Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)
Prague, Czech Republic – July 28- 30, 2022
Paper No. 117
DOI: 10.11159/icsta22.117

A Theoretical Formulism for Evidential Reasoning and Logic Based Bias Reduction in Geo-Intelligence Processing

Nicholas V. Scott[1]
[1]Riverside Research Institute
2640 Hibiscus Way, Beavercreek, Ohio 45431, USA
nscott@riversideresearch.org

***Abstract* -** Geo-intelligence processing is strongly based on the need to bring together analytical viewpoints from multiple members comprising a geo-intelligence team so that unified answers to problems can be provided to leadership responsible for decision making. A three-tier evidential reasoning formulism is proposed and explained embodying a guide for the statistical/cognitive processing of geo-intelligence sensor information to facilitate this aim. The first tier comprises computational modeling used in conjunction with informal logic-based bias reduction by a multiple analyst team to interpret geo-intelligence information and create geo-intelligence reports. In the second tier, Bayesian belief networks over distinct provinces under geo-intelligence analytical investigation are created by each analyst through the amalgamation of statistical information provided by geo-intelligence reports. Bayesian belief network (BBN) results coupled with ancillary intelligence and analyst beliefs provide a set of propositions and probability masses summarizing the state of each province analyzed by each team member. The BBN state levels denote the three conditions of lack of nefarious substance presence, probable nefarious substance presence, and definite nefarious substance presence and are taken to be related, via a one-to-one mapping, directly to a new set of decision-based propositions – lack of adversary attack, probable adversary attack, and definite adversary attack. In the third tier, team member probability masses associated with these propositions, along with conjunctive and disjunctive combinations, are gradually amalgamated using Dezert-Smarandache (DS) evidential theory. A numerical example demonstrates the mechanics of the third-tier information fusion process which takes into account logical paradoxes and results in a single virtual analyst probability mass distribution associated with the geo-intelligence information amalgamation problem.

***Keywords*:** geo-intelligence reports, proposition field, Bayesian belief network, probability mass distribution, information amalgamation, Dezer-Smarandache evidential theory

1. Introduction

The objectives of many geo-intelligence organizations are the improvement of geo-intelligence information processing through information fusion and knowledge generation in addition to the facilitation of the assessment of environmental state and human activities. To support this, efforts in modulating how geo-intelligence processing is performed have focused on innovative ways to improve the computational and cognitive techniques used in the interpretation of geo-intelligence information. Past work on information fusion and knowledge generation in support of aquatic geo-intelligence demonstrated how fuzzy-logic and Dempster-Shafer (DS) evidential theory could be used to arrive at rigorous propositional assessments that ultimately stem from the analysis of captured geo-intelligence data/information [1,2]. While informative and a good first-step, the work failed to conceptually take into account *simultaneously* two very important aspects of geo-intelligence analysis crucial to uncertainty reduction in geo-intelligence processing. These are the incorporation of geo-intelligence-based bias *and* variance analysis. A multiple tier geo-intelligence processing formulism is proposed that addresses both bias and variance concurrently allowing for improved comprehension in geo-intelligence analysis. The objective of this short note is to explain how such a formulism, which is an amalgamation of three separate parts, would theoretically but practically work. Since much of geo-intelligence research is classified and lies outside the perimeter of what can be freely documented and presented, a heuristic-based explanation of the statistical, evidential, and logical analytical pipeline is only possible. Future work is based on the implementation of the formulism on real data.

2. Problem and Overall Approach

A fictitious geo-intelligence model problem for the comprehensive assessment of whether enemy forces in a specific region are going to attack, not attack, or retreat is addressed in this formulism whose theoretical structure is expounded upon. The state assessment requires first the diagnosis of the overall presence of nefarious substances for the region consisting of four provinces. The region yielding the remotely sensed geo-intelligence information is shown in Figure 1. Two provinces in the north and two provinces in the south are shown where information is examined by four different analysts comprising a geo-intelligence team. The provinces are further sub-divided into sub-regions where geo-intelligence information has been captured. The imagery acquired in the north consists of remotely sensed hyperspectral imagery (HSI) and infrared thermal imagery (IR). The imagery acquired in the south consists of remotely sensed HSI and synthetic aperture radar (SAR) imagery. Figure 2 provides a schematic of the overall process of how geo-intelligence information is gradually combined via a multiple tier pyramid shaped computational/cognitive formulism to yield a robust assessment of adversarial or enemy state. All tiers contain structures which have important aspects which function to address cognitive analyst bias and information variance.

