PROCEEDINGS OF SPIE

Signal Processing, Sensor Fusion, and Target Recognition XXI

Ivan Kadar
Editor

23–25 April 2012
Baltimore, Maryland, United States

Sponsored and Published by
SPIE

Volume 8392

Proceedings of SPIE, 0277-786X, v. 8392

Signal Processing, Sensor Fusion, and Target Recognition XXI, edited by Ivan Kadar,
Proc. of SPIE Vol. 8392, 839201 · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.970996

Please use the following format to cite material from this book:
Author(s), "Title of Paper," in *Signal Processing, Sensor Fusion, and Target Recognition XXI*, edited by Ivan Kadar, Proceedings of SPIE Vol. 8392 (SPIE, Bellingham, WA, 2012) Article CID Number.

ISSN 0277-786X
ISBN 9780819490704

Published by
SPIE
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone +1 360 676 3290 (Pacific Time) · Fax +1 360 647 1445
SPIE.org

Printed in the United States of America.

Publication of record for individual papers is online in the SPIE Digital Library.

SPIE Digital Library

SPIEDigitalLibrary.org

Paper Numbering: Proceedings of SPIE follow an e-First publication model, with papers published first online and then in print and on CD-ROM. Papers are published as they are submitted and meet publication criteria. A unique, consistent, permanent citation identifier (CID) number is assigned to each article at the time of the first publication. Utilization of CIDs allows articles to be fully citable as soon as they are published online, and connects the same identifier to all online, print, and electronic versions of the publication. SPIE uses a six-digit CID article numbering system in which:

- The first four digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B … 0Z, followed by 10-1Z, 20-2Z, etc.

The CID number appears on each page of the manuscript. The complete citation is used on the first page, and an abbreviated version on subsequent pages. Numbers in the index correspond to the last two digits of the six-digit CID number.

Contents

Conference Committee

Symposium Chair

Kevin P. Meiners, Office of the Secretary of Defense (United States)

Symposium Cochair

Kenneth R. Israel, Lockheed Martin Corporation (United States)

Conference Chair

Ivan Kadar, Interlink Systems Sciences, Inc. (United States)

Program Committee

Mark G. Alford, Air Force Research Laboratory (United States)
William D. Blair, Georgia Tech Research Institute (United States)
Erik P. Blasch, Air Force Research Laboratory (United States)
Mark J. Carlotto, General Dynamics Advanced Information Systems (United States)
Kuo-Chu Chang, George Mason University (United States)
Chee-Yee Chong, BAE Systems (United States)
Marvin N. Cohen, Georgia Tech Research Institute (United States)
Frederick E. Daum, Raytheon Company (United States)
Mohammad Farooq, AA Scientific Consultants Inc. (Canada)
Charles W. Glover, Oak Ridge National Laboratory (United States)
I. R. Goodman, Consultant (United States)
Lynne L. Grewe, California State University, East Bay (United States)
David L. Hall, The Pennsylvania State University (United States)
Michael L. Hinman, Air Force Research Laboratory (United States)
Kenneth Hintz, George Mason University (United States)
Jon S. Jones, Air Force Research Laboratory (United States)
Thiagalingam Kirubarajan, McMaster University (Canada)
Martin E. Liggins II, The MITRE Corporation (United States)
James Llinas, University at Buffalo (United States)
Ronald P. Mahler, Lockheed Martin Maritime Systems & Sensors (United States)
Raj P. Malhotra, Air Force Research Laboratory (United States)
Alastair D. McAulay, Lehigh University (United States)
Raman K. Mehra, Scientific Systems Company, Inc. (United States)
Harley R. Myler, Lamar University (United States)
David Nicholson, BAE Systems (United Kingdom)
Les Novak, Scientific Systems Company, Inc. (United States)

John J. Salerno, Jr., Air Force Research Laboratory (United States)
Andrew G. Tescher, AGT Associates (United States)
Stelios C. A. Thomopoulos, National Center for Scientific Research Demokritos (Greece)
Wiley E. Thompson, New Mexico State University (United States)
Pierre Valin, Defence Research and Development Canada, Valcartier (Canada)

Session Chairs

1 Multisensor Fusion, Multitarget Tracking, and Resource Management I
Ivan Kadar, Interlink Systems Sciences, Inc. (United States)
Thiagalingam Kirubarajan, McMaster University (Canada)
Kenneth Hintz, George Mason University (United States)

2 Multisensor Fusion, Multitarget Tracking, and Resource Management II
Thiagalingam Kirubarajan, McMaster University (Canada)
Kenneth Hintz, George Mason University (United States)
Ivan Kadar, Interlink Systems Sciences, Inc. (United States)

3 Multisensor Fusion Methodologies and Applications I
Adel I. El-Fallah, Scientific Systems Company, Inc. (United States)

4 Multisensor Fusion Methodologies and Applications II
Michael L. Hinman, Air Force Research Laboratory (United States)
Chee-Yee Chong, BAE Systems (United States)
Ivan Kadar, Interlink Systems Sciences, Inc. (United States)

5 Multisensor Fusion Methodologies and Applications III
Erik P. Blasch, Air Force Research Laboratory (United States)
Chee-Yee Chong, BAE Systems (United States)
Michael L. Hinman, Air Force Research Laboratory (United States)
Kenneth Hintz, George Mason University (United States)

6 Multisensor Fusion Methodologies and Applications IV
Chee-Yee Chong, BAE Systems (United States)
Michael L. Hinman, Air Force Research Laboratory (United States)
Kenneth Hintz, George Mason University (United States)
Ivan Kadar, Interlink Systems Sciences, Inc. (United States)

7 Multisensor Fusion Methodologies and Applications V
Kenneth Hintz, George Mason University (United States)
Michael L. Hinman, Air Force Research Laboratory (United States)
Chee-Yee Chong, BAE Systems (United States)
Erik P. Blasch, Air Force Research Laboratory (United States)

