Data Uncertainty Handling in High Level Information Fusion

R. Woodley[1], M. Gosnell[1], and A. Fischer [1]
[1]21st Century Systems, Inc., 6500 Prairie Ave, Omaha, Nebraska, USA

Abstract - *Situation/threat modeling and threat prediction require higher levels of data fusion to provide actionable information to the warfighter. A significant challenge to the fusion of information into higher levels of knowledge is the uncertainty in the underlying data. This uncertainty may be in the form of trust pedigree, sensor noise, and data relevancy. Handling these elements within the fusion structure is vital in order to develop high level information fusion (HLIF) systems for multi-sensory, multi-use applications. 21st Century Systems, Inc. has developed the initial concepts for what we call Fusion with Uncertainty Reasoning using Nested Assessment Characterizer Elements (FURNACE). FURNACE utilizes nested fusion loops building higher levels of information fusion without losing sight of the potential weaknesses of the underlying data. FURNACE uses advanced technologies in information filtering and reasoning to provide the levels of fusion. These reduce bias, disambiguate, and fill gaps in the data. FURNACE handles uncertainty through an innovative evidential reasoning technology that provides the necessary data to the analyst, such that they can account for the pedigree of the information supplied as it is aggregated and fused. Our preliminary results indicate this uncertainty handling scheme is capable of maintaining process standards such that actionable information is produced for the warfighter.*

Keywords: High Level Information Fusion (HLIF), Nested Fusion Loops, Situation Assessment and Modeling, Threat and Impact Assessment, Bias and Ambiguity and Uncertainty Handling

1 Introduction

The FURNACE effort is focused on the process and algorithms for high level data fusion with improved handling for bias, ambiguity, and uncertainty (BAU). Figure 1 shows our conceptual diagram of how FURNACE will support higher level fusion processes. FURNACE changes the paradigm in that it does not view the fusion levels as a sequential hierarchy. Instead, FURNACE parallels a meta-tagging scheme that adds the fused information as metadata to the existing data. Each level of FURNACE deals with the direct data plus the added metadata generated by each level of the fusion. FURNACE is able to account for data that is controlled by the analyst (i.e., a set of sensors or known data sources where the analyst can direct the content) as well as 'outsourced' data sources where the data is being repurposed for the analyst's needs and not collected specifically for those needs. The concept is applicable to all levels of data fusion. Initial work has developed the underlying algorithmic needs to produce a fusion system able to handle BAU, repurposed data, and do so in a cohesive manner. FURNACE creates the framework by which the user can connect feeds, define the domain, utilize repurposed data, and add context to information.

FURNACE takes a cue from the way human situational awareness is modeled to create an innovative data fusion system. By parallelizing the fusion process (i.e., getting away from the sequential hierarchy paradigm), the higher level fusion emerges from the data. The continual feedback

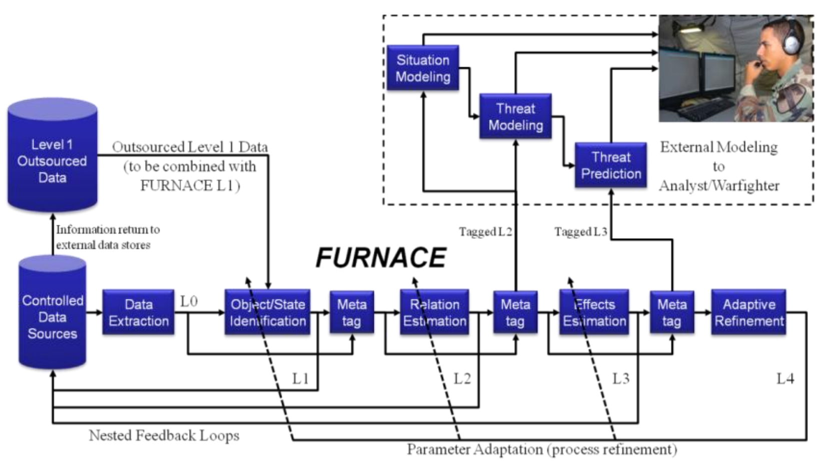

Figure 1: Conceptual diagram of the nested fusion loops.