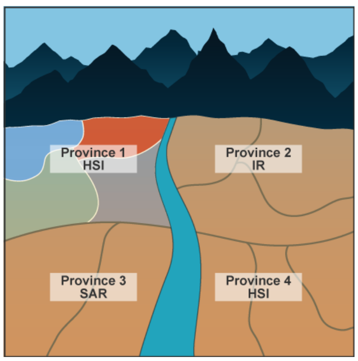


Figure 1. Spatial area over which remotely sensed imagery is acquired. Four provinces with spatial sub-regions shown.

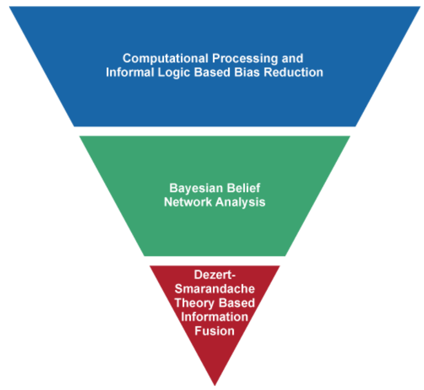


Figure 2. Information fusion formulism consisting of computational processing and informal logic-based bias reduction, Bayesian belief network analysis, and Dezert-Smarandache evidential theory.

The first tier of the pyramid formulism is where all analysts of a geo-intelligence analysis team operate individually to process incoming geo-intelligence information. Data is accepted by analysts and is processed using computational processing tools to assess the probabilistic presence of nefarious substances. Included in this analysis is informal logical analysis of the multiple propositions about nefarious substance presence which is then placed into a geo-intelligence report. Informal logic bias reduction is applied to investigate whether the arguments given in support of different propositions posed by analysts are well-grounded. (The use of 'well- grounded' here is the classic logic definition of the exploration of the likely truth of the premises of a proposition [3]).

Geo-intelligence reports created by each analyst, which cover a specific province, are used to create Bayesian belief networks over the information sources distributed throughout each province. This second-tier processing functions as a way of achieving low variance estimates of nefarious substance state via exploitation of spatially distributed geo-intelligence information.

Whereas the characteristic aspect of the first and second tier is bias and spatial variance reduction, the dominant aspect of the third-tier analysis is based on achieving low information variance across different analysts. New propositions about the state of the adversary (as opposed to the presence of nefarious substances) are made. Probabilistic beliefs about the state of the adversary associated with each member of the geo-intelligence analysis team are amalgamated. In particular, the beliefs of team members are combined using a set-theory based logic which takes into account logical paradoxes which may arise in the proposition consolidation process. The merging of information gradually groups the results of the individual analyses of team members until a super or 'virtual' analyst is created which represents the gestalt of the geo-intelligence processing. The super or 'virtual' analyst result embodies the group perspective of all team members with the final result being a low bias and low variance solution of the geo-intelligence problem. It is this result which is used in decision making by geo-intelligence leadership.

The sections to follow provide a qualitative and quantitative explanation of the geo-intelligence processing carried out at each stage displayed in Figure 2. The work provides a heuristic conception of how geo-intelligence processing could occur based on the implementation of logic, Bayesian belief networks, and Dezert-Smarandache **(**DSm) evidential theory. Numerical calculations are additionally given which illustrate the mechanics of the information amalgamation carried out in the third-tier processing.

