Approaches to Multisensor Data Fusion

Joseph S. J. Peri

As part of an Office of Naval Research–funded science and technology development task, APL is developing an identification (ID) sensor data fusion testbed. The testbed is driven by an APL-modified version of the Joint Composite Tracking Network pilot benchmark called the Composite Combat ID Analysis Testbed (CAT). The CAT provides accurate tracking for realistic scenarios involving multiple targets and netted radar and ID sensors. Track state outputs from the CAT include feature information from electronic support measures and noncooperating target recognition sensors. These data are combined to improve the confidence of aircraft-type declarations using both Bayesian and evidential reasoning–based algorithms. We use measure theoretic methods to describe the relationship between Bayesian theory and the Dempster-Shafer evidential reasoning theory.

BACKGROUND

Positive identification (ID) of targets as friend, foe, or neutral is recognized as a major operational problem for military forces. Friendly ID is assisted by various resources, such as identification, friend or foe (IFF) Mode 4 or operational procedures such as return to force corridors, air tasking orders, or unencrypted IFF modes. These techniques are clearly inadequate for neutral platforms, hostile platforms using deceptive tactics, or friendly aircraft operating with damaged equipment or forced to operate outside normal or agreed upon operating procedures. To support the development of more robust ID capabilities, the Office of Naval Research is funding a science and technology task, Composite Combat Identification (CCID), to develop improvements in techniques and algorithms.

The goal of the CCID project is to develop a method for networking multiple sensors and integrating observational data to provide target classification and ID. Target classification addresses the issue of the type of target observed; for example, is the target a MiG, cruise missile, or airbus? Target ID addresses the issue of deciding whether the target is friendly, hostile, or neutral. Positive target ID supports the engagement decision in ship self-defense. The project will demonstrate the CCID concept by using the existing Cooperative Engagement Capability (CEC) to net various sensors and to provide composite tracks together with composite ID.

As part of this task, APL is teamed with several other organizations, including the Naval Research Laboratory (NRL) and General Dynamics Electronics Systems, to demonstrate the effectiveness of integrating electronic warfare and other ID sensors and sources into a high-capacity, real-time, netted sensor system.

The demonstration will initially be performed using non-real-time modeling and simulation. This will be followed by real-time demonstrations driven by high-fidelity models of sensors. Finally, a real-time demonstration using live sensors and sources will be performed in the field.

The problem of target ID must address real-world issues. For example, Fig. 1 dramatizes a possible setting in a littoral environment in which there could be several non-co-located Navy platforms, each with its local sensor picture of the real world. The sensors could be radar, electronic support measures (ESMs), or noncooperative target recognition (NCTR). Netting these platforms, as in a CEC setting, would allow the sharing of information and the creation of both composite tracks and composite ID. Thus, each Navy platform benefits from the shared information in case physical factors (e.g., terrain obscuration) should cause loss of local tracks and local ID.

Multiple targets would normally be present in a real scenario, the nature of the targets being neutral, friendly, or hostile. Composite ID would help greatly to resolve targets. The problem of unresolved targets addresses the issue of multiple targets in close proximity within the uncertainty region of the sensor or multiple tracks that cross. After such track interactions, one must solve the problem of which track goes with which target.

This article discusses the CCID status, technical demonstrations, and conceptual approaches. Highlights of mathematical machinery used in the target classifier's reasoning engine are also addressed.

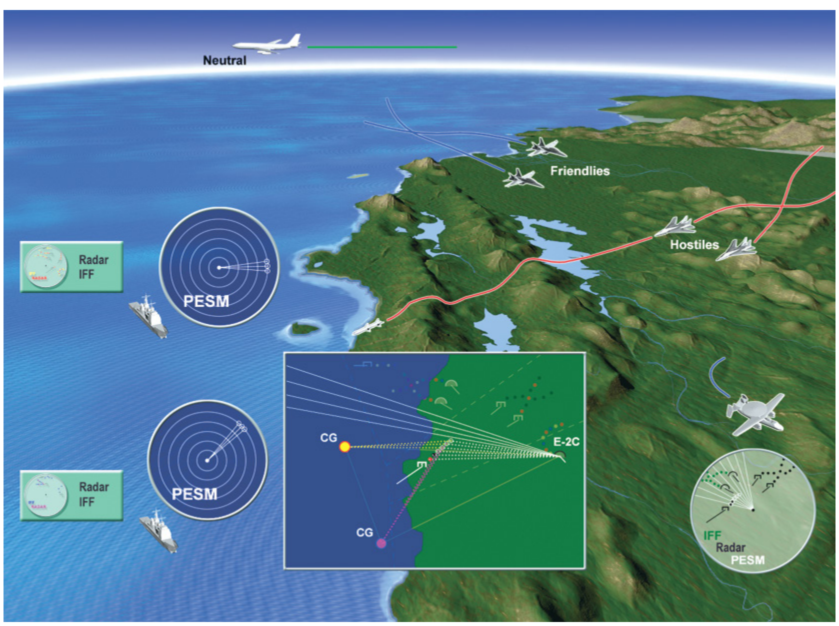

Figure 1. A possible scenario in which netted sensors and target ID could improve the warfighter's decision criteria. Netted ID sensors provide integrated passive electronic support measures (PESM), improved classification, and an enhanced air picture via passive tracking. Limitations include a restricted set of sensors, limited noncooperative target recognition capability, and aspect angle/terrain blockage.

APL'S CCID EFFORTS

The Laboratory's CCID work has focused on the creation of a high-fidelity model of netted sensors. As a starting point, the development team used the existing Joint Composite Tracking Network (JCTN) pilot benchmark. Developed for the Ballistic Missile Defense Office, the pilot benchmark served as the basis for the CCID Analysis Testbed (CAT). Certain modules of the pilot benchmark were modified and new modules were added for the development of the testbed.

Figure 2 shows the basic files and processing modules of the pilot benchmark, and highlights those areas that were modified or added to support the CAT. Pilot benchmark code is written in MATLAB and is relatively easy to modify and use.