8 Signal and Image Processing, and Information Fusion Applications I
Lynne L. Grewe, California State University, East Bay (United States)
Alastair D. McAulay, Lehigh University (United States)
Mark G. Alford, Air Force Research Laboratory (United States)

9 Signal and Image Processing, and Information Fusion Applications II
Alastair D. McAulay, Lehigh University (United States)
Lynne L. Grewe, California State University, East Bay (United States)
Mark G. Alford, Air Force Research Laboratory (United States)

10 Signal and Image Processing, and Information Fusion Applications III
Mark G. Alford, Air Force Research Laboratory (United States)
Alastair D. McAulay, Lehigh University (United States)
Lynne L. Grewe, California State University, East Bay (United States)

Invited Panel Discussion
Real-World Issues and Challenges in Social/Cultural Modeling with Application to Information Fusion

Organizer

John Salerno, Air Force Research Lab., Air Force Research Lab., Rome Research Site.

Co-Organizer

Ivan Kadar, Interlink Systems Sciences, Inc.

Moderators

Ivan Kadar, Interlink Systems Sciences, Inc.

John Salerno, Air Force Research Lab., Air Force Research Lab., Rome Research Site.

April 24, 2012

SPIE Conference 8392

"Signal Processing, Sensor Fusion and Target Recognition XXI"

Baltimore, MD April 23-25, 2012

Invited Panel Discussion

(Participants)

Dr. Erik Blasch, Air Force Research Lab., Rome Research Site

Dr. Mica Endsley, SA Technologies, USA

Dr. Laurie Fenstermacher, Air Force Research Lab., Human Effectiveness Directorate

Professor Lynne Grewe, California State Univ., East Bay

Dr. Ivan Kadar, Interlink Systems Sciences, Inc., USA

Dr. John Salerno, Air Force Research Lab., Rome Research Site

Professor Shanchieh Jay Yang, Rochester Institute of Technology

Invited Panel Discussion

(Topics)

Behavior Modeling for Sensemaking and Fusion
Dr. Laurie Fenstermacher, Air Force Research Lab., Human Effectiveness Directorate

Exploitation of Social Behavioral Modeling Understanding for Information Fusion
Drs. Erik Blasch and Guna Seetharaman, Air Force Research Lab., Rome Research Site

The Interest Graph and Beyond – Social Modeling and Information Fusion
Professor Lynne Grewe, California State Univ., East Bay

Issues and Challenges in Intent Modeling in Social/Cultural Networking Domain
Dr. Ivan Kadar, Interlink Systems Sciences, Inc.

Invited Panel Discussion

(Topics, Cont'd)

Characterizing Large-Scale Attack Behavior for Predictive Cyber Situation Awareness
Professor Shanchieh Jay Yang, Rochester Institute of Technology

Building Robust Situation Models for Higher Level Information Fusion & Social-Cultural Modeling
Dr. Mica Endsley, SA Technologies

Modeling Human Behavior within the Fusion Process
Dr. John Salerno, Air Force Research Lab., Rome Research Site

THE AIR FORCE RESEARCH LABORATORY
LEAD | DISCOVER | DEVELOP | DELIVER

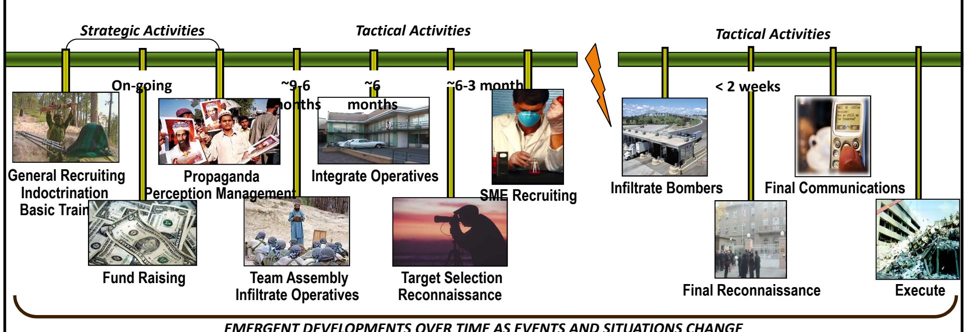
Behavior Modeling for Sensemaking and Fusion

24 April 2012

Laurie Fenstermacher
Air Force Research Laboratory
Human Effectiveness Directorate

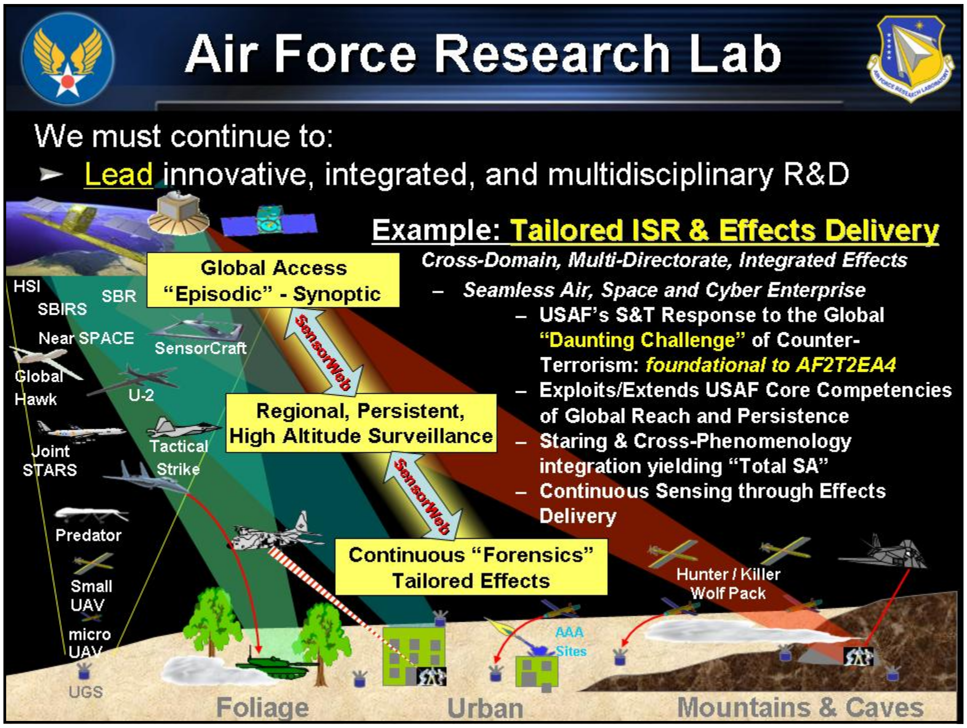
"Wicked" Problems: Operational Challenges

