigma(s) F(v \mid s) dv}{\sum_{S'} \sigma(s') \int_{V_\gamma} F(v \mid s') dv} \right) \tag{1}$$

which gives a cost for each decision γ, given a measurement v_1. A decision system is designed by assigning each region in feature space to a specific decision. Through the application of information theory, the assignment of the decisions to the feature-space sub-volumes V_γ is done so as to minimize the information loss (equivocation) between the prior set S and the decision set Γ.

The cost function consists of two terms, the conditional probability of a measurement v_1 given prior class s,

$$p(s \mid v_1) = \frac{\sigma(s) F(v_1 \mid s)}{\sum_{S'} \sigma(s') F(v_1 \mid s')}, \tag{2}$$

and the conditional probability of prior class s given decision γ,

$$p(s \mid \gamma) = \frac{\sigma(s) \int_{V_\gamma} F(v \mid s) dv}{\sum_{s'} \sigma(s') \int_{V_\gamma} F(v \mid s') dv}. \quad (3)$$

A belief matrix may be constructed from the collected set of conditional probabilities of Equation 3.

The cost function $c(\gamma \mid v_1)$ in Equation 1 can be viewed as a logarithmic distance measure from the probability $p(s \mid \gamma)$ to the probability $p(s \mid v_1)$ in a probability space with N_S dimensions. With the costs determined, the optimal decision is the one with the minimum cost,

$$D(v_1) = \min_{\gamma}(c(\gamma, v_1)). \quad (4)$$

In the event of tie minimum costs, the decision rule can be extended for decision systems with one-to-one correspondence between the prior class elements and the decision elements. In these systems, the tie may be broken by selecting the decision γ that associates with the prior class s with the greatest conditional probability $p(s \mid v_1)$. If tie decisions still remain and a forced decision is required, random selection can be used to force the decision.

A decision fusion cost function can be obtained from Equation 1. For independent, orthogonal decision subsystems, it is assumed that the probability density functions $F(v \mid s)$ are separable into products of probability density functions in k orthogonal feature subspaces v_j,

$$F(v \mid s) = \prod_{j=1}^{k} F_j(v_j \mid s). \quad (5)$$

In addition, it is assumed that the decision volume integrals can be decomposed into products of integrals over the decision sub-volumes $V_{j\gamma}$,

$$\int_{V\gamma} F_j(v_j \mid s) dv_j = \prod_{j=1}^{k} \int_{Vj\gamma} F_j(v_j \mid s) dv_j \,. \quad (6)$$

In general, the integrals of Equation 6 construct a confusion matrix. The decomposition assumes that the full-feature-space belief matrix can be constructed from the element-wise product of the subspace confusion matrices.

Given these assumptions, the cost function becomes

$$c(\gamma, v_1, v_2 \ldots, v_k) = -\sum_{S} \frac{\sigma(s) \prod_{j=1}^{k} F_j(v_j \mid s)}{\sum_{S'} \sigma(s') \prod_{j=1}^{k} F_j(v_j \mid s)} \ln\left(\frac{\sigma(s) \prod_{j=1}^{k} \int_{V\gamma} F_j(v_j \mid s) dv_j}{\sum_{S'} \sigma(s') \prod_{j=1}^{k} \int_{V\gamma} F_j(v_j \mid s) dv_j} \right). \quad (7)$$

The relative differences between the costs, and not the absolute magnitudes, indicate the strength of conviction for the optimal decision. Absolute cost is not a good indicator of conviction strength for a given decision because the costs are logarithmic distances from the decision vectors. A prior vector with a higher probability for a given decision than the matching decision vector can have a higher cost than that of a prior vector that is identical to the decision vector. However, the costs of the other decisions continue to increase at a greater rate than the cost of the optimal decision as the probability associated with the optimal decision increases.

The absolute magnitudes of the costs are still useful in that they indirectly indicate the degree of disagreement between the contributors of the fused decision. Fusion of agreeing contributors will decrease the winning decision cost while conflicting contributors will increase the cost. With sufficient conflict, the optimal decision may be different from those that would be selected by the individual contributors. The optimal decision is a compromise between the contributors in conflict situations. When there is total disagreement among the contributors, all costs are infinite.