FY2000 Demonstration (00 DEMO)

The CCID is presently in its second year of development. The first year's efforts consisted of the development of the CAT as well as in-depth analysis of the various mathematical methods used in the reasoning engine. FY2000 culminated in the 00 DEMO, highlighting the infrastructure of the CAT with key components such as ESM interface modules; ESM-to-ESM and ESM-to-radar track correlation logic; evidential reasoning classifier; and data extraction, reduction, and display capabilities.

The 00 DEMO scenario was a derivative of the Northeast Asian Design Reference Mission, with emphasis on air targets and the netted sensor platforms. The network consisted of two cruisers, one aircraft carrier, and an E-2C surveillance aircraft. Figure 3 is a plot of track state update messages for a 20-min period of this 6-h scenario.

The infrastructure demonstration portion of 00 DEMO revealed the system's ability to form high-quality tracks using the line-of-sight measurements of the ESM sets before the radars acquired the tracks. The system successfully correlated many ESM reports to already existing composite radar tracks. Although we used simulated ESM electronic notation data, the system demonstrated its ability to combine various pieces of evidence to classify and identify the composite tracks.

FY2001 Demonstration (01 DEMO)

The success of the 00 DEMO served as the launching point for the 01 DEMO planned for the end of

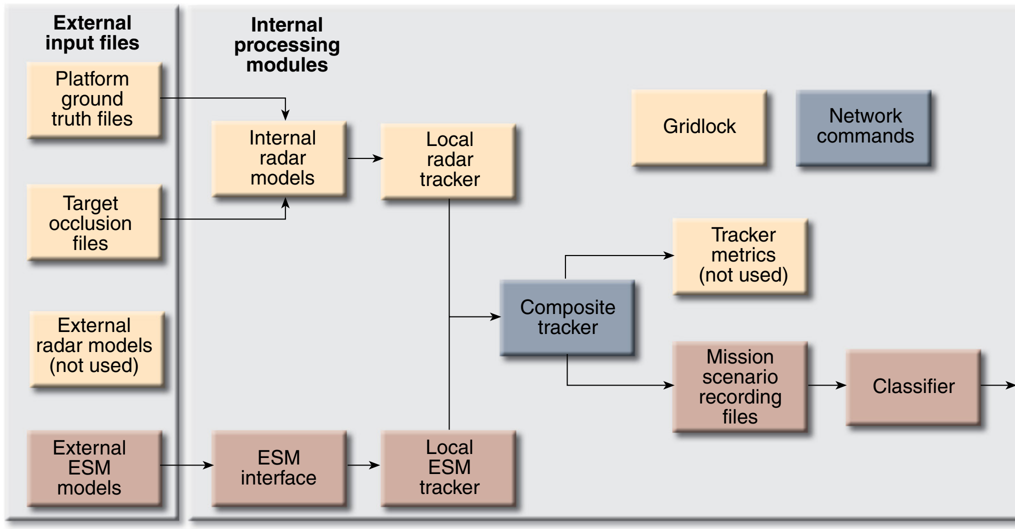

Figure 2. Joint Composite Tracking Network pilot benchmark with CAT modifications (blue = modified, cream = existing, brown = new).

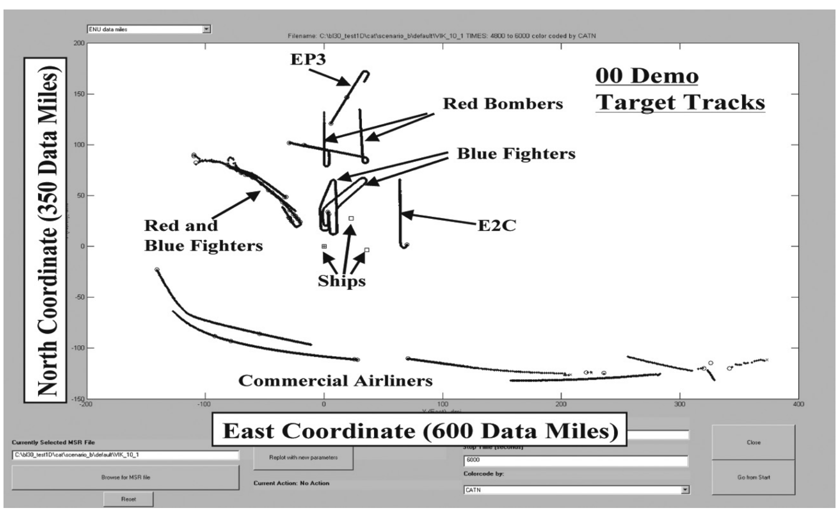

Figure 3. Target tracks from the 00 DEMO showing 20 min of data.

the 2001 time frame. The 01 DEMO will include (1) the implementation of a multi-hypothesis tracker (MHT) process within the CAT for improved ESM contact to composite track data association, (2) relative performance comparison of the two classifier reasoning approaches, i.e., Dempster-Shafer theory and Bayesian theory, and (3) implementation of "real-time" ID processing driven by outputs from the CAT, including operator-entered doctrine, use of geographic and air route information, etc. The ID improvements will not be discussed further in this article.

The MHT process is based on a structure-branching approach to hypothesis updating. In collaboration with NRL, the developers of the MHT tracker, APL has been integrating the MHT algorithm within the CAT to address the issue of associating ESM contact data with composite track data. The 01 DEMO implementation will use all remote radar and ESM reports in its hypotheses formulations. Future implementations of MHT functionality will be expanded to include all aspects of track formation and maintenance.

DATA FUSION APPROACHES/ISSUES, THEORETICAL ASPECTS OF REASONING ENGINES

To achieve target classification, a reasoning engine must be used to combine the various pieces of evidence (i.e., information from multiple netted sensors) and to produce the target classification and ID. Figure 4 contains a simple image of our basic problem. For example, the ESM detects radiation from an active radar on an airborne platform. It analyzes the attributes of the detected radiation, i.e., frequency, pulse width, pulse repetition interval, etc.; compares the attributes to those in its library; and outputs a list of interpretations for the detected evidence. The list of interpretations takes the form of a list of possible emitters, together with relative probabilities, that could have produced the physical evidence. Similarly, for any other netted sensor (e.g., an NCTR sensor such as an electro-optic imaging system or high-resolution radar), the reasoning engine combines all the interpretations of the physical evidence to provide platform classification and ID.