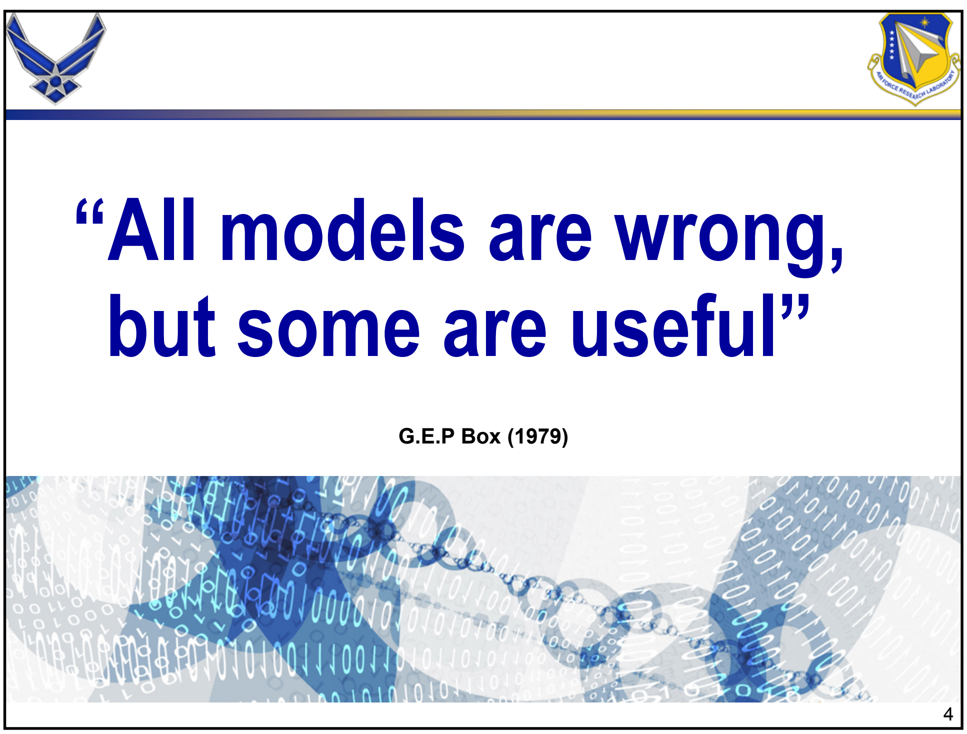
- Hostile events are rare and hidden within 'normal' city activity 'noise'
- Sophisticated adversary will conceal operations and intensions in 'plain sight'
- Anomalies may never appear or may appear too late to disrupt
- Adversary activities will change and adapt
  - *Strategic* adaptation occurs in response to and anticipation of US/Allied activity
  - *Tactical* adaptation supports flexibility, uses deception, misinformation, disinformation
  - *Emergent* properties evolve over time as actors/situations/environments change



Graphic Ref: Weinstein,"Warning Terms and Usage", 5 Nov 2007

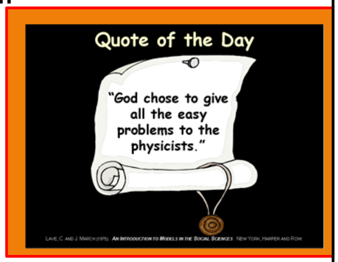
2


Air Force Research Lab
We must continue to:
➢ Lead innovative, integrated, and multidisciplinary R&D
Example: Tailored ISR & Effects Delivery
Cross-Domain, Multi-Directorate, Integrated Effects
– Seamless Air, Space and Cyber Enterprise
– USAF's S&T Response to the Global "Daunting Challenge" of Counter-Terrorism: foundational to AF2T2EA4
– Exploits/Extends USAF Core Competencies of Global Reach and Persistence
– Staring & Cross-Phenomenology integration yielding "Total SA"
– Continuous Sensing through Effects Delivery
Global Access
"Episodic" - Synoptic
SensorWeb
Regional, Persistent, High Altitude Surveillance
SensorWeb
Continuous "Forensics" Tailored Effects
HSI
SBIRS
SBR
Near SPACE
SensorCraft
Global Hawk
U-2
Joint STARS
Tactical Strike
Predator
Small UAV
micro UAV
UGS
Hunter / Killer Wolf Pack
AAA Sites
Foliage
Urban
Mountains & Caves