bolsters true evidence, while ambiguous, or even fraudulent, data is suppressed. The structure of FURNACE provides a wealth of power to reduce bias and disambiguate data, but we also add filtering and fusion processes to the FURNACE concept. These algorithms help FURNACE further identify relevant data and helps fill in the gaps caused by missing data. A data uncertainty handling capability rounds out FURNACE's arsenal by helping the analyst understand the trustworthiness of the data so that proper decisions are made from the fused information FURNACE generates.

Our initial results show that the FURNACE concept is technically feasible. The resulting design and proof-of-concept form the basis for future development and a testbed to showcase FURNACE's abilities. We highlight here an example scenario from the Global Intelligence, Surveillance and Reconnaissance (GISR) domain to demonstrate the algorithms and concepts. FURNACE's design is different from previous fusion systems in that each fusion level bears symmetry with the other levels in the form and function of the design. Using a common fusion engine [1], [2], abstracted for both high and low level information fusion, provides a unique opportunity to setup an advanced nested feedback system that drives the reduction of BAU, as well as stabilizes the fusion results. This system is designed to handle BAU and repurposed data at an intrinsic level rather than treat it as an outside calculation. Given the period of performance constraint on initial Phase I work, we were able to show FURNACE operating up to Level 2. However, we show that the abstractions made in the design should be able to be adapted for any fusion level which we will show in Phase II development.

2 Example Scenario and Evidence Reasoning

We now describe the example scenario and data reasoning algorithm. The scenario is designed to test the data reasoning component of FURNACE. While the scenario is not overly complex, it does show where the data reasoning is able to modify the fusion results as it is applied to the feedback mechanism. The change in the fused data can be seen in the Results Section.

2.1 Example Scenario

This Scenario showcases the feedback concept in that higher level fusion reduces the uncertainty at the lower level to make additional combinations. To date, the feedback is designed for the Level 2 to Level 1 path, but the concept is general enough to be used at higher levels.

Figure 2 is the conceptual drawing of the scenario and Figure 3 is a screen capture from the simulation. The images show two entities (E1, E2) entering a building. A few minutes later two more entities (E3, E4) exit the building. The Area of Interest (AOI) is covered by three sensors. A GMTI-radar detects movement in a large area around the building. An EO-camera (EO1) has a Field of View (FOV) covering the front entrance. A second EO-camera (EO2) has a FOV covering the parking lot, but does not see the exit. E1 and E2 are detected by GMTI and EO1, while E3 is first detected by the GMTI and then EO2. E4 is only ever detected by the GMTI. There exists domain knowledge that no vehicle may enter the building. E1 and E3 have similar appearance, but there is no information about what happens inside the building to connect the two directly.