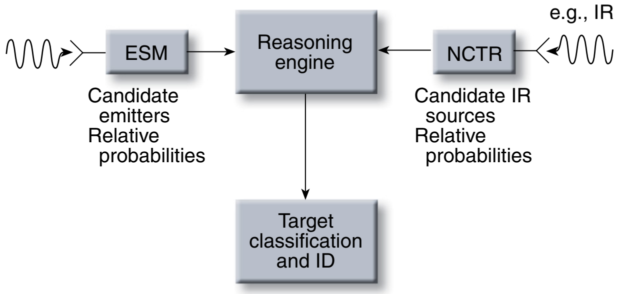

Figure 4. Basic problem.

One can consider many mathematical techniques to address the CCID problem.[1,2] However, based on the recommendations of the Multi-Source Integration System Engineering Team,[3] the Dempster-Shafer fusion approach is well-suited for handling partial probability attributes, while the Bayesian fusion approach is well-suited for handling complete probability attributes. An example of an operational Dempster-Shafer ID system is the Canadian Navy ESM Warning System, Version 2 (CANEWS 2).[4]

The decision to use one approach versus another depends on many factors such as the characteristics of targets, sensors, and systems. One must also consider how to handle various kinds of information (e.g., probabilistic, nonprobabilistic). Finally, the output of the reasoning engine, such as posterior probability or evidential interval, must be interpreted properly. Thus, the emphasis of CCID is to achieve a robust ID system while investigating the basics of system performance, algorithms, network limitations, and mathematical approaches.

There are apparent differences between the Dempster-Shafer theory and the Bayesian theory that have led to misunderstandings of the various terms used in them. For example, suppose evidence indicates a set of aircraft types without pointing to any particular member of the set. The Bayesian representation would consider a uniform distribution of the weight of evidence over all members of the set. However, in the Dempster-Shafer representation the weight of evidence would be attributed to the set itself, and not to any particular member. Not surprisingly, the two approaches would give results differing in their fundamental structures. One can *mistakenly* conclude that the Dempster-Shafer theory is *ad hoc*, competing against the well-established Bayesian theory, and that because its implementation can be computationally intensive, one should therefore not use it.

Bayesian theory is fundamentally classical probability theory. Although it has evolved for over 250 years, not all problems have been solved. The very concept of independence gives probability theory its unique characteristics and familiar language. One of the highlights is Bayes' theorem, which relates posterior probability to its prior probability,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} ,$$

where, for $i = 1, 2, \ldots, n$, A_i and B are random events, and $P(B|A_i)$ is the conditional probability of event B given that event A_i has occurred. A similar interpretation goes with the posterior probability $P(A_i|B)$. The conditional probabilities are used to model the stochastic nature of sensor response given some target. In a recursive manner, the posterior probabilities are then used as priors for the next cycle. Kolmogorov[5] shows how one can develop the theory of probability axiomatically, and therefore place it on a solid mathematical foundation.

The Dempster-Shafer theory introduces new concepts such as basic probability assignments, belief, plausibility, multivalued functions, upper and lower probabilities, and Dempster's combination rule. There has been much discussion about this theory, largely stemming from confusion about its terms, their origin, and how they are related. Dempster's original paper[6] introduced upper and lower probabilities, multivalued mappings, and a rule for combining multiple sources of information (see Appendix A, Dempster's Construction). Shafer's approach postulates the existence of a belief function satisfying some axiomatic conditions.[7] He also introduces the notions of the basic probability assignment and plausibility (see Appendix B, Shafer's Approach). The most recurring questions about the theory are: What do belief and plausibility mean? From where did the Dempster combination rule come? Under what conditions is it applicable? All these questions can be answered in the context of measure theory. Dempster, Shafer, and others[8,9] are aware of the measure theoretic connections.

Measure theory is a branch of mathematics that is intimately related to integration theory. One can construct a theory of measure and then develop integration based on measure. Measure is a generalization of familiar concepts such as the length of a line segment, the area of a plane figure, or the integral of a non-negative function over a region of space. Appendix C, Elements of Measure Theory, gives some highlights of measure theory relevant to demonstrating the close relationship between evidential reasoning and Bayesian theory.

In a Bayesian setting, one begins with a set X of elements called elementary events, a *σ-algebra* $\boldsymbol{S}$ of subsets of X called random events, and a probability function p that maps elements of $\boldsymbol{S}$ to the unit interval $[0, 1]$. For the CCID problem, X is the set of emitters output by the ESM, with relative probabilities, p, of producing the detected physical evidence. The *σ-algebra* $\boldsymbol{S}$ represents a collection of subsets of X such that the members of $\boldsymbol{S}$ are sets of emitters and their set complements; and for any two sets of emitters in $\boldsymbol{S}$, $\boldsymbol{S}$ also contains their union. Thus, the mathematical structure permits addressing questions about the occurrence probability of emitters and of classes of emitters. The system consisting of the three items X, $\boldsymbol{S}$, and p—denoted (X, $\boldsymbol{S}$, p), and called a probability space—satisfies the axioms of classical probability. The structure of $\boldsymbol{S}$ depends on the nature of the experiment. When the experiment takes place, one of the elementary events x_0 is realized. If x_0 is an element of the random event A, then we say that the random event A has occurred as a result of the experiment. So far, the material is the familiar classical probability theory.