"All models are wrong, but some are useful"
G.E.P Box (1979)
4

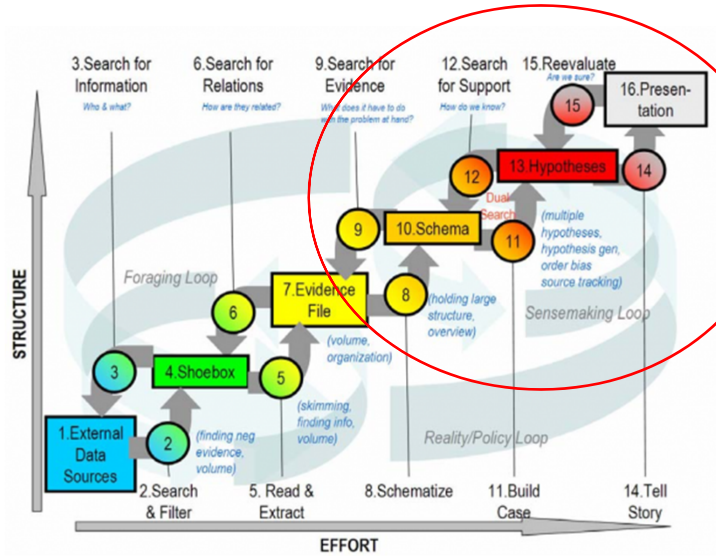
Socio-Cultural Modeling Benefits

- **Multi-scale representations of behaviors -- ranging from cognitive systems to individuals, groups, organizations, societies, nations, civilizations, global**
- **Models can clarify complex situations, aid decision making and bound the expected and outlier behaviors**
- **Modeling can capture dynamical, non-linear processes**
- **Models CAN address weaknesses in mental models by making assumptions explicit, helping calculate consequences of assumptions**
- **Models can provide a formal mechanism to attach evidence to models (relative likelihoods of outcomes) and/or enable the human-in-the-loop to link and assess evidence**
- **Models enable the exploration of data, "what-if" analysis (e.g., alteration of social network)**



5

Sensemaking for Analysis



Ref: Pirolli and Card, 2005

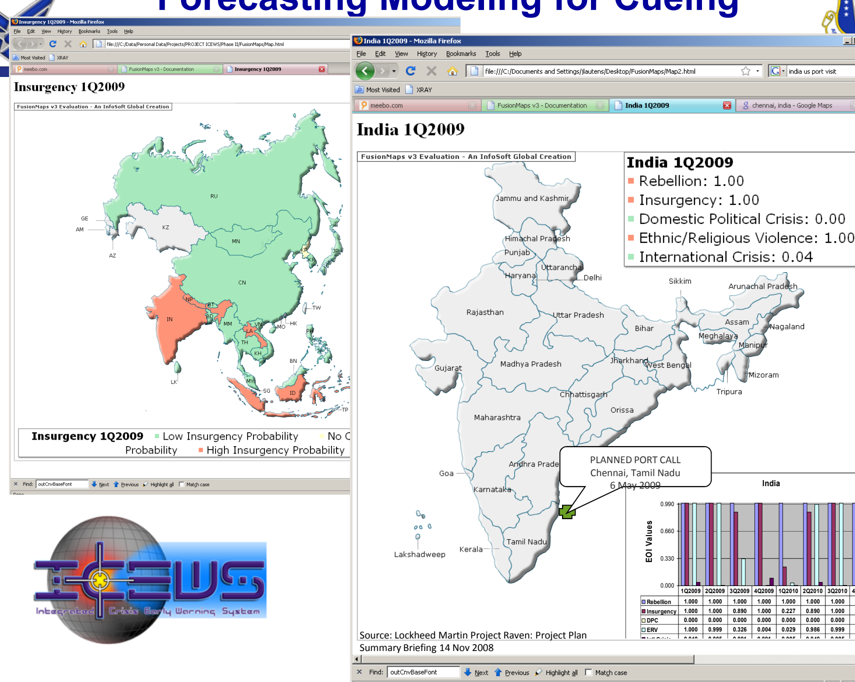
6

Socio-Cultural Modeling for Sensemaking and Fusion

- Modeling can cue data collection
  - Models for forecasting events of interest, advent/cessation of violence, etc. enable anticipatory Intelligence, Surveillance and Reconnaissance (ISR) and proactive shaping, action
  - Discourse analysis based on social identity, categorization provide leading indicators of action
- Modeling can facilitate the understanding and interpretation of data in a relevant context
  - Understanding how vigilance/precaution varies cross-culture
  - Understanding of normative behavior(s)
  - Provide indirect information (human/organizational behaviors and socio-cultural context)
- Modeling can enable sensemaking by informing the development of mental schema
  - Identifying entities, transactions for detection and tracking of networks
  - Human/vehicle behaviors can enable inferencing about static facilities (buildings, geographic areas)
  - Generative models can identify new indicators, signatures, hypotheses and assist in overcoming heuristics and biases

7

Forecasting Modeling for Cueing



8

Forecasting based on Discourse Analysis

- **Model based methodology and semi-automatic text analytic tool for cueing detects and helps interpret language related to social identity (in-group/out-group)**
- **Used operationally for forecasting by NASIC**

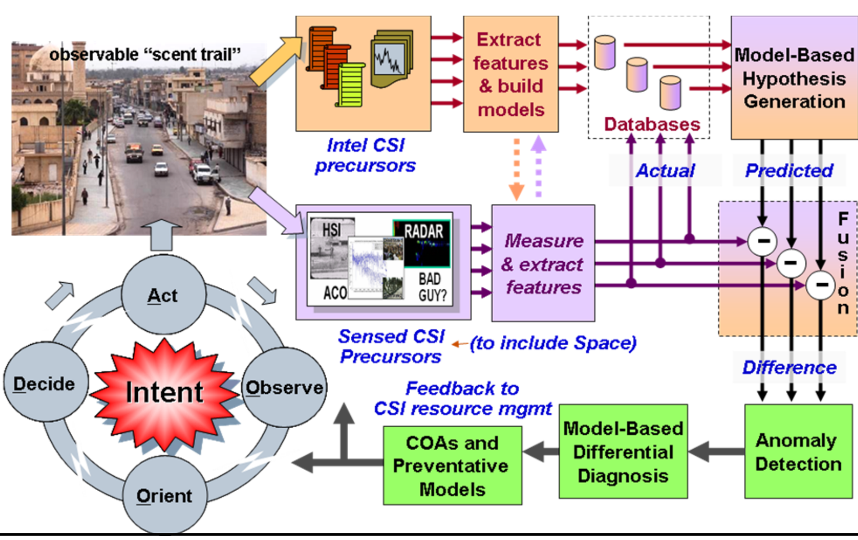
يزهو حزب الله بإنجازه. ويحق له ذلك. فقد وعد أن يحرر القنطار ووفى . وهو بهذا الإنجاز يستعيد ما خسره عربيا من شعبية بعد أن وجه سلاحه إلى الداخل ليحسم لصالحه صراعا سياسيا فرضته الأطماع الإقليمية على لبنان .كلفة الإنجاز لا تدخل في الحسبة. لا تذكر لها في احتفالات نصر تحرير الأسرى. ولا قيمة لها عند مئات الألوف من العرب المحبطين المتعطشين لأي نصر وبأي ثمن.

