Assessing Confidence in Situation Awareness

John Palmer
Overwatch Tactical Operations
Textron Systems
Austin, TX U.S.A.
jpalmer@overwatch.textron.com

Abstract* - *Situation Awareness enables the discovery of aggregations and the identification of interesting patterns in underlying data that can be leveraged to further the understanding of the battlespace. While there have been steady efforts within the information fusion community to increase the level of automated reasoning supporting Situation Awareness, there remain unresolved issues such as the development of standard metrics of trust.

Aggregation methods applicable to one sort of analysis may not be parameterized in the same fashion as another. New methods may be introduced. Together, these belie the possibility of a universal a-priori understanding of the factors that may temper a method's reliability. To accommodate such variability, this paper adopts a non-parametric approach to the assignment of a confidence metric. It introduces a measure similar to Hubert's Γ but which incorporates a measure previously shown helpful in assessing the effectiveness of object refinement engines. Results illustrate its application.

Keywords: Level 2 Fusion, Situation Awareness, confidence, trust, aggregation, cluster analysis.

1 Introduction

Situation Awareness (SA) is a process within data fusion that supports the "orientation" phase of the intelligence cycle oftentimes labeled the Observe Orient Decide Act (OODA) loop. Among its objectives, SA reasoning paradigms attempt to discern the associations among a collection of objects, relationships, or some higher order interacting derivative of these components, such as events.

While there have been steady efforts within the information fusion community to increase the level of automated reasoning that effectively supports Situation Awareness, there remain challenges regarding the most viable methods of information elicitation and representation [1]. One of these concerns embraces issues related to information uncertainty. Being intertwined with user trust, the manner and perceived fairness with which uncertainty is treated is critical in forging a bond between the analytical tool and the user. The criticality and sometimes tenuous nature of trust has been previously documented [2].

Modern situation awareness engines may employ one or more aggregation methods that resist being readily pigeonholed into disciplines having particular parameter sets. Rather, these engines employ techniques that may consider similarities over a wide range of attributes or behaviors such as entities sharing a common motion, a similar destination, or particular commonalities in modes of communication. Aggregation methods applicable to one sort of analysis may not be parameterized in the same fashion as those of another. Moreover, new methods such as affinity propagation and Markov clustering are introduced from time to time which belie the possibility of a universal a-priori understanding of the factors affecting a method's reliability.

For this reason, this paper advocates using a non-parametric approach in assessing the validity of non hierarchical aggregation methods. It introduces a measure similar to Hubert's Γ but which incorporates a quantifier, termed *Clarity*, previously found helpful in assessing the effectiveness of object refinement engines [3][4].

The method proposes using Monte Carlo techniques to vary the input data by randomly holding out a minimum number of observations in order to assess the stability of the aggregation method. This assessment of a method's stability with respect to a particular dataset measures the level of confidence that may be ascribed to the application of a particular method to a particular collection of data.

2 Background

Interest in this topic arose as a natural consequence of the development of a high profile fusion system that included requirements not only to develop a broad range of meaningful aggregations but assign confidence metrics to the resulting products. This paper looks at those subsets of aggregations that can be fairly characterized as non-hierarchical clusters. In fairness, it should be noted that the concepts of "aggregating" and "clustering" are not synonymous in that the latter includes nuances of collecting together entities of a more uniform character than does the former. However, the approach proposed for validity assessment is not sensitive to this distinction.

Substantial effort has been expended in exploring methods for assessing the validity of cluster analysis procedures and this work was taken as the starting point of this inquiry. Of particular interest are the comprehensive reviews of applicable methods for validity assessment that can be found in [5], [6] and also [7].

A persistent dilemma repeatedly surfaces during any discussion regarding the propriety of assessing cluster validity. Central to this problem is the difficulty in identifying an unbiased frame of reference in which to form an assessment. A partition P imposed upon the data by a clustering algorithm purportedly mirrors a user's intuition or other standard about what is thought to be the "natural" clustering of the data. A recurring concern with terms like "natural clustering" is that the approach itself may subtly be inscribing human values and prejudice into the concept of the term "natural" and that, in fact, there may be nothing natural about the structure at all. Moreover, if the correct structure of P can be algorithmically determined beforehand, then that same algorithm could be used as the basis of the clustering method and produce flawless results – thus rendering the concept of cluster validity a moot point.