What happens when a function, say Γ, maps the elementary events of X to another set, say Φ, whose elements are actually subsets of some set Θ? Then Γ and p induce a probability structure on Φ, say (Φ, $\boldsymbol{T}$, μ), that satisfies all the axioms of probability. $\boldsymbol{T}$ is a σ-*algebra* of subsets of Φ, and μ is the probability function that maps the elements of $\boldsymbol{T}$ to the unit interval. These concepts are at the foundation of the Dempster-Shafer theory. We apply these mathematical constructs to the target ID problem as follows. When the classifier consults the platform database to determine a relationship between the list of emitters and the sets of possible platforms that carry such emitters, the mapping Γ is established between X, the set of emitters, and a set Φ of subsets of Θ, the collection of platforms. Thus, the elements of Φ, which are subsets of Θ, behave like elementary events, and the elements of $\boldsymbol{T}$ behave like random events. Because the elements of Φ are subsets of Θ, Γ is called a random set.[10] Hence, μ is the (generalized) probability law of the random set Γ. The Dempster-Shafer theory is precisely a theory of random sets, and the CCID classification problem is perfectly described by this theory.

In summary, the ESM detects radiation and claims that the evidence could have come from several emitter types, accompanied by relative probabilities. This situation is described by the probability space (X, $\boldsymbol{S}$, p). The classifier checks the platform database to see which platforms are associated with each emitter type. This association is described by the mapping Γ. A natural ambiguity can easily arise in that each emitter type may be associated with multiple platforms. The ambiguity is described by the fact that each element of X, i.e., each emitter type, is mapped to a subset of Θ, i.e., is associated to multiple platforms. The set Φ is that collection of elements, each representing multiple platforms, with which the emitters are associated. $\boldsymbol{T}$ is the collection of subsets of Φ on which probability μ is constructed, and μ is the probability function related to Γ and p. We therefore have the induced probability space (Φ, $\boldsymbol{T}$, μ).

COMPARISON OF CLASSIFIER APPROACHES

Preliminary comparison of the classifier approaches involves the need for fair scoring methods and proper interpretation of the classifier outputs. The comparison is modeled after the example given by Blackman and Popoli.[11] Figure 5 shows the feature space for this example, which consists of four targets, each capable of operating in one of two equally likely modes: Mode A and Mode B. For each target, the modes are associated with a dynamic feature. For example, if the mode represents radiation frequency, then the dynamic feature is the frequency band associated with that mode. Various regions of the dynamic feature arise because all targets are considered potential sources of the detected evidence. If Mode A evidence is detected in Region IV, a natural ambiguity arises in that the evidence may have come from Target 2 or Target 3.

Blackman and Popoli describe two sets of probability density functions (PDFs) for these modes. Both sets of PDFs are described in Fig. 6, the first as mixed uniform and the second as mixed normal. The mixed normal distributions are truncated Gaussians with the same means and limits as those of the mixed uniform cases, and with the standard deviations one-sixth those of the corresponding mixed uniform distributions.

The performance comparison for the two classifier approaches consisted of Monte Carlo runs for four cases:

1. Target truth based on mixed uniform and classifier tuned to mixed uniform (u-u)
2. Target truth based on mixed normal and classifier tuned to mixed uniform (n-u)
3. Target truth based on mixed uniform and classifier tuned to mixed normal (u-n)
4. Target truth based on mixed normal and classifier tuned to mixed normal (n-n)

Figure 7 contains a plot of the probability of correct classification versus independent sample number for the Bayesian classifier. The horizontal axis spans 20 samples, but the arrow indicates a more realistic situation consisting of 5 samples.

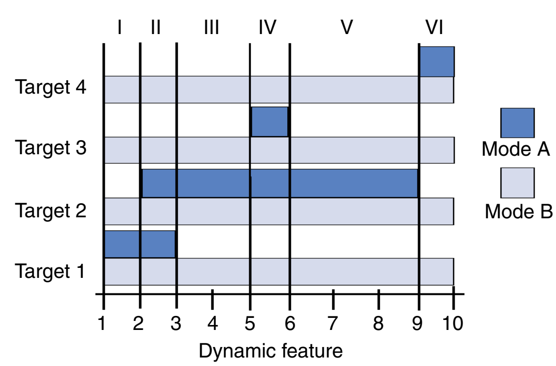

Figure 5. Target modes and features intervals. Each mode is assumed to be equally likely.

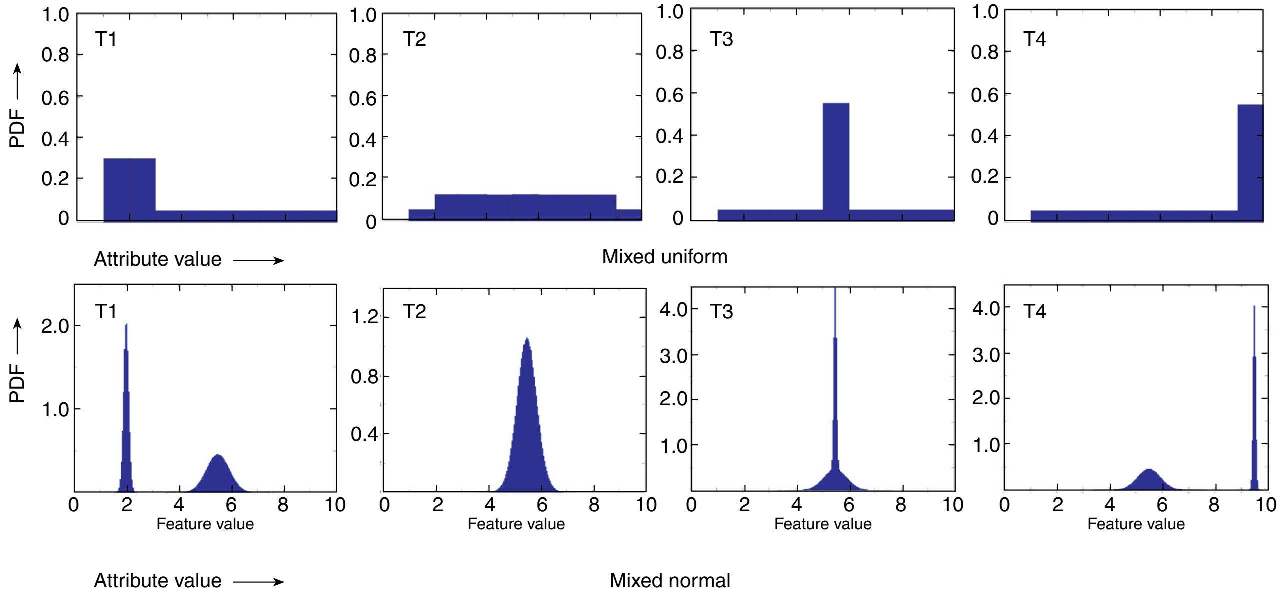

Figure 6. Mixed uniform and mixed normal distributions used by Blackman and Popoli.[11] The authors used these same distributions for initial testing of Bayesian and evidential reasoning classifier algorithms. T1 refers to Target 1, T2 to Target 2, etc.