The assumption of Equation 5 is directly related to the requirement that data fusion systems not process redundant data. This is algorithmically equivalent to redundantly processing a subspace of the full feature space. A second requirement for data fusion systems is that the contributed data be statistically independent. This is not always accomplished in practice and so reduced performance can be anticipated in those cases.

Equation 6 is the more stressing assumption of the two. It implies not only that the feature-space distribution functions of the decision subsystems should be independent, but that the performance of the decision subsystems, as reflected in the confusion matrices, should be independent. This is unfortunately almost never true in practice. Violation of this assumption will generally result in information loss and performance degradation. This can only be avoided by fusion at the feature level. Feature level fusion, in contrast, is confronted with a problem that has been termed the "curse of dimensionality". The complexity of the feature space increases so rapidly with each additional feature that the multi-dimensional probability density functions cannot be accurately estimated. For the decision function of Equation 1, the minimization of equivocation in large-dimensioned feature spaces becomes another significant challenge in addition to the probability density function estimation problem.

In light of the challenges presented by feature level fusion, the losses from violation of Equations 5 and 6 may often be acceptable. It is always possible to select specific subspaces of the features space for feature level fusion if they violate the independence assumptions too severely. Resources devoted to feature level fusion for such subsystems may be well spent whereas those devoted to subsystems that satisfy the independence assumptions may be poorly spent. Violation of the assumptions in Equations 5 and 6 for decision level fusion and the curse of dimensionality for feature level fusion are what make the decision fusion problem so difficult.

The decision fusion function of Equation 7 assumes that the contributing subsystems are trustworthy (at least to the degree specified in the confusion matrices) and that all probability density functions and the confusion matrices for each subsystem are conservatively and truthfully estimated. Maximum entropy techniques provide one possible method for generating the density functions since the resulting functions should only capture statistically significant details in the training data sets.[3] The extension of the decision theory to include decision fusion problems with untruthful contributors is an intriguing thread that has currently not been followed. The influence of untruthful subsystems on the cost magnitudes suggests one possible approach for detecting untruthful contributors. Another possible approach is to analyze the performance of the subsystems over repeated trials to obtain statistically significant performance

measures that can be used to verify that the actual contributor's performance matches the reported performance reflected in the confusion matrices.

3. Decision and Prior Class Hierarchies

Because the cost function in Equation 7 is fundamentally a distance measure, procedures can be developed to change the levels of abstraction of the prior class and decision set hierarchies by mapping probability-space vectors from a space with one dimensionality to one with another. The elements of the prior-class set are now considered to be independent subsets of a global set of prior classes containing one or more elements. The union of all the subsets must be the global set and the intersection of any two subsets must be the empty set. The same relations hold for the global set of decisions. The assignment

$$P_{\gamma,s} = \int_{v\gamma} F(v \mid s) dv \tag{8}$$

will be adopted to simplify the following notation. For prior class expansion (for example $\{s_2\}$ to $\{s_{2A}, s_{2B}\}$), the components of the prior probabilities, density functions, and confusion matrices are expanded through mappings such as

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}, \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}, \begin{pmatrix} P_{1,1} & P_{1,2} \\ P_{2,1} & P_{2,2} \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma_1 \\ \sigma_{2A} \\ \sigma_{2B} \end{pmatrix}, \begin{pmatrix} F_1 \\ F_2 \\ F_2 \end{pmatrix}, \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,2} \\ P_{2,1} & P_{2,2} & P_{2,2} \end{pmatrix}. \tag{9}$$

If the expanded prior probabilities σ_{2A} and σ_{2B} are known, they are used in the expansion. When they are unknown, a reasonable option is to distribute σ_2 equally between s_{2A} and s_{2B}. The fundamental assumption in the expansion is that the probability density functions for the expanded classes are identical, leading to a basic duplication of terms for the probability density functions and integrals.