To circumvent the lack of a universal naturally available exemplar, there are three general approaches used in assessing cluster validity. One approach relies upon comparisons with external pre-existing criteria.

2.1 External Criteria

In this case, it is assumed that an algorithm is applied to a set $X = \{X_1, X_2, \ldots, X_N\}$ of entities and produces a clustering $C = \{C_1, C_2, \ldots, C_c\}$ which is then evaluated against an externally imposed partition of the same data $P = \{P_1, P_2, \ldots, P_p\}$. Statistics are applied that assess the degree of concordance between the derived clustering C and an externally supplied partition P. These statistics generally depend upon the tabulation of similar and dissimilar membership in the subsets of these aggregations. For each pair $\{X_i, X_j\}$ whose components are selected from X, it may be the case that both of the components in $\{X_i, X_j\}$ belong to the same cluster C_k and the same partition subset P_l ; or they may belong to the same cluster in C but different partitions in P, or vice versa, or share no joint membership in either C or P. By tabulating the membership outcomes for all the $\frac{N(N-1)}{2}$ combinations of possible paired choices from X and then combining the results, a multitude of statistics have been developed to measure the concordance between the cluster and the partition. The most notable of these are the Rand Statistic, the Jaccard Coefficient, the Folkes and Mallows Index, the Hubert' Γ, and normalized Γ statistic. The interested reader may find a description of these and other cluster validity measures in [5][6][7][8].

Apart from those cases wherein the user constructs their own partition and desires to assess its degree of similarity to a system rendered product, it is difficult to see, within a deployment environment, what mechanism would provide the external validation criteria needed to enable an assessment of aggregation confidence.

2.2 Internal Criteria

A second approach for assessing cluster validity utilizes quantities that are construed as inherent in the data. Most frequently, this internal quantity is taken as the proximity matrix which organizes all the pairwise similarities among the data. Similarity functions provides a measure in the closed interval [0, 1] and that behave as an inverse function of distance. Using the previous population notation $X = \{X_1, X_2, \ldots, X_N\}$, the proximity matrix is denoted as $S = (s_{ij})$ where $s_{ij} = \text{similarity}\{X_i, X_j\}$.

To illustrate the application of Hubert's Γ as an internal measure using the proximity matrix, denote by Y a matrix $Y = (y_{ij})$ where $y_{ij} = 1$ if x_i and x_j belong to distinct clusters and 0 otherwise. Applying Hubert's Γ to P and Y gives a measure of conformance between the clustering and the proximity matrix. In terms of the proximity matrix S and the matrix Y, Hubert's Γ statistic is:

$$\Gamma = (1/M) \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} s(i,j)\, y(i,j) \qquad (1)$$

where $M = \frac{N(N-1)}{2}$.

2.3 Relative Criteria

Both the internal and external based methods of developing cluster validity measures are perceived to suffer from significant computational overhead [5].

A third approach that skirts these objections measures cluster validity through the use of a relative criteria. Historically, the focal idea in this approach is that the evaluation of a clustering structure is found by comparing it to other clustering schemes, obtained by using the same algorithm but utilizing different parameter values [5]. Variations on this concept that circumvent the need to vary parameters motivate the approach described in the sequel. Instead of performing a variation of parameters, the proposed approach uses Monte Carlo technique to assess the stability of the aggregation method.

3 Methodology

Denote the aggregation algorithm of interest by K. This algorithm operates on various collections of reports provided by Level 1 fusion -- an object refinement engine. Apart from the collection of attributes which may be used in aggregation, each report is augmented with an additional measure assigned by the Object Refinement Service (ORS) that indicates the confidence or degree of belief that the attributes, as reported, correctly characterize the entity.