Figure 7 shows good performance for the Bayesian classifier properly tuned to the target distributions. However, the plot indicates the performance degradation that could occur with mismatches. Figure 8 contains a plot of the probability of correct classification versus independent sample number for the evidential reasoning classifier. It was generated using the basic probability assignments indicated in Ref. 11. In general, the Bayesian results are better except in the (u-n) case, in which the data were generated using mixed uniform distributions but the classifier was tuned to expect mixed normal distributions.

There are many ways to define regions in the target feature space. Each of the region definitions results in different mass assignments. Figure 9 shows the performance changes in the evidential classifier that result with a variant set of masses differing from those of the Blackman and Popoli example. It contains only the (u-u) and (n-u) cases for both the Bayesian and the evidential reasoning classifiers. The matching cases indicate that both classifiers perform identically, on the average. However, the mismatched cases show that the evidential classifier slightly outperforms the Bayesian classifier. These analyses give a flavor of the many ways that masses can be assigned and reasoning approaches compared.

J. Altoft[4] and the NRL group suggest another approach to mass assignment whereby masses can be estimated using plausibility outputs from an evidential reasoning system. Also, J. K. Haspert of IDA (personal communication, Feb 2000) noted the equality of the ratio of plausibilities to the ratio of likelihoods. The first relationship states that the plausibility for a given set is related to the belief of the set's complement as follows:

$$\mathrm{Pls}(A) = 1 - \mathrm{Bel}(A^C) ,$$

where Pls(A) = plausibility of Set A and $\mathrm{Bel}(A^C)$ = belief of the complement of Set A. Note that the belief in A^C is equal to the basic mass assignment of A^C using the relationship

$$\mathrm{Bel}(A^C) = \sum_{D \subset A^C} m(D) ,$$

where $m(D)$ = basic mass assignment of D, a subset of the complement of A.

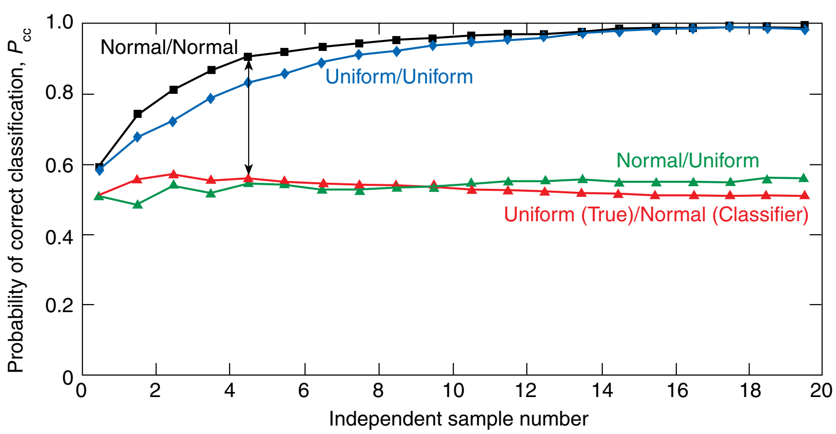

Figure 7. Bayesian classifier probability of correct classification versus independent sample number for all the combinations of target distributions and classifier tuning. (Reference point = sample no. 5; mean P_{cc} = 0.70; spread = 0.36.)

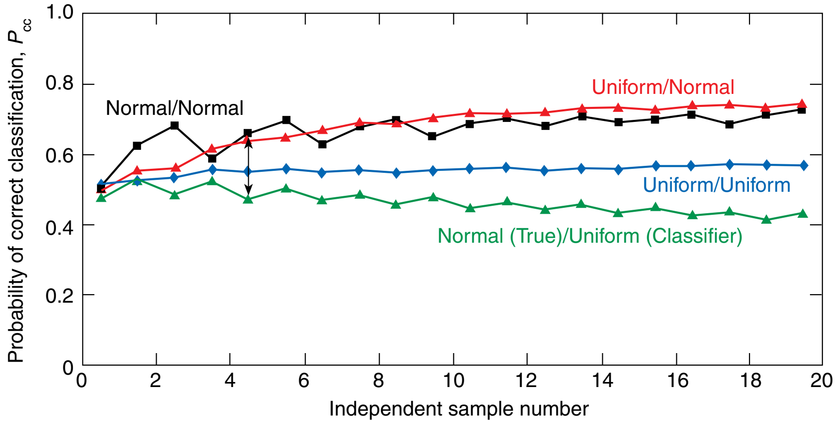

Figure 8. Evidential reasoning classifier probability of correct classification versus independent sample number for all the combinations of target distributions and classifier tuning. (Reference point = sample no. 5; mean P_{cc} = 0.58; spread = 0.19.)

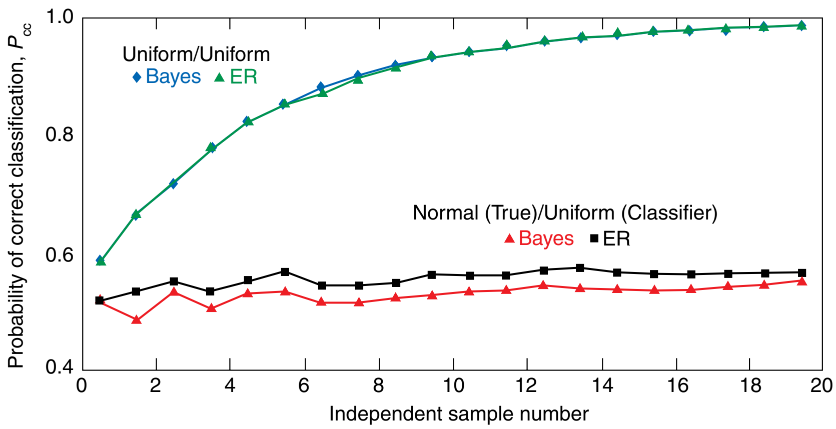

Figure 9. Comparison of Bayesian and evidential reasoning (ER)–based classifiers assuming target features are mixed uniform distributions.