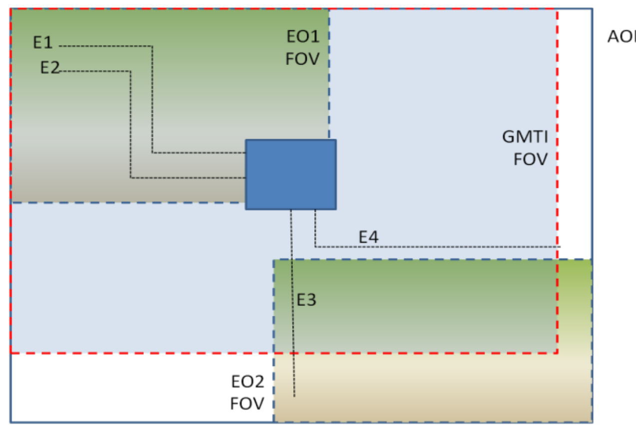

Figure 2: Concept drawing of Example Scenario

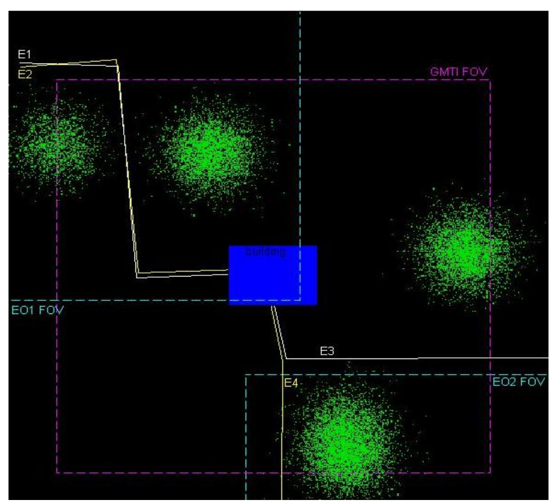

Figure 3: Example Scenario screen capture

2.2 Evidential Reasoning Network (ERN®)

Typical decision-support approaches will use either a simplistic uncertainty tracking method or something along the lines of a Bayesian probability approach. Simple uncertainty tracking does not fully account for the propagation and combination of uncertainty. It does not propagate the error whereby it may allow potentially erroneous data to bias the results. Bayesian approaches are better and account for the error propagation, but have the basic need of *a priori* probability measures on the uncertain elements. What is sometimes needed is a way to incorporate various degrees of uncertainty ranging from simple percent unknown up to probabilistic measures, where available. 21CSi's Evidential Reasoning Network (ERN®) technology is designed for this purpose.

ERN technology uses a belief algebra structure for providing a mathematically rigorous representation and manipulation of uncertainty within the evidential reasoning network. Since the introduction of the Dempster-Shafer Theory of Evidence [3], new evidential reasoning methods have been, and continue to be, developed, including fuzzy logic [4] and Subjective Logic [5], [6]. Recently, Mr. Steven O'Hara from 21CSi worked with Dr. Jøsang (creator of Subjective Logic) to develop Hypothesis Abduction using Subjective Logic (Analysis of Competing Hypothesis) [7], [8]. An evidential reasoning framework was needed to ensure that evidential reasoning expressions are coherent, consistent, and computationally tractable. 21CSi's Evidential Reasoning Network is a novel structure that addresses these needs. The two prime belief algebra operators required are *consensus* and *discount*. These operators allow the propagation of belief values through the network amongst various opinion generating authorities, such as human subject matter experts or software agents that perform some sort of data analysis, processing, and reasoning. The belief algebra structure is capable of using probabilistic belief mass assignments through the use of belief frames. The ERN Toolkit includes a Subjective Logic and Dempster-Shafer belief algebra implementation.

Subjective Logic (SL) [9] is a way of thinking about uncertainty that builds upon the basic ideas presented by Dempster and Shafer to incorporate the subjectivity of all observations. In Subjective Logic, we operate on opinions as opposed to facts. An *opinion* ω_x^A on a subject x by a party A is a 4-tuple of the belief (b_x^A), disbelief (d_x^A), uncertainty (u_x^A), and relative atomicity (a_x^A) (with respect to all possible states) about subject x. Note that $b_x + d_x + u_x = 1$, so while it is not necessary to specify all three of these values, it is convenient when performing certain calculations.

SL introduces the *consensus* operator to combine opinions and the *discount* operator to support the belief in the *source* of an opinion. It has been shown that the consensus combination rule generates more intuitively correct results than common variants of Dempster's rule [5], [6]. Subjective Logic can be viewed as an extension to binary logic and probability calculus.