Decision expansion (for example $\{\gamma_2\}$ to $\{\gamma_{2A}, \gamma_{2B}\}$) is accomplished with mappings such as

$$\begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,2} \\ P_{2,1} & P_{2,2} & P_{2,2} \end{pmatrix} \Rightarrow \begin{pmatrix} P_{1,1} & P_{1,2} & P_{1,2} \\ \frac{\sigma_{2A} P_{2,1}}{\sigma_2} & \frac{\sigma_{2A} P_{2,2}}{\sigma_2} & \frac{\sigma_{2A} P_{2,2}}{\sigma_2} \\ \frac{\sigma_{2B} P_{2,1}}{\sigma_2} & \frac{\sigma_{2B} P_{2,2}}{\sigma_2} & \frac{\sigma_{2B} P_{2,2}}{\sigma_2} \end{pmatrix}. \tag{10}$$

Only the confusion matrix is modified for decision expansion. The distribution of the rows of the confusion matrix may be scaled by the prior probabilities although the scale factors cancel in the cost function. The scaling is done to keep the sum of the confusion matrix columns equal to one.

Prior class contraction (for example the contraction of $\{s_2, s_3\}$) is accomplished with a mapping such as

$$\begin{pmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\end{pmatrix},\ \begin{pmatrix}F_1\\ F_2\\ F_3\end{pmatrix},\ \begin{pmatrix}P_{1,1} & P_{1,2} & P_{1,3}\\ P_{2,1} & P_{2,2} & P_{2,3}\\ P_{3,1} & P_{3,2} & P_{3,3}\end{pmatrix}\Rightarrow$$

$$\begin{pmatrix}\sigma_1\\ \sigma_2+\sigma_3\end{pmatrix},\ \begin{pmatrix}F_1\\ \dfrac{\sigma_2F_2+\sigma_3F_3}{\sigma_2+\sigma_3}\end{pmatrix},\ \begin{pmatrix}P_{1,1} & \dfrac{\sigma_2P_{1,2}+\sigma_3P_{1,3}}{\sigma_2+\sigma_3}\\ P_{2,1} & \dfrac{\sigma_2P_{2,2}+\sigma_3P_{2,3}}{\sigma_2+\sigma_3}\\ P_{3,1} & \dfrac{\sigma_2P_{3,2}+\sigma_3P_{3,3}}{\sigma_2+\sigma_3}\end{pmatrix}. \qquad (11)$$

The combined density function is a prior-weighted sum of the original density functions.

Decision contraction (for example the contraction of $\{\gamma_2,\gamma_3\}$) is a mapping that again uses prior-weighted sums, such as

$$\begin{pmatrix}P_{1,1} & \dfrac{\sigma_2P_{1,2}+\sigma_3P_{1,3}}{\sigma_2+\sigma_3}\\ P_{2,1} & \dfrac{\sigma_2P_{2,2}+\sigma_3P_{2,3}}{\sigma_2+\sigma_3}\\ P_{3,1} & \dfrac{\sigma_2P_{3,2}+\sigma_3P_{3,3}}{\sigma_2+\sigma_3}\end{pmatrix}\Rightarrow\begin{pmatrix}P_{1,1} & \dfrac{\sigma_2P_{1,2}+\sigma_3P_{1,3}}{\sigma_2+\sigma_3}\\ P_{2,1}+P_{3,1} & \dfrac{\sigma_2P_{2,2}+\sigma_3P_{2,3}+\sigma_2P_{3,2}+\sigma_3P_{3,3}}{\sigma_2+\sigma_3}\end{pmatrix}. \qquad (12)$$

A primary benefit obtained from the expansion rules is that the expansion of an optimal decision subset to multiple decision subsets results in the expanded decision subsets being equally optimal in terms of cost. Expansion by traditional Bayesian methods usually reduces the probabilities assigned to the prior classes to the point that an unexpanded class may be selected as the optimal decision. This pitfall is avoided because the characteristic decision vectors undergo a dilution comparable to that of the measurement based probabilities. The contraction process avoids the same problems to a lesser degree.

It should be noted that contraction and expansion are not inverse operations. Expansion followed by contraction will result in the original parameters, but contraction followed by expansion will generally result in different matrix and vector elements. This is because information is lost during the contraction operation that cannot be restored through the expansion operation. The information lost during contraction leads to the possibility that the optimal decision after re-expansion may not correspond to the optimal decision prior to expansion.