One wishes to estimate the confidence level associated with a particular aggregation methodology (K)applied to this data. This is accomplished by repetitively self-pairing the aggregation engine with itself by utilizing two separate samples from the original data, where in forming each sample one report, selected at random, is withheld from the sample.

3.1 Precision of Confidence Estimate

As developed in the sequel, each Monte Carlo iteration produces a single confidence estimate. Each estimate relates to the particular aggregation method and the associated dataset being considered. Because sampling variability introduces a degree of scintillation in the computed confidence measure, it is helpful to characterize the precision which may be expected in the estimate that will be obtained from a particular number of iterations. Computing the appropriate number of iterations to be performed requires specifying both the desired precision of the estimate as well as the minimum probability that the precision is achieved, in light of the available computational budget.

In deployment environments wherein computational resources are in great demand, the aforementioned sequencing may properly be viewed as an instance of the "tail wagging the dog" and an inverse sequencing of considerations put in order. In a more realistic scenario, the computational budget is set which then dictates the number of iterations which yields the precision of the estimate.

If N repeated Monte Carlo samples are obtained and $\hat{a}$ denotes the arithmetic mean of the individual confidence estimates, then the precision E of $\hat{a}$ can be inferred from the dimensions of its corresponding Wilson confidence interval to be [9][10]:

$$E = \Phi^{-1}\left(\frac{1+C}{2}\right)\frac{\sqrt{\frac{\hat{a}\,(1-\hat{a}\,)}{N}+\frac{\left[\Phi^{-1}\left(\frac{1+C}{2}\right)\right]^2}{4N^2}}}{1+\frac{1}{N}\left(\Phi^{-1}\left(\frac{1+C}{2}\right)\right)^2} \qquad (2)$$

Here, $\Phi^{-1}\left(\frac{1+C}{2}\right)$ denotes the inverse of the cumulative normal distribution function evaluated at a specified certainty C, usually taken as .90, while $\hat{a}$ is the current confidence estimate, and N is the number of iterations performed. The formula for precision, given above, requires the samples to be obtained independently and this will not be the case in the Monte Carlo sampling procedure discussed. For this reason, equation (2), which is already an approximation, must be considered even more so in this setting. (Engineering Statistics Handbook 7.2.4.1)

3.2 Constructing the Monte Carlo Sample

Denote the input report data provided by the object refinement service(s) as $R = \{R_1, R_2, \ldots, R_n\}$. Let N denote the number of iterations that will be performed in computing the aggregation confidence estimate.

On each iteration, two samples are obtained from the collection of reports R. Designate the samples obtained on the i th iteration as R_1^i and R_2^i. Each sample R_j^i is formed by selecting an index k_j at random from R's original set of indices $\{1, 2, \ldots, n\}$ and then removing that report from the original set to create the sample. The paired samples on the i th iteration are:

$R_j^i = R - \left\{R_{k_j}\right\}$ for j = 1,2. Each of the subsamples R_1^i and R_2^i are of size $n-1$ and differ from R by one report. (The same report is not to be removed in a single pairing.) This is an application of the *Leave One Out* method similar to that described in [8 Pattern Analysis p 570]

Label the clusters resulting from the application of aggregation algorithm K to the two samples R_1^i and R_2^i as $C_{1\,1}, C_{1\,2}, \ldots C_{1\,n_1}$ and $C_{2\,1}, C_{2\,2}, \ldots C_{2\,n_2}$. Note that the method K may produce a differing number of clusters: n_1 *and* n_2, when it is applied to the two different samples. (Notational consistency would be better served by the denoting a cluster on the i th iteration as $C_{k\,j}^i$ but for simplicity the iteration number is cloaked when discussing the individual clusters.)