3. First Tier Processing Methodology – Cognitive Processing and Informal Logic Based Bias Reduction

The pyramidal shaped processing formulism shown in Figure 2 begins with the computational and cognitive analysis of different types of imagery captured in the different provinces. This starting point, depicted in Figure 3, shows the geo-intelligence processing of HSI, IR, and SAR imagery data by four different analysts responsible for each of the four different provinces. The data processed by each analyst emanates from the different sub-regions within a province shown in Figure 1. The geo-intelligence processing by each analyst is aimed at assessing the probability of the presence of nefarious substances. Three mutually exclusive possibilities $\theta = \{\theta_1, \theta_2, \theta_3\}$ exist which consist of definitive presence, plausible presence, and no presence of nefarious substances respectively. Specific details surrounding the actual computational modeling tools and algorithms aimed at quantifying the state of nefarious substances are beyond the scope of this paper and are not given here.

In addition to the computational uncertainty emanating from the computational analysis, there is a cognitive uncertainty attributed to the way in which an analyst perceives computationally processed geo-intelligence information. An analyst's preconceived ideas, previous experience, and other ancillary intelligence particular to the geo-intelligence problem influence estimates of the certainty surrounding the three propositions. This cognitive uncertainty translates into variability that is superimposed on top of the computational variance.

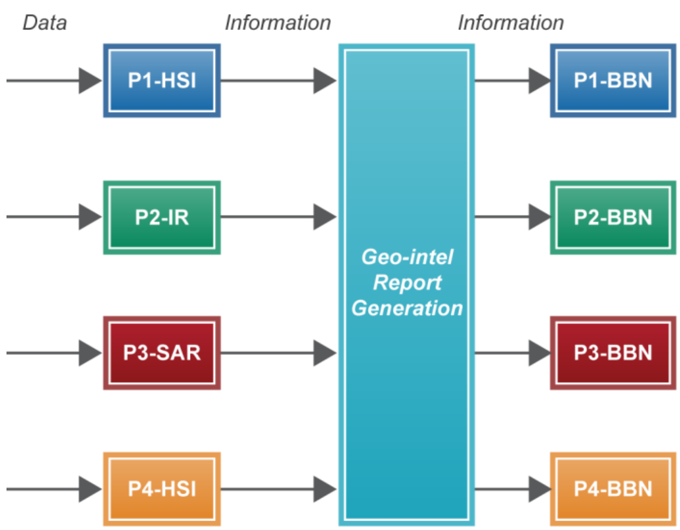


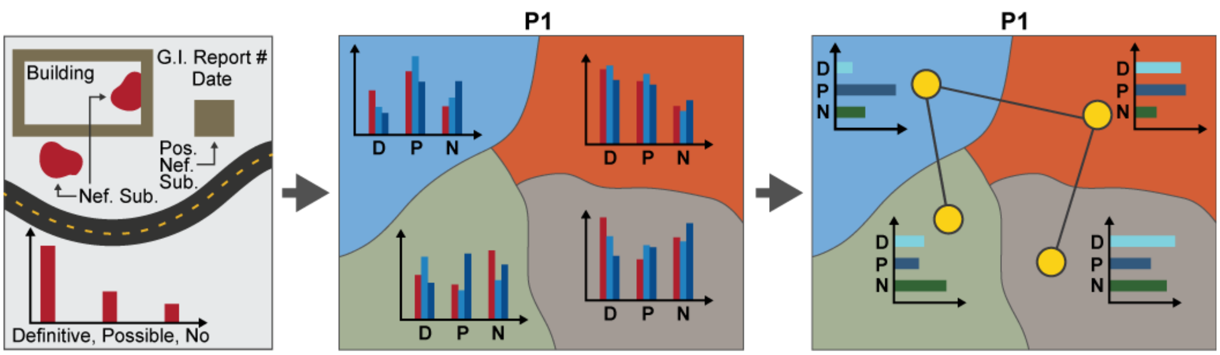
Figure 3. First and second tier geo-intelligence processing. Provinces designated P1-P4.

Informal logical bias analysis is used to rigorously assess each proposition to reduce cognitive uncertainty associated with the computational results. This is performed before geo-intelligence reports are produced. While not all bias can be eliminated, there are some traditionally ones which should be addressed. For example, the circular reasoning fallacy and ad hominem reasoning fallacy [3] are two potent informal logic biases endemic to geo-intelligence which can profoundly influence an analyst's final assessment of the relative probabilities for a set of propositions.