The consensus between opinions ω_x^A and ω_x^B is defined by the formulas in Figure 4. In the case where we are dealing with dogmatic opinions (those with no uncertainty), then K=0, and a slightly different form of these equations is needed, and can be found in the referenced literature on Subjective Logic [9]. We use the $\oplus$ symbol to represent the consensus operator. The discount operator represents an opinion about another opinion, or the source of the opinion. The opinion ω_B^A represents the opinion of B by A. This is a model for the concept of trust, where an opinion/source you trust would be discounted slightly, while an opinion that is not trustworthy would be discounted greatly. Figure 5shows the SL Discount operator. We use a $\otimes$ symbol to represent a discount operator.

$$
\begin{aligned}
K &= u_x^A + u_x^B - u_x^A u_x^B \\
b_x^{A,B} &= \frac{b_x^A u_x^B + b_x^B u_x^A}{K} \\
d_x^{A,B} &= \frac{d_x^A u_x^B + d_x^B u_x^A}{K} \\
u_x^{A,B} &= \frac{u_x^A u_x^B}{K} \\
a_x^{A,B} &= \frac{a_x^A u_x^B + a_x^B u_x^A - (a_x^A + a_x^B) u_x^A u_x^B}{K - u_x^A u_x^B}
\end{aligned}
$$

Figure 4: Consensus Operation

$$
\begin{aligned}
b_x^{A,B} &= b_B^A b_x^B \\
d_x^{A,B} &= b_B^A d_x^B \\
u_x^{A,B} &= d_B^A + u_B^A + b_B^A u_x^B \\
a_x^{A,B} &= a_x^B
\end{aligned}
$$

Figure 5: Discount Operation

The expressivity of the belief algebra is important in a heterogeneous system that may be incorporating some mixture of probabilistic and evidential reasoning. When working in known probability measure spaces, the belief algebra should reduce to probability calculus to preserve the accuracy and functionality of the supporting probabilistic systems—and Subjective Logic is easily shown to do so.

3 Results

The design of the feedback mechanism has two components: the threshold function to determine if feedback is necessary and the uncertainty adjustment to produce the actual feedback information. The first thing that the feedback mechanism does is determine if there is need to send back any information. This is done for two reasons: First, since every relationship (including all primary, secondary, etc.) can potentially generate feedback, we would quickly create a logjam of data that would slow the process. By forcing the relationships to pass a threshold (i.e., a sniff-test to see if there is anything unusual that hampers the relationship (either in believability or uncertainty)) we only require the system to analyze that data and not everything. Second, we also eliminate many race conditions. By forcing the system to stabilize once it hits a threshold, feedback oscillations are attenuated while the system identifies and updates characteristics about an entity.

To calculate the threshold we utilize 21CSi's ERN technology. If we consider each relationship from the scenario as an opinion, then the data that forms the relationship is the evidence. Equation 1 shows an example of the calculation for relationship R1 based upon the entities E1 and E2 which form R1:

$$\omega_{R1}^{E1+E2} = \omega_{R1}^{E1} \wedge \omega_{R1}^{E2} \qquad \text{Eq. 1}$$

Equation 1, in belief algebra, says that the opinion on R1 is the opinion multiplication of the opinion of E1 about R1 and the opinion of E2 about R1. For our purposes, we are using the multiplication operator since the consensus operator is too sensitive to calculations with a dogmatic condition either in belief or disbelief. The opinion multiplication operator ^ is calculated as:

$$b_{x\wedge y} = b_x b_y + \frac{(1-a_x)a_y b_x u_y + a_x(1-a_y)u_x b_y}{1-a_x a_y},$$

$$d_{x\wedge y} = d_x + d_y - d_x d_y,$$

$$u_{x\wedge y} = u_x u_y + \frac{(1-a_y)b_x u_y + (1-a_y)u_x b_y}{1-a_x a_y},$$

$$a_{x\wedge y} = a_x a_y.$$

The opinion E1 about R1 takes into account how much E1 effects R1, the uncertainty of E1, and any additional domain knowledge that can affect the relationship. Then by analyzing the resulting opinion's disbelief (which is a function of both the level of dissimilarity and uncertainty in the opinion) we have a measure of the need for re-evaluating the entities that formed the relationship.