3.3 Constructing the Association Matrix

Knowing the clusters contents developed from the two samples allow the creation of an association matrix that records the consistency of cluster assignments between the two samples. (This matrix is often called the "Confusion Matrix".) This analysis excludes the reports withheld in sample formation. The association matrix, denoted $A_{m\times m}$, where $m = max(n_1, n_2)$ guides the assessment of how consistently the algorithm performed. Measuring the consistency of behavior gives some sense of the methods constancy or stability. These perceptions, in turn, bear directly upon a user's confidence in the algorithm results.

The matrix $A_{m\times m} = (c_{\,i\,j})$ is initialized to zero. During the h th iteration, the paired sample R_1^h and R_2^h is constructed as described above by holding out a single random report. Each report is reviewed which is both a member of R_1^h and R_2^h. The count c_{ij} in A is incremented each time a report appears in cluster i formed by the aggregation method applied to first sample of the pair, $K(R_1^h)$, and in cluster j formed by the aggregation method being applied to the second sample, $K(R_2^h)$. No tabulation occurs when encountering a report that has been held out of either sample.

Augment this matrix by adjoining the row totals (O_i) and column totals (C_j). Denote by T the grand total of the elements of A.

$$A = \begin{pmatrix} c_{1\,1} & \cdots & c_{1\,m} \\ \vdots & \ddots & \vdots \\ c_{m\,1} & \cdots & c_{m\,m} \end{pmatrix}\begin{pmatrix} O_1 \\ \vdots \\ O_m \end{pmatrix}$$
$$(C_1 \quad \ldots \quad C_m)$$

3.4 Estimating Confidence

For this sample, an estimate of the confidence or reliability of K can be ascertained by assessing the probability of a one-to-one correspondence between the aggregations generated from the distinct samples. This measure, previously called *Clarity,* measures the probability that there is a one-to-one correspondence between the clusters produced from the two samples and has previously been used as a measure of effectiveness for evaluating object refinement engines.[4]

To illustrate, consider computing the probability of a one-to-one correspondence between the i^{th} cluster derived from the first sample and the j^{th} cluster generated from the second sample. Conceptually, if one selects a report at random from the i^{th} cluster of the first sample [i.e.$K(R_1)$], then, by inspecting the augmented A matrix, it is easily observed that the probability that this particular choice is associated with the j^{th} cluster formed from the second sample is:

$$\frac{c_{i\,j}}{O_i}$$

Given this probability that this selected report is associated with the $j\ th$ cluster of $K(R_2)$, select another report at random, this time from the j th cluster of $K(R_2)$. The probability that this newly selected report will reciprocally be associated with the initial i^{th} cluster in $K(R_1)$ is:

$$\frac{c_{i\,j}}{C_j}$$

Multiplying these two probabilities provides the joint probability that a report selected from the i^{th} cluster of $K(R_1)$ traced to the j^{th} cluster of $K(R_2)$ will be reciprocally associated in a random round-trip back to the i^{th} cluster of $K(R_1)$. This is taken to be the probability of a one-to-one correspondence between these aggregations. The higher the degree of correspondence, then the higher the clarity with which one collection of aggregates mirrors another.

Clarity is just the weighted average of these cell-based one-to-one correspondence probabilities, each cell being weighted by its relative frequency with respect to the entire matrix. This weighted average is:

$$\gamma = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{c_{ij}^{3}}{O_i C_j}}{T} \tag{3}$$

This is the probability of a one-to-one correspondence between the clusters' memberships derived from the two samples. Repetitively computing this statistic over many iterations provides a sense of the consistency with which the given aggregation algorithm K can be expected to perform on the dataset R.

Denote by $\Gamma_K(R)$ the confidence ascribed to the aggregation algorithm K applied to the set R. Then, denoting the value of *Clarity* computed on the $q\ th$ iteration as γ_q, the mean value of $Clarity$ over the N iterations are:

$$\widehat{\Gamma_K(R)} = \frac{\sum_{q=1}^{N}\gamma_q}{N} \tag{4}$$

and this provides a measure of confidence of the aggregation process applied to the available data. These calculations assume data provided by the object refinement service is error free. In some cases, the object refinement service reports a confidence level in its certainty of its classification. Let L_i denote the reported confidence provided by the object refinement service associated with reportR_i. Then it is a simple matter to characterize the average confidence over R, the set of n reports by using the geometric mean.