Constructed geo-intelligence reports of each analyst contain a histogram of propositions which quantify the relative probabilities associated with each of the three propositions. Figure 4a depicts a sample geo-intelligence report for one analyst. All geo-intelligence reports, irrespective of the type of geo-intelligence sensor information used to produce them, fundamentally have a similar form. This consists of an image of the area containing the substance under scrutiny, and labels which highlight the presence of structures of interest. It also contains the date and time of capture, and a histogram, appearing in left lower corner of the geo-intelligence report, quantifying the probability of whether a nefarious substance is definitively, possibly, and not possibly present. It is noted again that the uncertainty inherent in the histogram probability estimates for the propositions reflects the variance inherent to the computational modeling algorithms for the different geo-intelligence information types *and* the uncertainty afforded by an analyst's bias emanating from the cognitive analysis of the geo-intelligence information.

4. Second Tier Processing Methodology – Bayesian Belief Network Analysis of Provincial Geo-Intelligence Information

The second-tier processing scheme is the same for all analysts and is illustrated here for the HSI analyst of province 1. The processing of geo-intelligence information emanating from the sub-regions of a particular province yields multiple geo-intelligence reports. Over a finite time interval, these reports yield a distribution of histograms detailing the probabilistic state of the presence of nefarious substances for the different sub-regions within province 1. This is shown in Figure 4b. In other words, within each sub-region of a province and for each proposition there is an array of frequency values obtained over a set amount of time which stems from the gleaning of information from a multitude of geo-intelligence reports accrued over the same time period. Frequency histogram values associated with each proposition in a sub-region are averaged over the set time interval that they were acquired to form average proposition histograms over the propositions for the time interval. This is shown in Figure 4c. Formation of conditional probability distribution tables, detailing the statistical relationship between sub-regions, can be distilled from a series of average proposition histograms for each time interval taken from each sub- region. This is done via computation of the covariance existing between the sub-region information. These in turn yields the joint probabilistic information necessary for the construction of a Bayesian belief network (BBN) for that province.



a)

b)

c)

Figure 4. a) Geo-intelligence report emanating from first tier geo-intelligence processing. Nef. Sub. stands for nefarious substance. Pos. Nef. Sub. stands for possible nefarious substance. Definitive, Possible, and No stands for definitive nefarious substance presence, possible nefarious substance presence, and no nefarious substance presence respectively. b) Histograms delineate average relative frequency for each of the categories of D, P, and N which stand for definite, possible, and no nefarious substance presence. c) Schematic of a BBN for a province. Nodes situated at sub-regions and edges indicate covariance relationships between sub-regions.

The nodes and edges, comprising the probabilistic graphical network structure of the BBN, allow for understanding the statistical interrelationships between the different sub-regions of a province [4, 5]. Each node of the network shown in Figure 4 has a probability distribution over the same three propositions concerning nefarious substance presence $\theta = \{\theta_1, \theta_2, \theta_3\}$. The propositional fields for any node, irrespective of the sub-region and province it emanates from, are equivalent. However, the particular proposition values for the nodes of the BBNs for each province and how they change may differ structurally. This is because the BBN quantifies the joint probability distribution relationships that exists over the sub-regions within a province which may differ from province to province. Specific agreed upon nodal evidence can be applied to the BBN via nodal instantiations [5] at each of the four nodes within each province. Each member of the geo-intelligence team then uses their respective BBN for their province to produce statistical diagnostics. The universality of the proposition field across all information nodes from any province is what allows for amalgamating statistical diagnostic results across provinces in the third-tier processing which is the next stage.