With the ability to change the hierarchical levels of the prior classes and the decisions, the next natural step is to contemplate the existence of unknown elements in the prior class set and decision set. A fully degenerate global prior-class set or a decision set consists of a single subset containing all the global set elements. Expansion of the single subset to more descriptive levels creates additional subsets that provide greater detail and focuses the decision system on the elements of interest. The expansion process can be assumed to always contain a subset with a collection of elements that consists of "everything else". This set can also be considered to be a subset of "unknowns". A subset of "unknowns" provides a means to account for uncertainty. Dempster-Shafer evidential reasoning accounts for uncertainty through the

power set Θ (the set of all sets).[2] A significant distinction between Dempster-Shafer's Θ and our unknown subset is that Θ contains "everything" and our unknown subset contains "everything else." This distinction is due to the requirement that our subsets contain no common elements whereas the subsets in Dempster-Shafer evidential reasoning are allowed to contain common elements.

4. Recursive Decision Algorithms

Recasting Equation 7 as a recursive algorithm proves to be instructive. A decision fusion process for time series information from a single sensor is used as the model to develop the recursive algorithm. The resulting recursive algorithm has traits that are shared with the Kalman filter algorithm. As with the Kalman filter algorithm, there is an initial state estimate,

$$\sigma_{0,0}(s) = \begin{pmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(n) \end{pmatrix}, \quad B_{0,0}(s \mid \gamma) = \begin{pmatrix} & \sigma(2) & \cdots & \sigma(n) \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}. \tag{13}$$

The initial probability state estimate $\sigma_{0,0}(s)$, and belief matrix $B_{0,0}(s \mid \gamma)$ are formed from the prior probabilities. The belief matrix is somewhat synonymous with the covariance matrix of the Kalman filter. The state is propagated to the next time step, which, for now, is an identity operation,

$$\sigma_{t,t-1}(s) = \sigma_{t-1,t-1}(s), \tag{14}$$

$$B_{t,t-1}(s \mid \gamma) = B_{t-1,t-1}(s \mid \gamma). \tag{15}$$

Next, the prior probabilities and belief matrices are updated with new data,

$$\sigma_{t,t}(s) = \frac{\sigma_{t,t-1}(s) F(v_t \mid s)}{\sum_S \sigma_{t,t-1}(s) F(v_t \mid s)}, \tag{16}$$

$$B_{t,t}(s \mid \gamma) = \frac{B_{t,t-1}(s \mid \gamma) P_t(\gamma \mid s)}{\sum_S B_{t,t-1}(s \mid \gamma) P_t(\gamma \mid s)}. \tag{17}$$

A decision can then be selected at this point in the cycle through the use of the cost function,

$$c_t(\gamma, v_t) = \sum_s \left[-\ln\left(B_{t,t}(s \mid \gamma)\right) \right] \sigma_{t,t}(s). \tag{18}$$

The recursive algorithm returns to Equation 14 to begin the next time step. Examination of Equations 13 through 18 shows that the recursive algorithm is identical to the fusion function in Equation 7.

With a basic recursive function, extensions can be considered. The first extension that can be contemplated is to change the initial state values in Equation 13. An important characteristic of the decision fusion function of Equation 7 is that an identity operator exists. Fusion with the identity operator does not modify the resulting costs. A simple interpretation in terms of decision fusion is that the identity

operator represents a maximally indifferent expert, who is always incapable of making a decision and thus does not influence the resulting decision. The identity operator is

$$\sigma_I(s) = \frac{1}{N_S}, \quad P_I(s \mid \gamma) = \frac{1}{N_S} \tag{19}$$

for all prior classes and decisions, where N_S is the number of elements in the set of prior classes. The initial state in the recursive algorithm can be selected to be the maximally indifferent state,

$$\sigma_{0,0}(s) = \frac{1}{N_S}, \quad B_{0,0}(s \mid \gamma) = \begin{pmatrix} \frac{1}{N_S} & \frac{1}{N_S} & \cdots & \frac{1}{N_S} \\ \frac{1}{N_S} & \frac{1}{N_S} & \cdots & \frac{1}{N_S} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N_S} & \frac{1}{N_S} & \cdots & \frac{1}{N_S} \end{pmatrix}, \tag{20}$$

The resulting decisions that are obtained as the algorithm executes arise from the information accumulated in the probability state estimates and the belief matrix. The prior probabilities in Equation 7 can be given a strict interpretation as being obtained entirely from previously accumulated evidence. This interpretation leads to the next extension. If $\sigma(s)$ arises from accumulated prior information, then the same $\sigma(s)$ terms in the initial belief matrix $B_{0,0}(s \mid \gamma)$ of Equation 13 are not necessarily correct and should be replaced with an accumulated belief matrix. With the assumption that initial, non-maximally indifferent probability state estimates are due to accumulated information, then the decision fusion cost function of Equation 7 should be amended to allow for an accumulated belief matrix, such as would result from the repeated application of Equation 17. Such a modified system could make decisions before any actual information is processed. Decision systems can thus be created that are biased to a particular decision and must accumulate sufficient evidence in support of an alternate decision before that alternate decision can be selected.