- فقد وعد أن يحرر القنطار ووفى: Hezbollah follows through and is effective (HEZBOLLAH:POS_POWER)
- بهذا الإنجاز: the phrasing as “this achievement” indicates authorial alignment with Hezbollah in two ways: 1) the use of “this” instead of “that” (HEZBOLLAH:POS_INTIMACY), 2) the use “achievement” to spin the occurrence as something positive (HEZBOLLAH:POS_VIRTUE)
- صراعا سياسيا فرضته الأطماع الإقليمية: the word choice (“political conflict imposed by territorial ambitions/greed”) makes Hezbollah seem petty and greedy (HEZBOLLAH:NEG_VIRTUE)

9

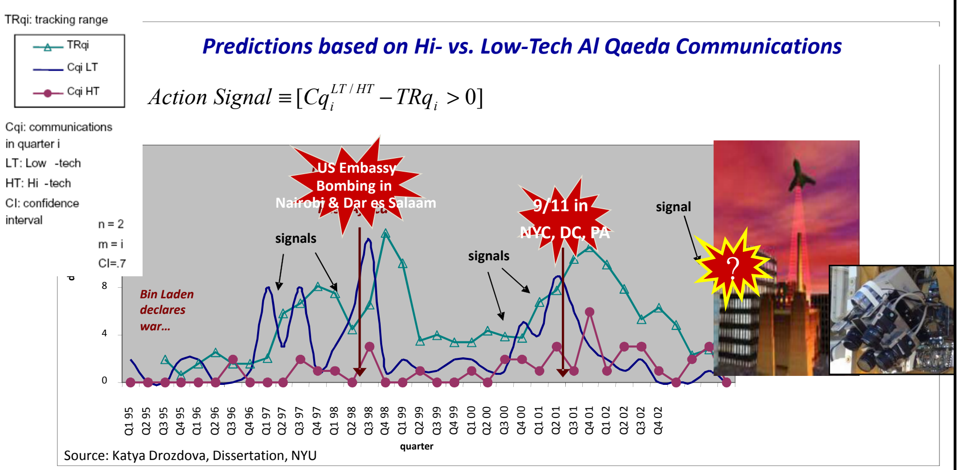
Fusion of Indirect & Direct Sensing

Anticipation/ Fusion of Sensing & “Intel” to Disrupt Adversary’s OODA Loop (Universal Situational Awareness)



10

Example: Detecting Terrorist Attack Precursors



11

Modeling for Context, Interpretation

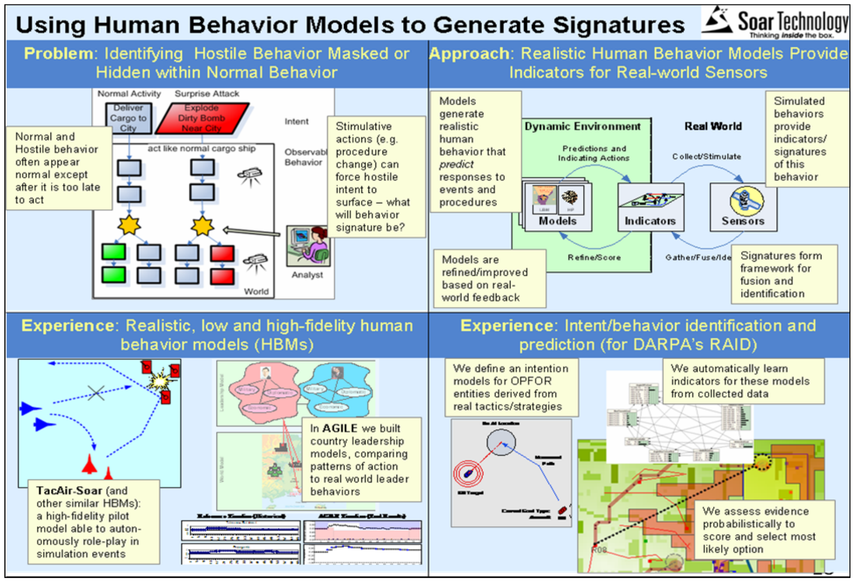
Ref: M&C News: "**In photos: 'People in Abkhazia Celebrate Russia's recognition of Independence" By James Wray,** Aug 26, 2008 (EPA/MAXIM SHIPENKOV)

11 May 2012

12

xx

Modeling for Informing Development of Mental Schema



13

Potential Issues

Numerous issues, including

- **Models need data**
  - Right resolution, right format, current, sufficient amount
  - Data (ingest/coding/characterization)
- **Usability**
- **Generalizability, Ease of adaptability**
- **Resolution of model is often not suited to the use (e.g., macro model for micro level analysis)**
- **Many models based on sparse data or Subject Matter Expert input (potential issues of bias, agreement, etc.)**

➢ **Must educate users on "boundary conditions" of model (assumptions, impact of missing/inaccurate data, resolution of model)**

14

Summary

Socio-cultural models can serve a variety of purposes for sensemaking and fusion (especially situation awareness and threat detection) by:

- Providing indirect information to layered sensing
- Enabling the detection of potential threat signatures and narratives and more nuanced understanding
- Identifying precursor signals to assist in resource allocation and cueing
- Providing a frame or lens through which to view and interpret data/information
- Assisting humans in sensemaking by informing the development of mental schema, consider possibilities and understand connections in order to anticipate and act
- Fostering a contextual understanding of a situation, actors that enables pattern interpretation and discovery of hidden meaning

11 May 2012 15

Behavior Modeling for Sensemaking and Fusion

Laurie Fenstermacher
Air Force Research Laboratory, 711 HPW/RHXM, 2255 H Street,
Wright Patterson AFB, OH 45433

ABSTRACT

Socio-cultural behavior modeling can serve a variety of purposes in supporting sensing and fusion for situation awareness and threat detection. Various analytic approaches, methods and technologies can provide the indirect information to layered sensing and enable the detection of potential threat signatures and threat narratives, identify precursor signals that assist in resource allocation and cueing, and provide a frame or lens through which to view and interpret data. In addition, behavior models can assist humans in sensemaking by informing the development of mental schema and helping them to consider possibilities (collection, analysis, synthesis) and understand connections (people, places, events) in order to anticipate/forecast and act effectively. Behavior modeling ultimately fosters a contextual understanding of a situation and actors in a way that enables an analyst to interpret patterns and uncover hidden meaning, resulting in a nuanced understanding of the "as-is" in a way that elucidates plausible futures and effects needed to achieve a desired "to-be".

Keywords: socio-cultural modeling, behavior modeling, fusion, sensemaking

1. INTRODUCTION

Developing situation awareness and understanding threat and intent in many of the environments in which the US military is employed currently is extraordinarily difficult. The goal, in many cases, is to sense/detect/track an individual in the midst of other individuals in a classically high clutter environment. Adversaries conceal operations and intentions in "plain sight" and adapt their activities. Individuals or groups who pose a threat are no longer so easily distinguished from other individuals or groups by a picture, a transaction, a conversation. Anomalies may never appear or appear too late. Emergent properties evolve over time as actors, situations and environments change. It will take a more sophisticated approach, one that is informed by social science theory and computational social science methods and models at various levels of analysis ranging from cognitive to society level in order to piece together the available information and enable sensemaking or meaning making.

1.1 Layered Sensing

In 2008, the Sensors Directorate of the Air Force Research Laboratory published a paper on the layered sensing. The paper stated that in order to achieve the goal of universal situational awareness (the ability to understand the capabilities, intents and locations)[1] would require both direct (physical sensors) and indirect sensing. Indirect sensing incorporates information from non-physical measures including "semantic reasoning and cultural nuances"[2]. Among other things, socio-cultural methods and models enable semantic reasoning and the development of a nuanced understanding of culture, attitudes, beliefs and motivations. The desired capability of universal situational awareness entails requirements for detecting and countering non-traditional threats (e.g., lone actors, non-state actors, network organizations) by predicting conditions that spawn them, detecting connectivity to other adversaries and known behavior patterns/roles, rhetoric, identifying radicalization trajectories, signatures, routes of communication, procurement, etc.

1.2 Behavior Modeling

Models do not replace people, they augment their ability. Socio-cultural models can serve a variety of uses in supporting operations. Models can provide information about social dynamics and associated behaviors, potential "surprises" and unintended consequences (e.g., emergent behaviors). This information allows an analyst or planner to answer "what if's" and manage the risks associated with plausible futures. Modeling approaches offer systematic methods for data collection, foraging, classification and analysis, supporting assessments within situations, across situations and over time. Models can help compensate for the inherent limits of human information processing, such as decision heuristics and biases[3], providing alternative interpretations, hypotheses, forecasts (i.e., "have you thought about x?). Behavior models help to deal with complexity and uncertainty. Complexity in social systems is a frequent occurrence and it is very difficult for subject matter experts (SME's) to account for the effects of all possible interactions, particularly those that occur with low probability. Models help an analyst deal with and reason in spite of uncertainty (What don't I know? What's the impact of not knowing?). There is no "silver bullet" model for providing answers/insights to analysts/planners.

1.3 Sensemaking

Ultimately situation refinement/assessment and impact refinement/assessment, level 2 and level 3 fusion (respectively) in the Joint Director of Laboratories model, is sensemaking or meaning making. Sensemaking can be defined as an "approach to creating situation awareness in situations of uncertainty"[4] or as a "continuous effort to understand connections in order to anticipate their trajectories and act effectively[5]. The goal of any automation, including fusion, is to enable the human to better detect and interpret patterns and, through asking questions, foraging, analyzing, synthesizing and interpreting evidence, make sense of a situation or find meaning in the motivations and intent of various actors. The process is typically one in which information collection leads to a mental model or schema which best fits the evidence, which then leads to further collection or foraging, and then the strengthening of the mental schema or the development of a new one and so on. Due to the "wicked" nature of many situations of interest (e.g., deceptive adversaries, covert networks) and the questions being asked, it is necessary to look beneath the surface, beyond detection and identification or labeling of people (e.g., "friend or foe") to an understanding of the threat narrative. The threat narrative is the manifest dialectical and behavioral "storyline", elucidating the worldview through which people filter information from the cognitive, socio-technical and environment dimensions to create meaning regarding perceived threat and providing a "why" for the "who" and "what".

1.4 Socio-Cultural Modeling for Sensemaking and Fusion

Modeling can provide important cues for resource allocation, cueing data collection assets as to what to look for and where. For example, the forecasting models can identify key actors and forecast events of interest[6] cueing data collection for video, social media, communications, etc. Behavior variables, environmental variables (e.g., unemployment, food prices) and sentiment collectively can enable the capability to accurately forecast the likelihood of the advent or cessation of violence[7], enabling anticipatory Intelligence, Reconnaissance and Surveillance (ISR). Additionally, discourse (text, speech) analysis can be used to identify and interpret characteristic patterns in discourse related to social identity and categorization. These patterns over time can be used to forecast hostile action by a group[8], cueing the collection of other information to detect and respond or prevent the event. The in-group/out-group discourse is a thread in the fabric of a threat narrative that influences the decisions and actions of the individuals and groups. Behavior models can enable anticipation of potential actors; for example, prediction of potential of rogue organizations using leadership and process capacity models and prediction of lone actors based on behavior and social grievance and greed models.