If the opinion about R1 has zero disbelief there is no need to re-evaluate the entities E1 and E2. However, if a relationship generates a feedback request, FURNACE would then determine what should be fed back. The relationships in the example are:

- R1: E1 – E2 spatial close
- E1 – E2 temporal close
- R2: E1 – Bg location close
- R3: E2 – Bg location close
- R4: E3 – Bg location close
- R5: E4 – Bg location close
- R6: E3 – E4 spatial close
- E3 – E4 temporal close

Suppose Relationship R5 reaches a threshold. With domain knowledge from the GISR examples stating that vehicles cannot exit the building, the opinion generated on R5 from the Building (Bg) will contain relatively high disbelief that E4 is a vehicle. Suppose E4 is initially classified as possibly vehicular (with high uncertainty). In this case, when the opinion multiplication combines the E4 and building opinions into R5, the disbelief reaches the threshold to be re-evaluated. When the updated R5 opinion is fed back, it forces the uncertainty that E4 is a human to decrease. When Level 1 processes the new uncertainty levels, it is able to determine that E4 is a human and not a vehicle.

The R1 opinion from Equation 1 is a primary relationship, dealing only with direct connections. Secondary relationships are harder to show and may or may not be useful. Table 1 shows a connectivity matrix example where the relationship label indicates a primary connection. If we look at the table, we see that E4 has primary connections to E3 and the building (Bg). However, Bg has primary connections to all four entities which implies that E4 has secondary connections E1 and E2. We can then form secondary opinion equations similar to Equation 1 using the opinion of E1 about E4 and so forth.

Table 1: Connectivity matrix

E1	-				
E2	R1	-			
E3			-		
E4			R6	-	
Bg	R2	R3	R4	R5	-
	E1	E2	E3	E4	Bg

The feedback due to primary connections affects the *entities* individually that formed the relationship. However, the secondary connections affect the *relationship* between the entities. Equation 2 shows the opinion of the secondary relationship between E1 and E3.

$$\omega_{E1,E3}^{R2+R4} = \omega_{E1,E3}^{R2} \wedge \omega_{E1,E3}^{R4} \qquad \text{Eq. 2}$$

When this is calculated it shows a high belief that a relationship exists between E1 and E3. The relationship is due to the secondary connection with building between the two entities as well as the similarity in appearance (which is obtained as metadata when processed in Level 1). This secondary connection opinion reaches the threshold to activate the feedback mechanism. This time, however, the feedback reduces the uncertainty such that the data for E1 and the data for E3 are the same entity. When Level 1 re-calculates the entities, it creates a new entity list.

Table 2: Base rate and probability expectation calculations

	Veh	person	Bldg	Irrelevant	Uncertainty
Person - E4	0.22	0.22		0.22	0.34
E4(a)	0.15	0.2	0.05	0.6	
P(E4)	0.271	0.288	0.017	0.424	
Building B - E5			0.97		0.03
E5(a)	0.15	0.2	0.05	0.6	
P(E5)	0.0045	0.006	0.9715	0.018	

Table 3: Context truth table

Relationships	Veh	person	Bldg	Irrelevant	
Building A	T	T	F	T	Garage, can take cars and people
Building B	F	T	F	T	Regular building, only people can interact
Building S	F	T	F	F	Secure Facility, need to know everything
Person	T	T	T	T	
Vehicle	F	T	T	T	

Table 4: **Relationship opinion calculation**

Relationship Opinions - Building A

R5		B	D	U	A	P(x)
O1	E5 Opinion of R5	0.983	0.017		0.5	98%
O2	E4 Opinion of R5	1	0		0.5	100%
P(R5) = 98%	O1 ^ O2	0.98	0.02		0.25	**98%**