5. Third Tier Processing Methodology – Dezert–Smarandache Theory Based Information Fusion

5.1 Theory and Geo-Intelligence Implementation

Information fusion in the third tier begins with the formulation of a new geo-intelligence question that each analyst seeks to answer in support of assisting geo-intelligence leadership decision making. The question is whether the enemy or adversary is going to attack, not attack, or retreat based on the understanding and insight afforded to each analyst through the analysis over their respective province. The probabilistic solution assumes a one-to-one mapping of the old proposition field $\theta = \{\theta_1, \theta_2, \theta_3\}$ to a new proposition field $\beta = \{\beta_1\ \beta_2\ \beta_3\}$. In other words, the assumed mapping, used as the foundation for addressing the new geo-intelligence question, is based on $\theta_1 \rightarrow \beta_1$, $\theta_2 \rightarrow \beta_2$, and $\theta_3 \rightarrow \beta_3$. The probabilistic solution to the new geo-intelligence question is addressed by each analyst using the probabilistic information furnished from their respective BBN, combined with probabilistic information furnished from personal beliefs, perspectives, and ancillary intelligence. Through this cognitive fusion, each analyst furnishes a new set of probabilities or beliefs that are pertinent to addressing the geo-intelligence question. Note that the new propositions formulated by each analyst in addressing the new geo-intelligence question not only can emanate from the three-element set $\beta = \{\beta_1\ \beta_2\ \beta_3\}$, but also from conjunctive and disjunctive combinations of these elements. In addition, the number of new propositions is, in principle, not limited in number.

The sets of propositions created by each analyst have probability masses which are merged to form an overall assessment of enemy state over all four provinces. This fusion is attained via DSm evidential theory. It allows for the combination of probability masses across different analysts or provinces giving, in the end, a single solution with represents the beliefs and perspectives of the entire geo-intelligence team. The application of DSm evidential theory represents an extension of Dempster-Shafer evidential theory, where the property of exhausivity of propositions is still present, but where mutual exclusivity is dispensed with [6]. In DSm evidential theory conjunctive propositional relationships are allowed with the rule of combination of probability mass being the following.

$$m(C) = \sum_{A \cap B = C} \left[m_N(A) m_M(B) \right] \tag{1}$$

Here A and B are members of the hyperset D^β which consists of an exhaustive frame of discernment β that is closed under operations of disjunction and conjunction [6, 7]. The quantities $m_N(A)$ and $m_M(B)$ are probability mass assignments which are a function of the propositions A and B for analyst N and M. The quantity C is a new proposition member created under the operations of conjunction $\cap$ or disjunction $\cup$ lying in the hyperset D^β. The probability mass emanating from the combination of information from two different analysts, analyst N and M, is defined by the multiplication of probability mass values represented by $m_N(A)$ and $m_M(B)$ [7, 8]. A numerical example is given below to illustrate the mechanics of the theory.

5.2 Numerical Simulation of DSm Evidential Reasoning

A decision maker engages in DSm reasoning using the information gathered from the array of four analysts in order to gain an overall understanding of enemy state pervading the four provinces. Probability mass components or mass distributions for analyst 1 and 2 are shown in the left blue vertical column and horizontal blue column respectively of the multiple box table in Figure 5a. Probability mass components for analyst 3 and 4 are shown in the same respective positions in Figure 5b. The probability mass distributions for analyst 1 and 2 are combined first followed by that of analyst 3 and 4. Note the presence of conjunctive and disjunctive propositions which seek to quantify the fuzziness inherent in the geo-intelligence problem. The explicit method by which the probability mass is assigned initially by analysts depends on the geo-intelligence team procedures and is an issue not taken up here. Probability mass allocated to the conjunctive intersection of propositions appear as elements inside the matrix table shown in Figures 5a-c. The results of the consolidation for analysts 1 and 2 are shown in the horizontally long rectangle near the bottom in Figure 5a. The results of the consolidation for analysts 3 and 4 are shown in the same respective position in Figure 5b. In both cases the proposition with the largest probability is β_1, or the enemy is going to attack. However, the amalgamation of the propositional probabilities from Figure 5a and 5b to yield the virtual analyst result, shown in Figure 5c, demonstrates that the proposition $\beta_1 \cap \beta_2$ has the highest probability mass at 47%. High probability associated with attack and also not attack possibly suggests a feint maneuver by the enemy in this model problem. The proposition with the next highest probability mass is that of enemy attack, β_1 at 39 %. These results may suggest that if a definitive proposition emanating just from the original three-element set $\beta = \{\beta_1\ \beta_2\ \beta_3\}$ is necessary, then β_1 is the optimal single solution. A decision based on this proposition can subsequently be made by leadership.