All the mass is assigned to A^C under the assumption that the information at hand does not justify distributing mass to any of the subsets of A^C. This concept is consistent with a maximum entropy approach. Therefore

$$m(A^C) = \mathrm{Bel}(A^C) \ .$$

The remaining mass is assigned to the frame of discernment Θ, i.e., $m(\Theta) = 1 - m(A^C)$. We have made extensive use of both ideas in our current mass assignment algorithm, deriving the plausibilities within regions from likelihoods related directly to the PDFs for each target type. Implementation of these concepts has improved our algorithm's performance by allowing the algorithm to take advantage of the plausibility output of CANEWS 2. Much of this work uses relationships from Shafer's book.[7]

FUTURE EFFORTS

As gleaned from the preceding discussion, much effort has gone into the development of the algorithm infrastructure, analysis of the various mathematical approaches for the reasoning engine, and testing and demonstration of the system. The data used so far have been simplistic for development purposes. To test the various reasoning approaches, we need more realistic data. Extensive testing still needs to be done using distributions drawn from existing and future ESM databases. More effort must be expended on improving the fair scoring method for accurate and realistic algorithm testing. In the near future, NRL will provide APL with ESM likelihood data for more realist input data. The NRL effort to get CANEWS 2 up and running is a critical step in establishing realistic ESM databases.

More emphasis will be placed on understanding the details of sensor performance. Until now, we have assumed that various modes are uniformly distributed. This is a simplification adapted because actual distributions are poorly documented. We have learned many lessons from the implementation of the Blackman and Popoli example, but we must go beyond the simple distributions to the distributions from actual ESM data.

We have focused on the mathematical foundations of evidential reasoning and Bayesian theory in order to give the ID technology an unassailable foundation. The interpretations of the theory still need clarification. Probabilities on sets of sets still need proper interpretations. The human engineering aspects, such as how to effectively display classification and ID information to the warfighter, are also being evaluated.

CONCLUSIONS

We have presented a description of an analysis testbed employed to develop and test algorithms for use in a network of radar and ID sensors. The complexity of the environment requires high-fidelity models of the targets, sensors, and processing algorithms. The processing algorithms first demonstrated in this testbed will be the basis for real-time code implementations. This code will then be demonstrated in real-time laboratory and field tests.

FY2000 activity culminated in a successful demonstration of the CAT with several new processing algorithms, including an ESM interface module, ESM-to-ESM and ESM-to-radar track correlation, and an evidential reasoning–based classification algorithm. In FY2001 these capabilities will be extended to include an MHT module for associating ESM tracks to the composite track database. A Bayesian-based classifier is being added for comparison with the evidential reasoning–based classifier that already exists.

Belief, plausibility, lower and upper probabilities, and Dempster's rule of combination are firmly based on measure theoretic concepts. Dempster's multivalued mapping is actually a random set (i.e., a mapping analogous to the familiar concept of random variable), and the combination rule originates as a product measure on the Cartesian product of measure spaces. The origin of Dempster's independence assumption is the need to construct a joint cumulative distribution associated with the frame of discernment. The problem allows us to know the marginal distributions; the theory tells us how to combine them to get the joint distribution. Further analysis of the correct interpretation of evidential reasoning engine output must take place.

Many processing and algorithmic issues remain to be resolved. The MHT algorithm being integrated into the CAT promises to provide improved performance in correctly associating ESM emitter reports to composite tracks. MHT testing will also provide valuable insights into implementation issues such as number of maintainable hypotheses and speed of processing. With respect to classifier implementation, the CAT will provide valuable insights into the relative performance of the Bayesian and evidential reasoning approaches. There are many significant issues that need to be resolved. Two of these are the focus of much activity: (1) What, if any, is the optimal way to calculate the basic mass assignments, and (2) How should one interpret the evidential intervals associated with the target sets?

The CAT provides a valuable means to evaluate the effectiveness of netting ID sensors, developing and testing alternative data association and tracking algorithms, and comparing and selecting target classification approaches. The algorithms developed in this testbed can serve as the basis for real-time code, testable at land-based sites and at sea, that can ultimately be transitioned to the Fleet to provide much needed improvement in target identification.

APPENDIX A. DEMPSTER'S CONSTRUCTION

Let $(X, \boldsymbol{S}, p)$ be a probability space and let Θ be the frame of discernment, i.e., the collection of platforms that fits the problem. For example, X may be a set of interpretations of data gathered in the field.[12] The data could be detected radiation together with attributes. The interpretations could be the possible emitter, or set of emitters, that could produce such data together with measures of probability of occurrence of the interpretations. $\boldsymbol{S}$ is a σ-algebra of subsets of X. Θ could be a collection of possible answers to a question, e.g., what platforms could possess emitters that would produce the data gathered that would lead us to form our set X of interpretations of these data?

Let $\Gamma: X \rightarrow \Phi$ be a multivalued map from X to a collection Φ of subsets of Θ. The mapping Γ is the association between the interpretations, i.e., emitters, of the measured data and the potential answers to the question of what platform could produce the evidence. For our problem, the platform database provides this association. Given this relationship between the probability space $(X, \boldsymbol{S}, p)$ and Θ, what probability statements can we make about Θ?

The mappings Γ and p induce the probability space $(\Phi, \boldsymbol{T}, \mu)$, where the induced probability function μ is the composition of p and Γ^{-1}, denoted $\mu = p \circ \Gamma^{-1}$, as shown in the figure.