The next extension to the recursive algorithm is not nearly so obvious, given the limited discussion on information theory and the construction of the confusion matrices. The confusion matrix $P_t(\gamma \mid s)$ is obtained by minimizing the equivocation (information loss) between the prior classes and the decisions. The equivocation minimization is influenced by the values of the prior probabilities. This minimization is achieved through the optimal assignment of decision regions throughout feature space. Given this, the natural extension to the recursive algorithm is to define new feature-space decision regions so as to minimize equivocation at each recursive cycle. This minimization would occur between Equations 15 and 16. Equivocation for the recursive algorithm is

$$H(S \mid \Gamma) = -\sum_{S}\sum_{\Gamma}\left[\int_{V} \sigma_{t,t-1}(s) F_t(v \mid s)\delta_t(\gamma \mid v)dv \times \right.$$
$$\left. \ln\left(\frac{B_{t,t-1}(s \mid \gamma) \int_{V} F_t(v \mid s)\delta_t(\gamma \mid v)dv}{\sum_{S'} B_{t,t-1}(s' \mid \gamma) \int_{V} F_t(v \mid s')\delta_t(\gamma \mid v)dv} \right)\right], \qquad (21)$$

where $\delta_t(\gamma \mid v)$ is the function that maps each element of feature space to a decision. Only the decision function may be modified to minimized information loss since all other functions are assumed fixed. Assuming that other functions can be modified breaks the recursive nature of the system.

It is doubtful that recursive equivocation minimization will find much use in real-time applications due to the difficult and time consuming nature of the minimization. It does however maximize the additional information that can be accumulated at each cycle of the recursive algorithm without abandoning the algorithm's recursive nature. It also demonstrates the non-linear character of the generic decision problem and why it is so difficult to design optimal decision systems of any reasonable complexity.

The last extension to be considered involves Equations 14 and 15. Given the previous remarks on the recursive algorithm's relationship to Kalman filters, it is natural to replace the state propagation equations with an operation other than an identity operation. A natural choice is a Bayes-Markov process that modifies the prior probabilities and belief matrices at each time increment. A common response that has been encountered with this proposal, viewed in terms of identification systems, is "Why would one ever wish to do this?" In the view of more general decision systems, there are a number of reasons that are immediately obvious. For decision systems where states may change with time, it is a requirement. An example of such a decision system would be one that is not only required to identify targets, but decide on the intentions of the targets. Clearly the intentions will change with time and can be modeled with a Bayes-Markov process. Additionally, decision systems could be designed that are predisposed to evolve to a given decision unless sufficiently convincing information is received to force a different decision. When the supporting information is no longer received, the system will eventually return to a default decision.

A third possible use of a Bayes-Markov process is in cases where the incoming information is correlated between cycles. It may be possible to account for this correlation by reducing the information content of the accumulated state and belief matrix or the probability densities and confusion matrices that are updating the state. Degrading the information in the accumulated state and belief matrix would give a decision system that avoids becoming locked into a decision.

Research is currently underway to identify the specific processes that will generate desirable time dependent behaviors. For example, a desirable feature of information decay functions is that if no new information is received during a cycle, or maximally indifferent information is received, the resulting decision does not change. It remains to be determined if the vectors of the belief matrix and the prior probabilities can both use the same decay function and satisfy this requirement. It may be necessary to use different functions for the belief matrix vectors and the prior probabilities, in which case the functions should be related through some guiding principle. In addition, a correspondence between the level of correlation between prior accumulated data and the update information and the de-correlation functions remains to be identified.