$$\hat{L} = \left(\prod_{i=1}^{n} L_i\right)^{1/n} \tag{5}$$

Seeing as the confidence measure ascribed to the aggregation $K(R)$in Equation 4 is conditioned upon an assumed unerring performance of the object refinement service, a more complete measure of confidence will include estimates of uncertainty assessed during earlier stages of processing. Including these considerations gives:

$$\widehat{\Gamma_K(R)}_{Adj} = \widehat{\Gamma_K(R)} \times \hat{\mathrm{L}} \tag{6}$$

The adjusted value provides an overall assessment of the confidence level attributable to aggregation method being applied to the data that is available.

3.5 Illustrative Results

The following illustration uses simulated data, both in the representation of reports that mimic the product of the object refinement service, as well as the repetitive application of a k-means clustering algorithm. The k-means algorithm mirrors a commonly required situation awareness aggregation service that groups entities based upon proximity.

The collection of 37 reports, denoted R in §3.2, provide current estimates regarding specific platforms associated with an unknown number of units.

Table 1: Abridged Collection of Simulated L1 Reports

Id	Lat	Long	Course	Speed	ORS Confidence
101	34.4661	60.1862	140.4252	13.7966	0.571
102	34.4661	60.1862	140.4646	13.7982	0.6308
103	34.4661	60.1862	140.4719	13.801	0.5443
104	34.4661	60.1862	140.3793	13.8008	0.7153
105	34.4661	60.1863	140.4894	13.7978	0.6294
106	34.4661	60.1862	140.4433	13.7958	0.6495
107	34.4661	60.1862	140.5212	13.7944	0.713
201	34.4681	60.1819	254.3005	9.3281	0.8927
...					
408	34.4683	60.1837	157.1076	6.6521	0.8533
409	34.4683	60.1838	157.0744	6.6589	0.8229
410	34.4683	60.1838	156.9917	6.6505	0.7766

Apart from location and kinematic data, the last column includes a reported confidence of classification provided by the Object Refinement Service. These values theoretically range from 1, meaning absolute certainty as to classification to the other extreme of 0. By convention values below .50 are not reported by the Object Refinement Service.

Table 1 provides an abridged representation of the 37 simulated reports. Each platform is associated with a specific unit. The identity of the parent unit is withheld as "ground truth" since the goal of aggregation, in this particular case, is to recover a proxy unit identifier. The "Ground Truth Identifier" contains as its first digit a parent unit and then the platform identifier within the unit is represented by the remaining digits.

Using a simplified variant of equation 2, the estimated number of iterations needed to provide a 90% confidence of no more than a .10 error in confidence assessment required 42 iterations. Table 2 summarizes the results of a single iteration in the application of the k-means algorithm to two samples obtained from the data.

Table 2: Single Iteration Paired Proximity Clustering

		First Sample	Second Sample	Ground Truth:
	Report Number	Assigned Cluster Labels	Assigned Cluster Labels	Platform within Unit
	1	2	1	101
	2	2	1	102
	3	2	1	103
	4	2	1	104
	5	2	1	105
	6	2	1	106
	7	2	1	107
	8	2	1	108
Excluded From A	9	2	0	109
	10	1	2	201
	11	1	4	202
	12	1	4	203
	13	1	4	204
	14	1	4	205
	15	1	4	206
	16	1	2	207
	17	1	4	208
	18	1	4	209
	19	1	2	210
	20	4	3	301
	21	4	3	302
	22	4	3	303
	23	4	3	304
Excluded From A	24	0	3	305
	25	4	3	306
	26	4	3	307
	27	4	3	308
	28	3	3	401
	29	3	3	402
	30	3	3	403
	31	3	3	404
	32	3	3	405
	33	3	3	406
	34	3	3	407
	35	3	3	408
	36	3	3	409
	37	3	3	410

Per the discussion in §3.2, table 2 illustrates that two subsamples are produced from R, sample 1 is produced by withholding the 24th report, while sample 2 is obtained by withholding the 9th observation, these indices being chosen at random. Where the entry in "Assigned Cluster" column depicts a value of "0", this signals that this report was withheld when that sample was provided to the aggregation algorithm. Neither of the reports is used in formation of the association matrix for this iteration.