Socio-cultural modeling can facilitate the understanding and interpretation of data in a relevant context. For example, precaution and vigilance mechanisms vary across culture; thus, information on how they manifest in a certain culture/group is important in order to interpret whether a person/group is threatened, or is a threat or not. Information on normative cultural behaviors (e.g., shooting guns in the air as a celebratory act in Afghanistan) would have been useful in determining an appropriate response. Social knowledge (customs, common interpretations, mutual socio-cultural institutions, artifacts, shared perceived history, etc.) influences the motivating dynamics of behavior. Socio-cultural models can provide the "indirect" information referred to in the layered sensing concept paper; for example, IED related human/organizational behaviors (records of communications, finances, explosives, technologies used) and socio-cultural context (current events, historical data) can serve as a signature pattern to guide future data collection and interpretation of activities and behaviors (meetings, crowd/vehicle dispersal, etc.). Socio-cultural modeling of normalcy (e.g., understanding the background and adversary groups behaviors) can serve as the basis for the detection of anomalies (e.g., the absence of people in a normally crowded marketplace prior to an attack) and thus cue attention and support sensemaking. Karl Weick identified this ability as being central to sensemaking in which the analyst is "aware of something…(a) surprise, a discrepant set of cues or something that does not fit" which provides a useful understanding of phenomena.[9]

Socio-cultural modeling can enable sensemaking by informing the development and refinement of mental schema re: situation awareness and intent. For example, analysis of multi-media data (text, speech and websites) for identifying entities and transactions can enable the detection and tracking of networks.[10] Alternatively, analysis and modeling of behaviors (people, vehicles) can enable inferencing about the function of static facilities (buildings, geographic areas) and thus intent. In addition, socio-cultural modeling can compensate for sparse and noisy data by "acting out" possible adversary courses of action and network development, identifying new indicators and signatures, in order to enable the analyst or planner to project and prepare for possible actions. Generative models have another important use. One of the common pitfalls in sensemaking involves "locking" into a particular mental model/schema and ignoring new evidence that points to a different explanation, not questioning assumptions, and missing, discarded or inadequate hypotheses. Models (for example, agent based models) that identify potential emergent behaviors can assist in overcoming heuristics and biases (e.g., cognitive bias) in interpreting a situation or behavior and assessing intent.

1.5 Potential Issues

Potential issues in the use of socio-cultural modeling to support sensemaking are many. Models need data – the right resolution, currency and amount. Some models are not terribly user friendly – they aren't well suited to the technical sophistication of the user, lack usability, transparency/drilldown, don't generalize or are difficult to adapt. Many models are based on sparse data or Subject Matter Expert input (and those SME's may not agree or understand what they don't know and building their (conceptual) models often takes a lot of time). In addition, the resolution of many models in not suited to the use; therefore, a macro-level model (e.g., modeling government and economics, poverty and unemployment) is not going to provide a precise answer regarding the future behaviors of a particular person or group. These issues are not insurmountable, but require a thorough education of all those who use socio-cultural models or their results on the "boundary conditions"; that is, what assumptions they are based on, the impact of missing or inaccurate/deceptive data, the resolution of the model (i.e., what question can this model answer?), etc.

Without socio-cultural modeling to elucidate the worldview lens and threat narrative of the individuals and groups on which we are gathering data, we will tend to be reactive and misinterpret what we see, read and hear due to "mirroring" (interpretation based on our worldview, not on the worldview of the person/group of interest). In addition, we will largely be reactive, for being truly anticipatory entails a nuanced understanding of the environment and the "other" in order to proactively anticipate their perceptions and behaviors.

REFERENCES

[1] Bryant, M., Johnson, P., Kent, B., Nowak, Michael and Rogers, S., "Layered Sensing: Its Definition, Attributes, and Guiding Principles for AFRL Strategic Technology Development" (Version 6.0), AFRL Technical Document, Sensors Directorate, Wright Patterson AFB, Ohio, 1 May 2008.

[2] Ibid.

[3] Kahneman, D. and Tversky, A., "Prospect Theory: An Analysis of Decision under Risk", Econometrica, 47(2), 263-292, 1979.

[4] Moore, D., Sensemaking: A Structure For An Intelligence Revolution, Washington DC:Clift Series, March 2011.

[5] Klein, G., Moon, B., and Hoffman, R., "Making Sense of Sensemaking 1: A Macrocognitive Model", IEEE Intelligent Systems, 21(5), 2005.

[6] O'Brien, S., "Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research" International Studies Review, 12(1), 87-104, March 2010.

[7] Shellman, S. and Levey, B., Cascading Air Power Effects Simulation (CAPES), AFRL-RH-WP-TR-2011-046 (2011).

[8] Toman, P., L. Kuznar, and T. Baker, et. al., Analysis of Discursive Accent and Discursive Practices I&W. AFRL-RH-WP-TR-2010-0128 (2010).

[9] Weick, K. Sensemaking in Organizations. Thousand Oaks, CA: Sage Publications, 1995.

[10] C. Weinstein, W. Campbell, B. Delaney and G. O'Leary, " Modeling and detection techniques for counter-terror social network analysis and intent recognition," Proceedings of the IEEE Aerospace Conference, Big Sky, Montana (2009).