Given the recursive algorithm and its extensions, an information theoretic interpretation of the process can be made. The self-information $I(S)$ of the prior classes is

$$I(S) = -\sum_{s} \sigma(s) \ln \sigma(s) . \quad (22)$$

This is the information that an observer believes can be extracted from a system through measurement. The choice of the distribution in Equation 20 is a logical choice for the initial state since this is the state with maximum self-information. Measurements are conducted to extract this information from the system. Repeated, confirming measurements will increase the probability of one prior class and decrease the rest, reducing the self-information that the observer believes to remain for further extraction. The limit to the recursive process has a single element of S with a probability of one and the remaining with zero probabilities. The self-information in this limit is zero; the observer believes that there is no further information to extract from the system. Therefore, additional measurements have no effect on the prior probabilities or the optimal decision. Conflicting measurements have the opposite effect in comparison to confirming measurements, and cause the probabilities among the prior classes to converge to common values and the self-information to increase. The observer then believes that the self-information of the system was previously underestimated and that more information remains to be extracted.

The belief matrix also has information theoretic interpretations. The decision vectors comprising the rows of the belief matrix represent the average distribution of prior probabilities for each decision. This distribution captures the average amount of information remaining in a system after each time step, assuming that the series of decisions are all confirming decisions. The decision vectors of the belief matrix evolve with each increment of the recursive process and model the evolving averages of the information remaining to be extracted for each decision. In most decision systems, the decision vectors will approach the limits discussed for repeated, confirming measurements.

5. Decision System Design Options

The decision theory provides for a multitude of options for designing decision fusion systems. Regardless of the design, the theory requires that the feature subspaces be statistically independent. In addition, the decisions that result from the decision fusion function should reflect those decisions that would result from fusion at the feature level. When this second requirement is sufficiently violated such that the fusion system does not meet its performance specifications, fusion must be pursued at the feature level. In addition, the prior class hierarchy must be common throughout the fusion system. As will be seen, the same requirement in not necessarily imposed on the decision set hierarchy.

The first significant design decision is whether to design a centralized or decentralized fusion system. For centralized systems, the subsystems share a common mission objective that is determined by a control center. As for all decision fusion systems, the prior class hierarchy must be common throughout the system. The centralized design assumes a common decision hierarchy as well. If the subsystems minimize equivocation in real-time, the center must report the latest available prior probabilities to the subsystems. The subsystems then report updated confusion matrices along with each probability density vector. If no real-time minimization occurs, the confusion matrices can be reported at subsystem startup or maintained in a database at the control center. If communications bandwidth is a problem, the probability density vectors reported to the control center may be reduced to a few significant prior class probabilities that the center uses to reconstruct an approximation to the original vector. If communications

bandwidth limits are severe enough, a single prior class enumeration may instead be sent to the control center and used to select a characteristic probability density vector from a database.

For decentralized systems, there are more options to consider because the members of the system may be pursuing different missions. If the mission is common to the system, such as target detection and identification, then most design decisions are similar to those of the centralized system. The decentralized system requires more communications bandwidth than the centralized system because no control center is available to coordinate the network. As a subsystem joins the network, it may need to request priors and belief matrices from the active subsystems to bootstrap its internal decision system. The joining subsystem can select the best set of priors that it receives from the responding subsystems to complete its bootstrap. As each subsystem processes measurements, it broadcasts its probability density vectors and confusion matrices to the other subsystems in the network while receiving the same kind of information from the others. Each subsystem independently fuses its accumulated information and makes independent (and hopefully consistent) decisions.

Some of the theoretical underpinnings of the decision fusion theory may have to be abandoned to design decentralized systems with members that are pursuing independent missions. An example of such a system might be one with subsystems that make internal resource allocation decisions. Subsystems with unique missions imply that the subsystems are making different kinds of decisions. Decisions that do not share a common set hierarchy and therefore lead to incompatible belief and confusion matrices cannot be fused by the current theory. In this kind of system, the subsystems' confusion matrices may not be determined by minimizing information loss. In a decentralized system of this magnitude, only prior probabilities and probability density vectors can be exchanged between the members in the system. The set of decisions in each subsystem will be organized toward completing their unique missions. The control center may still appear as an element of a decentralized decision system, but with the role of mission coordination instead of decision coordination. The members of the network would independently select optimal decisions to satisfy their unique mission objectives.