Dempster constructs the upper and lower probabilities of an arbitrary subset A of Θ using three key ingredients. (1) Let A^* be the collection of elements of X such that the image of each of these elements, under Γ, meets A. (2) Similarly, let A_* be the collection of elements of X such that the image of each of these elements under Γ is a non-empty subset of A. (3) Define U to be that subset of X, each of the elements of which has a non-empty image in Φ. These three subsets of X are also elements of the σ-algebra $\boldsymbol{S}$, which are mapped to the unit interval. Then, the following ratios define the upper and lower probabilities of A,

$$P^*(A) = \frac{p(A^*)}{p(U)}$$

and

$$P_*(A) = \frac{p(A_*)}{p(U)},$$

respectively, such that

$$P_*(A) = 1 - P^*(A^C),$$

where A^C is the complement of A.

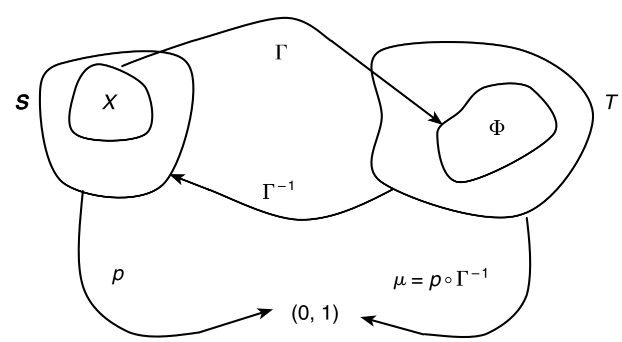

Diagram of mathematical structure.

Dempster formulates the combination of n pieces of evidence using n probability spaces, $(\mathcal{X}_i, \boldsymbol{S}_i, p_i)$, $i = 1, \ldots, n$ together with n multivalued mappings, $\Gamma_1, \ldots, \Gamma_n$. Assuming independence of evidence, Dempster defines the *combined* upper and lower probabilities as the ratios

$$P^*(A) = \frac{p_1(A_1^*)\cdots p_n(A_n^*)}{p_1(U_1)\cdots p_n(U_n)}$$

and

$$P_*(A) = \frac{p_1(A_{1*})\cdots p_n(A_{n*})}{p_1(U_1)\cdots p_n(U_n)} ,$$

respectively, where for each $i = 1, 2, \ldots, n$, (1) A_i^* is the collection of elements of $\mathcal{X}_i$ such that the image of each of these elements under Γ_i meets A, (2) A_{i*} is the collection of elements of $\mathcal{X}_i$ such that the image of each of these elements under Γ_i is a non-empty subset of A, and (3) U_i is the subset of $\mathcal{X}_i$, each of the elements of which has a non-empty image.

We explain Dempster's independence assumption as follows. Our target classification problem is posed such that we have knowledge of the sequence of marginal distributions associated with each of the measurements, but no knowledge of the joint cumulative distribution associated with the combination of measurements. Consequently, we are compelled to invoke independence because the joint cumulative distribution is not available from measurements.

APPENDIX B. SHAFER'S APPROACH

Let Θ be the frame of discernment representing a set of targets carrying the set of emitters that could produce the detected evidence. The basic probability assignment is the function m, with sum unity, from the subsets of Θ to the unit interval.

For any subset A of Θ, Shafer defines belief, Bel(A), and plausibility, Pl(A), of A as

$$\mathrm{Bel}(A) = \sum_{B \subset A} m(B), \ \text{(sum over all subsets } B \text{ of } A) ,$$

and

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B), \ \text{(sum over all subsets } B \text{ of } \Theta \text{ that meet } A) .$$

Belief and plausibility are related by the expression

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(A^C)$$

and numerically

$$\mathrm{Bel}(A) \leq \mathrm{Pl}(A) \ ,$$

where A^C is the complement of A.

If two independent pieces of information (measurement 1 and measurement 2) arrive, they can be combined using the Dempster combination rule

$$m_{12}(C) = \frac{\sum_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j)}{1 - \kappa} ,$$

$$\kappa = \sum_{\substack{i,j \\ A_i \cap B_j = \varnothing}} m_1(A_i) m_2(B_j) ,$$

where m_1 and m_2 are the basic probability assignments associated with *the first* and *second* measurements, respectively. Each A_i, $i = 1, \ldots, k$, is a collection of platforms associated with the *first* measurement, and each B_j, $j = 1, \ldots, n$, is a collection of platforms associated with the *second* measurement. C is a set of platforms for which we wish to calculate its combined basic probability assignment, $m_{12}(C)$. The summation in the numerator is over all pairs of platform sets, one associated with the *first* measurement and the other with the *second*, whose intersection is exactly the subset C of Θ. The factor κ represents the degree of evidential conflict. The summation in κ is over all pairs of platform sets, one associated with the *first* measurement and the other with the *second*, which have no elements in common. The combined mass can be used to calculate the combined belief and plausibility. The concept is used recursively as new information comes in.

APPENDIX C. ELEMENTS OF MEASURE THEORY

We present the following measure theoretic facts to give the reader a feel for some aspects of measure theory applicable to the CCID problem. My primary source for measure theory is Halmos.[13]

A *measure* is a mapping μ from a collection of sets to the non-negative real numbers. The measure of a set can be infinite, but the measure of the empty set $\varnothing$ is always zero.

A *measure space* is a set $\mathcal{X}$ together with a σ-algebra $\boldsymbol{S}$ of subsets of $\mathcal{X}$ such that $\mathcal{X}$ is the union of all the elements of $\boldsymbol{S}$, and a measure μ on $\boldsymbol{S}$. Elements of $\boldsymbol{S}$ are called measurable subsets of $\mathcal{X}$. A measure space is denoted as the triple $(\mathcal{X}, \boldsymbol{S}, \mu)$.

A *probability space* is a specialized measure space for which $\mu(\mathcal{X}) = 1$. Thus, the set of arrival is the unit interval. The measure μ is called a probability measure, and typically we use the symbol p to denote a probability function.