Relationship Opinions - Building B

R5		B	D	U	A	P(x)
O1	E5 Opinion of R5	0.712	0.288		0.5	71%
O2	E4 Opinion of R5	1	0		0.5	100%
P(R5) = 71%	O1 ^ O2	0.71	0.29		0.25	**71%**

The above example illustrates that the bias produced by the initially mislabeled entity E4 was reduced along with the uncertainty and ambiguity allowing the system to correctly classify the object. As the examples become more realistic and more complex we will be able to utilize the feedback system to iteratively correct the analysis and provide the best possible fusion results to the analyst. The actual calculations are somewhat more involved than portrayed above, but the principles still apply. Using data from the Example Scenario and the above analysis we calculate the base rate (Ex(a)) and probability expectation (P(Ex)) using the Evidential Reasoning Network for the entities involved in the relationship, shown in Table 2.

We next construct the *a priori* truth table of the possible context (which could possibly also be learned context). Table 3 shows three possible building types that could be included to extend this scenario. We will show first the calculation for a building (Type A) that would allow vehicles near it and then redo the calculation for the Type B building to illustrate the contextual aspect of FURNACE.

We can now calculate the opinion on the relationship as shown in Table 4. Note that when Entity E4 might be classified as a vehicle, the Type A building allows for vehicles, so it concludes that the relationship is 98% valid, so no feedback is needed. However, Type B does not allow vehicles and would trigger the feedback mechanism since the relationship is considered only 71% valid (see Table 4).

4 Conclusions

We investigated how a holistic fusion approach could be constructed and how uncertainty could be measured and then manipulated by the ERN technology. We also designed a feedback mechanism around the ERN technology which would help stabilize and prevent race conditions in the data feedback. By analyzing the level of disbelief in the fused output (along with its associated uncertainty) we could produce an intelligent threshold that would indicate if fused information is in need of additional processing. The actual feedback acts to either reduce or increase uncertainty such that lower fusion processes can make better decisions about the objects. Preliminary results show that the contextual information in the initial scenarios is successful in providing relevant feedback and reductions in uncertainty to provide fused output. These results indicate the approach to be feasible, but more work is needed to verify and increase the robustness of the concept through additional data and increasingly complex and higher level information fusion.

5 Acknowledgment

21st Century Systems, Inc. would like to acknowledge the support of the Office of the Secretary of Defense and the U.S. Air Force Research Laboratory (Contract No: FA8750-12-C-0168).

6 References

[1] R. Woodley, M. Gosnell, and A. Fischer, "High Level Information Fusion (HLIF) with Nested Fusion Loops," in *SPIE Signal Processing, Sensor Fusion, and Target Recognition XXII*, Baltimore MD.

[2] M. Hansen and C. Donahue, "Passive Sonar Fusion Multi-Model Ensemble Agents (MMEA)," 21st Century Systems, Inc, Technical Report, Oct. 2005.

[3] G. Shafer, *A mathematical theory of evidence*. Princeton NJ: Princeton University Press, 1976.

[4] P. Palacharla and P. Nelson, "Understanding relations between fuzzy logic and evidential reasoning methods," in *IEEE Proceedings of the Third IEEE Conference on World Congress on Computational Intelligence*, 1994, pp. 1933–1938.

[5] A. Jøsang, "A Logic for Uncertain Probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279–311, Jun. 2001.

[6] A. Jøsang, "Subjective Evidential Reasoning," *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002)*, pp. 1671–1678, Jul. 2002.

[7] A. Jøsang, S. O'Hara, and K. O'Grady, "Base Rates for Belief Functions," in *Proc. Workshop on the Theory of Belief Functions (Belief 2010)*, 2010.

[8] A. Jøsang and S. O'Hara, "Product of Multinomial Opinions," in *Proc. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*, 2010.

[9] A. Jøsang, *Subjective Logic: Draft*. Online at http://folk.uio.no/josang/papers/subjective_logic.pdf, 2013.