$m_1(\beta_1)=0.6$	$m_c(\beta_1)=0.24$	$m_c(\beta_1\cap\beta_2)=0.12$	$m_c(\beta_1)=0.18$	$m_c(\beta_1)=0.06$
$m_1(\beta_2)=0.2$	$m_c(\beta_1\cap\beta_2)=0.08$	$m_c(\beta_2)=0.04$	$m_c(\beta_2\cap(\beta_1\cup\beta_3))=0.06$	$m_c(\beta_2)=0.02$
$m_1(\Theta)=0.2$	$m_c(\beta_1)=0.08$	$m_c(\beta_2)=0.04$	$m_c(\beta_1\cup\beta_3)=0.06$	$m_c(\Theta)=0.02$
	$m_2(\beta_1)=0.4$	$m_2(\beta_2)=0.2$	$m_2(\beta_1\cup\beta_3)=0.3$	$m_2(\Theta)=0.1$

$m_c(\beta_1)=0.56$	$m_c(\beta_1)=0.10$	$m_c(\beta_1\cap\beta_2)=0.2$	$m_c(\beta_1\cup\beta_3)=0.12$	$m_c(\Theta)=0.02$

Figure 5a: DSm information amalgamation for analyst 1 and analyst 2. Probability mass resulting from the information amalgamation shown in yellow rectangle.

$m_3(\beta_1)=0.6$	$m'_c(\beta_1)=0.42$	$m'_c(\beta_1\cap\beta_2)=0.12$	$m'_c(\beta_1\cap(\beta_1\cap\beta_3))=0.06$
$m_3(\beta_2)=0.2$	$m'_c(\beta_1\cap\beta_2)=0.14$	$m'_c(\beta_2)=0.04$	$m'_c(\beta_2\cap(\beta_1\cap\beta_3))=0.02$
$m_3(\beta_1\cup\beta_3)=0.1$	$m'_c(\beta_1)=0.07$	$m'_c(\beta_2\cap(\beta_1\cup\beta_3))=0.02$	$m'_c[(\beta_1\cup\beta_3)\cap(\beta_1\cap\beta_3)]=0.01$
$m_3(\Theta)=0.1$	$m'_c(\beta_1)=0.07$	$m'_c(\beta_2)=0.02$	$m'_c(\beta_1\cap\beta_3)=0.01$
	$m_4(\beta_1)=0.7$	$m_4(\beta_2)=0.2$	$m_4(\beta_1\cap\beta_3)=0.1$

$m'_c(\beta_1)=0.56$	$m'_c(\beta_2)=0.06$	$m'_c(\beta_1\cap\beta_2)=0.2$	$m'_c(\beta_1\cap\beta_3)=0.1$	$m'_c[(\beta_1\cup\beta_3)\cap\beta_2]=0.02$

Figure 5b: DSm information amalgamation for analyst 3 and analyst 4. Probability mass resulting from the information amalgamation shown in yellow rectangle.