In the CCID application, $\mathcal{X}$ is the set of emitters and $\boldsymbol{S}$ is the σ-algebra of subsets of $\mathcal{X}$ determined solely by the evidence. Of course p is the probability function as determined by the output of the ESM. Dempster's construction begins at this fundamental stage.

Having introduced measurable sets, we now introduce the notion of a measurable mapping. Starting with a probability space $(\mathcal{X}, \boldsymbol{S}, p)$ and an arbitrary set Θ, let there be a mapping Γ from $\mathcal{X}$ to Θ. Γ, then, causes the formation of the σ-algebra $\boldsymbol{T}$ of subsets of Θ. Γ and p, together, create a probability measure μ on $\boldsymbol{T}$, and therefore induce the probability space $(\Theta, \boldsymbol{T}, \mu)$. The inverse images of measurable subsets of Θ are measurable subsets of $\mathcal{X}$. Γ is called a measurable map. In classical probability theory, a random variable is simply a measurable map. It may happen that the elements of Θ could be subsets of another set, in which case Γ is called a random set.[10]

Dempster's multivalued mapping is simply a random set. Shafer refers to this process of inducing probability spaces as the "allocation of probabilities." Shafer's basic probability assignment is nothing more than an induced probability function.

We now introduce the notions of inner and outer measures of a set. The figure gives an intuitive notion of these measures. Suppose the only measurable subsets of X are represented by small rectangles, and that we are interested in the measure of the arbitrary subset A. Since A is not measurable, we cannot calculate its measure. However, the outer measure of A is the measure of the smallest union of rectangles for which A is a subset. The inner measure of A is the measure of the largest union of rectangles for which A is a superset. Thus, the outer and inner measures represent the upper and lower bounds of the measure of A. The outer and inner measures are related in the following way:

$$\mu_*(A) = \mu(X) - \mu^*(A^C)$$

and

$$\mu_*(A) \le \mu^*(A) \; .$$

Hence, plausibility and upper probability are special cases of outer measures. Belief and lower probability are special cases of inner measures.

Measure theory provides an elegant way of combining several probability spaces. The technique utilizes the notion of the Cartesian product of sets. This results in what is called a product probability space, and the probability function on this space is called the product probability. Dempster's rule of combination is simply a special case of a product probability on a product space. The setting of measure theory unifies and simplifies the various concepts introduced by Dempster and Shafer in their studies of evidential reasoning.

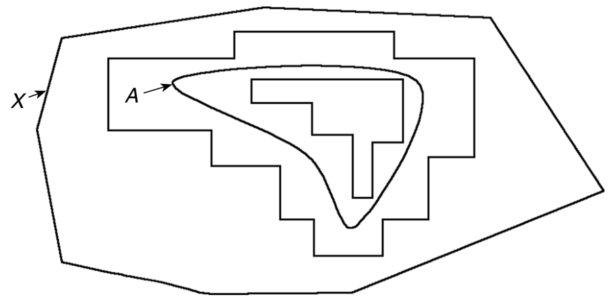

Diagram for the discussion of inner and outer measures.

REFERENCES

[1]Hall, D. L., *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Boston, MA (1992).

[2]Llinas, J., and Waltz, E., *Multisensor Data Fusion*, Artech House, Boston, MA (1990).

[3]*Multi-Source Integration System Engineering Team Final Report* (MSI SET), prepared for the Combat Systems Functional Allocation Board (CSFAB) (1997).

[4]Altoft, J., and Ford, B., *Applications of Dempster-Shafer Belief Theory to CANEWS 2*, No. 1243, Defense Research Establishment, Ottawa (1994).

[5]Kolmogorov, A. N., *Foundations of the Theory of Probability*, Chelsea Publishing Company, Providence, RI (1950).

[6]Dempster, A. P., "Upper and Lower Probabilities Induced by Multi-valued Mapping," *Ann. Math. Stat.* **38**, 325–339 (1967).

[7]Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press (1976).

[8]Fagin, R., and Halpern, J. Y., "Uncertainty, Belief, and Probability," in *Proc. Eleventh Int. Joint Conf. on Artificial Intelligence*, Volume 2, pp. 1161–1167 (1989).

[9]Nguyen, H. T., "On Random Sets and Belief Functions," *J. Math. Anal. Appl.* **65**, 531–542 (1978).

[10]Goodman, I. R., Mahler, R. P. S., and Nguyen, H. T., *Mathematics of Data Fusion*, Kluwer Academic Press, Dordrecht (1997).

[11]Blackman, S., and Popoli, R., *Design and Analysis of Modern Tracking Systems*, Artech House, Boston, MA (1991).

[12]Kohlas, J., and Monney, P. A., A *Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence*, Springer, Berlin (1995).

[13]Halmos, P. R., *Measure Theory*, Springer-Verlag, New York (1974).

ACKNOWLEDGMENTS: I have benefited from discussions with R. Rzemien, M. A. Friedman, F. C. Deal Jr., J. E. Conte, and C. R. Brooks Jr. of APL. R. Rzemien and J. W. Jackson have been analyzing various methods of comparing classifier approaches, and F. C. Deal Jr. has been analyzing various fair scoring issues. Gratitude is directed to the Office of Naval Research for its support of this project. I also thank my friend and math teacher Kwang Yu.

THE AUTHOR

JOSEPH S. J. PERI received a Ph.D. in physics in 1978 from The Catholic University of America. His thesis was on the development of a model of heat transfer between liquid helium and solids. After receiving his degree, he worked in industry for several years, and then joined APL in 1981. Dr. Peri has been involved in a variety of tasks that include analyzing the performance of communications systems that transmit emergency action messages to submerged submarines, developing an age-dependent radar track promotion algorithm still being used in the Cooperative Engagement Capability, modeling infrared (IR) propagation through the atmosphere, and predicting the performance of various IR systems. Currently his efforts support the Composite Combat Identification (CCID) Project, for which the goal is target ID by fusion of data from various sensors. His e-mail address is joseph.peri@jhuapl.edu.