A square association table is built using as its dimension the maximum number of clusters found in either sample -- here that value is 4. Cluster labels are used as the row and column cell identifiers in the association matrix and each cell counts the number of platform reports which relate the associated clusters. For example, when the first sample is input to the aggregation algorithm, the first report (ID=101) is assigned to a cluster labelled "2" while in the second sample, this same report is assigned to a cluster labelled "1". Accordingly, the cell in the second row and first column in the association table is incremented. Proceeding in this manner through the remainder of platform reports and tallying the counts, the association table is completed.

Table 3 is the association table derived from the single iteration shown in Table 2. The matrix is augmented with row and column totals to facilitate computation. Note that actual ground truth plays no role in the development of this table as the tableau relies solely upon the cluster labels assigned to sequential report numbers in paired samples.

Table 3: Single Iteration Association Matrix (Augmented)

$$A = \begin{pmatrix} 0 & 3 & 0 & 7 \\ 8 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 7 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 8 \\ 10 \\ 7 \end{pmatrix}$$

$$\begin{pmatrix} 8 & 3 & 17 & 7 \end{pmatrix}$$

This iteration's estimate of clarity is:

$$\gamma = \frac{\sum_1^m \sum_1^m \frac{c_{ij}^3}{O_i C_j}}{T} =$$

$$\frac{\frac{0^3}{(8)(10)} + \frac{3^3}{(3)(10)} + \frac{0^3}{(17)(10)} + \frac{7^3}{(7)(10)} + \frac{8^3}{(8)(8)} + \ldots + \frac{7^3}{(17)(7)} + \frac{0^3}{(7)(7)}}{35} = .64 \quad (7)$$

Over the 42 samples developed from Table 1, the average value of γ, $\widehat{\Gamma_K(R)}$, was .95 with an estimated standard deviation of this mean of $\sigma_{\widehat{\Gamma_K(R)}} = .0164$.

The geometric mean of "ORS Confidence" obtained from the unabridged version of Table 1 is $\hat{L}.= 6790.$. Factoring in this value, per equation 6, the confidence estimate of algorithm K applied to the dataset R is estimated to be $\widehat{\Gamma_K(R)}_{Adj} = .6451$.

These measures are not measures of accuracy or completeness, which do require ground truth, such as *Recall* and *Precision*.

Rather, this measure conveys to the user the aggregate uncertainty introduced by the totality of analytical processing. It encapsulates the variability introduced during data alignment, object refinement, as well as the current situation assessment processing. Although this value does not accurately portray the effectiveness of situation awareness processing itself, user's are best served by presenting the truth with the caveats as discerned, else intelligence tools stand in jeopardy of losing the greater trust. That said, maintaining Situation Awareness confidence measures do provide added value to a Process Refinement Service that is monitoring total system effectiveness. In this example, a diagnostic and optimizing service, looking at the intermediate results for $\widehat{\Gamma_K(R)}$, would be less inclined to focus on improving the Situation Awareness process but rather address the low confidence scores arising from Object Refinement and determine whether there is a flaw in that process or if ORS uncertainty is induced in prior processing such as erratic normalization or other data alignment processing.

4 Future Work

The previous material illustrates methods for safeguarding the users' trust by maintaining a global view of the factors that may introduce uncertainty into representations of the battlespace. Although not included in the example of §3.5, there are additional factors that merit attention in tempering the expression of confidence provided to a user.