For decision fusion systems of either type, recursive algorithms may be implemented in two principal ways. In the first, the subsystems report sequentially independent data that the receivers accumulate with a recursive algorithm. In the second, the subsystems accumulate data with a recursive algorithm and report the accumulated data. Here, the receiver incorporates the new data from an accumulating subsystem into the fusion system after it discards that subsystem's previous data. The recursive algorithm extensions lead to a wide range of options for system designs. Additional study is still required to determine best ways to implement the recursive algorithm and its possible extensions to meet the various needs of data fusion systems.

6. Summary

Continued study of a decision fusion theory, constructed from statistical decision theory and information theory, has revealed a number of desirable characteristics inherent in the theory. The theory allows for the prior classes and the decisions to be two distinct sets, without necessarily a one-to-one correspondence between them. The information theoretical connection to the theory is weakened without the one-to-one correspondence, but a more general decision theory can be developed. The decisions of the theory are represented as characteristic partitions of the prior classes. Selection of an optimal decision is accomplished by selecting the decision with a characteristic partition that most closely matches the probabilistic partitioning indicated by the measurements, through the probability density functions of the prior classes.

For decision systems composed of independent subsystems, a decision fusion function is obtained. The fusion function allows for the hierarchical levels of the sets of prior classes and decisions to be independently altered to suit the decision system requirements. Decision subsystems can be designed to operate at their natural level of abstraction in the prior class and decision set hierarchies. Decision fusion can then be accomplished after the subsystem data are transformed to the appropriate levels in the set hierarchies.

The decision fusion function has been recast as a recursive algorithm. The recursive algorithm provides information theoretical insights into the interpretation of the role of prior probabilities. Prior probabilities are simply the previous evidence that has been accrued in the recursive decision process. The recursive algorithm shows that the belief matrices also may accrue. Accrued belief matrices permit the design of biased decision systems. Biases are not to be considered as a negative characteristic in this type of application, but as a means to encapsulate previous information or to design the fusion system to meet performance specifications.

Work continues on the original motivator of this theoretical study, the development of an identification fusion system for Kwajalein Missile Range (KMR). This system will combine metric, beacon, and signature information from multiple radars and optical sensors, as well as from human operators, to create a fused picture of ballistic missile complexes. The fused picture will provide metric data and identity estimates for the objects in the complex. This information will be used to aid the sensors in satisfying their data collection requirements. Within the year, the KMR identification fusion system should be implemented as a real-time program and my collaborators and I will begin to obtain results for the system in an operational environment. A second paper, co-authored with Michael Seibert, is being presented at this conference and provides an informative overview of this fusion system.

A search has begun to identify other applications that might benefit from the recently developed decision fusion theory. Discussions have begun with researchers who are evaluating different discrimination systems for use in ballistic missile defense applications, as well as with experts evaluating combat ID systems. I hope to evaluate the decision fusion theory against other techniques in head-to-head tests in the near future.

Theoretical studies continue, in order to gain a better understanding of the recursive fusion algorithm and how to best implement more advanced recursive algorithms. Areas under investigation include Markov processes, exponential families, and control theory. Additional areas of study that relate to the decision fusion theory hold intense interest. One area relates to the probability density functions in the theory, which are fixed functions in the decision theory. An examination of techniques for the generation of probability density functions is on the list of topics to examine. Maximum entropy techniques, learning algorithms, and neural networks are possible fields that might prove fruitful. The theoretical analysis of decision fusion systems that do not have fully trustworthy contributors is another possible area of study. Theoretical results in this area could influence traditional intelligence gathering activities. It is possible that techniques could be developed to evaluate the reliability of information sources and identify sources that are supplying misleading information. Through the course of these future studies, I hope to gain a deeper understanding of the theory and how it relates to other theoretical efforts in decision fusion.

---

[1] Hurley, Michael B., "Information Fusion with Constrained Equivocation," Proc. 1998 IRIS National Symposium on Sensor and Data Fusion, ERIM, Ann Arbor, 1998, pp. 159-178.

[2] Bogler, Philip L., "Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-17, no. 6, pp. 968-977, Nov/Dec 1987.

[3] Bretthorst, G. L., "An Introduction to Model Selection using Probability Theory as Logic," in *Maximum Entropy and Bayesian Methods*, Glenn R. Heidbreder (ed.), Dordrecht, Kluwer Academic Publishers, 1996, pp. 1-42.