$m_c(\beta_1)=0.56$	$m_c(\beta_1)=0.3136$	$m_c(\beta_1\cap\beta_2)=0.0336$	$m_c[\beta_2\cap(\beta_1\cap\beta_2)]=$ 0.1456	$m_c[\beta_1\cap(\beta_1\cap\beta_3)]=$ 0.0056	$m_c[\beta_1\cap(\beta_2\cap(\beta_1\cup\beta_3))]=$ 0.0112
$m_c(\beta_2)=0.10$	$m_c(\beta_1\cap\beta_2)=0.056$	$m_c(\beta_2)=0.006$	$m_c[\beta_2\cap(\beta_1\cap\beta_2)]=$ 0.0026	$m_c[\beta_2\cap(\beta_1\cap\beta_3)]=0.01$	$m_c[\beta_2\cap(\beta_2\cap(\beta_1\cup\beta_3))]=$ 0.002
$m_c(\beta_1\cap\beta_2)=0.2$	$m_c[\beta_1\cap(\beta_1\cap\beta_2)]=0.112$	$m_c[\beta_2\cap(\beta_1\cap\beta_2)]=0.012$	$m_c(\beta_1\cap\beta_2)=0.052$	$m_c[(\beta_1\cap\beta_2)\cap(\beta_1\cap\beta_3)]=$ 0.02	$m_c[(\beta_1\cap\beta_2)\cap(\beta_2\cap(\beta_1\cup\beta_3))]=$ 0.004
$m_c(\beta_1\cup\beta_3)=0.12$	$m_c(\beta_1)=0.0672$	$m_c[\beta_2\cap(\beta_1\cup\beta_2)]=$ 0.0072	$m_c[(\beta_2\cup\beta_3)\cap(\beta_1\cap\beta_2)]=$ 0.0312	$m_c[(\beta_2\cup\beta_3)\cap(\beta_1\cap\beta_3)]=$ 0.1456	$m_c[(\beta_1\cap\beta_3)\cap(\beta_2\cap(\beta_1\cup\beta_3))]=$ 0.0024
$m_c(\Theta)=0.02$	$m_c(\beta_1)=0.0112$	$m_c(\beta_1)=0.0112$	$m_c(\beta_1\cap\beta_2)=0.0052$	$m_c(\beta_1\cap\beta_3)=0.002$	$m_c[\beta_2\cap(\beta_1\cup\beta_3)]=$ 0.0004
	$m_c(\beta_1)=0.56$	$m_c(\beta_2)=0.06$	$m_c(\beta_1\cap\beta_2)=0.26$	$m_c(\beta_1\cap\beta_3)=0.1$	$m_c[\beta_2\cap(\beta_1\cup\beta_3)]=0.02$

$m''_c(\beta_1)=0.3920$	$m''_c(\beta_2)=0.0072$	$m''_c(\beta_1\cap\beta_2)=0.4736$	$m''_c(\beta_1\cap\beta_3)=0.1$	$m''_c(\beta_1\cup\beta_3)=0.0272$

Figure 5c: Total DSm information amalgamation using the results from Figures 5a and 5b for all analysts. Inner matrix elements give values for the intersection of probability masses for all analysts. Total probability mass resulting from the information amalgamation outlined in yellow rectangle.

DOI: 10.11159/icsta22.117

6. Conclusions

This small technical note illustrates a method for uncertainty reduction in which computational analytics are fused with a logical framework to address uncertainty reduction. Computational variance, individual analyst cognitive bias, and geo-intelligence team variance are all used in a formulism created to address problems where absolute certainty cannot be assigned to propositions. It is noted that the DSm evidential theory is particularly helpful in dealing with situations where beliefs of different team members are vastly different making logical paradoxes likely [6]. Since no real data has been used in this work, a future step is to test the formulism to see if realistic and understandable results occur.

7. Acknowledgements and Disclaimer

The author gratefully acknowledges past colleagues and summer interns who assisted in this work. The views and research discussed in this work represent the author's personal viewpoint and in no way represent any agency or organization.

References

[1] N. V. Scott, A. Rutherford, and R. Bartell, "Fuzzy Logic/Dempster-Shafer Based Information Fusion Formulism for Land-Marine Decision Analysis," INFORMS Annual Meeting, Nashville, TN, 2016.
[2] N. V. Scott, "Fuzzy Logic/Dempster-Shafer Based Information Fusion Formulism for Land-Marine Decision Analysis," Military Operations Research Symposium, Colorado Springs, CO, 2019.
[3] I. Copi, *Introduction to Logic*, MacMillan Publishing, 590 pp., 1953.
[4] Sucar, L. E., *Probabilistic Graphical Models: Principles and Applications*, Springer-Verlag, 2015, 253 pp.
[5] K. E. Korb, and A. E. Nicholson, *Bayesian Artificial Intelligence*, CRC Press, 463 pp, 2011.
[6] L. A. Klein, *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*, Society for Photo-Optical Instrumentation Engineers, 337 pp., 2010.
[7] J. Dezert, "Foundations for a New Theory of Plausible and Paradoxical Reasoning, Information and Security," An International Journal, vol. 9, no. 200, pp. 13-57, 2002.
[8] J. Dezert, J. and F. Smarandache, "An Introduction to DSmT for Information Fusion, New Mathematics and Natural Computation," vol. 8, no.3, pp. 343-359, 2012.