4.1 Pedigree and Temporal Considerations

Although the exact meaning of source based pedigree is still maturing within the community, concepts of pedigree usually embody some notion of trust or confidence. As previously noted by others[11]. *"Because data may be useful but less precise, timely, or reliable than desirable, they ought to come with information on their reliability and methods (facts or algorithms) for authentication. After all, sensors err; analysts, human or automated, may be illogical, self-serving, premature, or simply wrong. Source-based pedigree is one option for both raw and processed estimates. Some estimates may be countersigned or otherwise vouched for. Command rules must also be authenticated at correct levels. Uncertainty must be correctly presented."* Different sources of information are either assigned a-priori confidence estimates by analysts or by the accumulated measures of trusts that have been tabulated over time, or possibly a mixture of these methods.

Once source confidence factors that are related to pedigree are introduced, their management and combination becomes an issue since multiple sources may report either similar or discordant findings. The management of discordant and reinforcing evidence may be addressed through several approaches such as fuzzy logic (which subsumes probability theory), Dempster-Shaefer evidentiary theory, Certainty Factors, or other approaches [12]

4.2 Temporal Considerations

It is routinely observed that the information value of intelligence observations decays with the passage of time. This is most often seen in the growth of areas of location uncertainty regarding an entities position or movement.

However, temporal based adjustments to the accuracy of other state estimates (e.g. activity) are encountered less frequently. Some researchers have considered this topic but requirements to address this issue are frequently lacking in systems under development and additional work is warranted. Some approaches illustrating both linear and exponential decay of information are referenced by Gacy and Dahn [13] and in their citations.

References

[1] E. Blasch, I. Kadar, J. Salerno, M. Kokar, S. Das, G. Powell, D. Corkill, and E. Ruspini, "Issues and Challenges of Knowledge Representation and Reasoning Methods in Situation Assessment (Level 2) fusion", Journal Of Advances In Information Fusion Vol. 1, No. 2 December 2006.

[2] Llinas, James (April, 1998). *Studies And Analyses Of Aided Adversarial Decision-Making Phase 2: Research On Human Trust In Automation*. Center for Multi-Source Information Fusion.

[3] United States. Naval Surface Warfare Center. Test Report: Correlation and Tracking algorithms to be used in the Advanced Tomahawk Weapon Control System. Port Hueneme, CA: Naval Surface Warfare Center, Port Hueneme, 1994.

[4] Palmer, J. (March 1992). *A Measure of Effectiveness for Selecting a Correlation Algorithm (WS:TWCS-92-005),* Austin, TX: Lockheed Missiles and Space Co.

[5] Halkidi, M., Batistakis, B. & Vazirgiannis M. (2002). *Cluster Validity Methods: Part I*. SIGMOD Record, 31(2), 40-45.

[6] Halkidi, M., Batistakis, B. & Vazirgiannis M. (2002). *Cluster Validity Methods: Part II*. SIGMOD Record, 31(3), 19-27.

[7] Theodoridis, Sergios, & Koutroumbas, Konstantinos. (2009). *Pattern Recognition*. Fourth Ed. Academic Press. 866 -87.

[8] Anderberg, M. R. Cluster Analysis for Applications. 1973. Academic Press, London. P 123

[9] Snedecor, G., & Cochran, W. (1967). *Statistical Methods*. Ames, Iowa: Iowa State University Press. P 210.

[10] Confidence Intervals (Sec. 7.2.4.1). (2006). Engineering Statistics Handbook. Retrieved (2010, April 6)http://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm (2 of 9) [5/1/2006 10:38:37 AM]

[11] Libicki , Martin. "Illuminating Tomorrow's War." CIAO. 1999. Institute for National Strategic Studies. 21 Jun 2006. http://www.ciaonet.org/wps/lim01/lim01.html.

[12] Klir, George, and Bo Yuan. Fuzzy Sets and Fuzzy Logic, Theory and Applications. 1st ed. Upper Saddle River: Prentice Hall, 1995.

[13] Gacy, A., & Dahn, D. (2010). An Information Taxonomy Combining Military Course Of Action. The Society For Modeling & Simulation International. Retrieved (2010, February 11) from www.scs.org/scsarchive/getDoc.cfm?id=1457