Exploitation of Social-Behavioral Modeling Understanding for Information Fusion

Erik Blasch, Guna Seetharaman

AFRL/RIEA

SPIE12, Panel Discussion

Real-World Issues and Challenges in Social/Cultural Modeling with Applications to Information Fusion

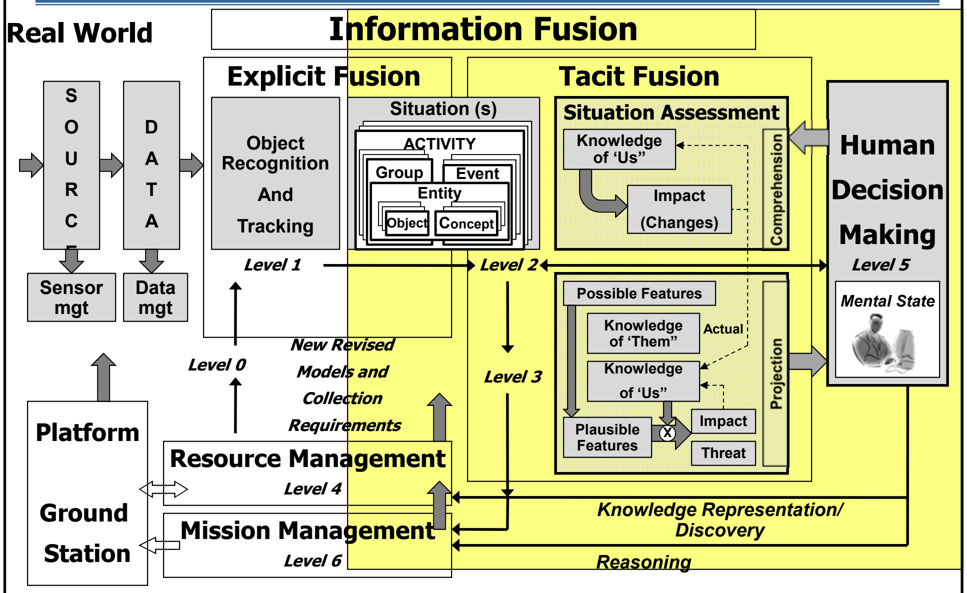
Panelists: **Erik Blasch, Air Force Research Lab. (Canada); Mica Endsley, SA Technologies (USA); Laurie H. Fenstermacher, Air Force Research Lab. (USA); Lynne L. Grewe, California State Univ., East Bay (USA); Ivan Kadar, Interlink Systems Sciences, Inc. (USA); John J. Salerno, Jr., Air Force Research Lab. (USA); Shanchieh Jay Yang, Rochester Institute of Technology (USA)**

Panel Organizers: **John J. Salerno, Jr., Air Force Research Lab. (USA); Ivan Kadar, Interlink Systems Sciences, Inc. (USA)**

Overview

- Data Fusion Information Group Model
  - Proposed additions to the JDL for actors
- Summary of past Fusion Panels
  - SA/TA/IA Assessment
  - Need: Social/Behavioral/Cultural Models
- Issues
  - Use of the SBC models and evaluation of/with for HLIF
  - EX: Human Observer: sensitivity, objectivity, veracity
- Challenges
  - Representations of SCB information using SBC models
  - Decision Support, Info Management, and evaluation

Data Fusion Information Group Model



Erik Blasch – SPIE 12

Summary – Top Ten Trends

SPIE 2012

- **Area A. Data/Knowledge Representation**
  - Reference Model
  - Taxonomy of notations, symbols, and meanings
- **Area B. SA/TA/IA Assessment**
  - Semantics/ontologies
  - Social/Behavioral/Cultural Models
- **Area C. Systems Design**
  - User/agent coordination
  - Display (interactive)
- **Area D. Evaluation**
  - Common scenario, Performance comparison
  - Metrics / Uncertainty analysis
- **Area E. Information Management**
  - Resource planning and information analysis
  - Joint theory of methods integration

Determined from 2000-2010 Panel Discussions at the International Conference on Information Fusion

Erik Blasch – SPIE 12 4

Summary

CTFG 2012

Topic / Panel	2000 Vision	2001 L2-4	2004 HLF	2005 KR-RM	2006 RM	2007 Agent	2008 HLIF	2009 Coalition	2009 TA/IA	2010 HLIF-GC
Reference Model		⊗	⊗	⊗			⊗	⊗	⊗	⊗
Data/Knowledge										
Representation				o	o		o	o	o	o
Semantics/Ontologies				o		o	o	o	o	o
SA/TA/IA Assessment	⊗	⊗	⊗	⊗		⊗		⊗	⊗	⊗
Social/Behavioral Model			⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
User/Agent Coordination	⊗	⊗	⊗	⊗	⊗	⊗		⊗	⊗	⊗
Display (Interactive)	x			x				x		x
Common Scenario			x					x	x	x
Performance Eval/Metrics	⊗	⊗	⊗	⊗	⊗			⊗	⊗	⊗
Uncertainty Analysis	o			o	o		o	o		o
Resource Planning	⊗		⊗	⊗	⊗	⊗	⊗	⊗		⊗
Joint Theory of Methods		x	x		x	x				x

⊗ Current and Consistent Theme
O Key Importance
X General Importance

Erik Blasch – SPIE 12 5

Issues

Issues for SCB modeling for HLIF :

Who are the actors and consumers of SBC models in an HLIF design?

What are the ontological discourse (framework) from which they perceive or respond to situations?

What IF methods/algorithms are important to process of social/behavioral cultural issues?

Where do we utilize the social/cultural information in the HLIF design?

When it is necessary to display (visual analytics) the SBC model results to an operator?

How do we create models of the unknown targets (behaviors of people)?

Evaluation Metrics for SCB modeling for HLIF :

What are the evaluation strategies

Erik Blasch – SPIE 12 6

Issues –Evaluation Metrics

Evaluation Metrics for SCB modeling for HLIF :

The SBC paradigm is well explored for systems (machines) monitoring other people; however, it can also apply to humans as sensor inputs.

Human actors are also sensors that need to be evaluated as to their influences from social, cultural and behavioral contexts, and their impact on the evidence these actors deliver.

For example, evaluating the credibility of a source providing testimonial evidence involves three variables: [19].

- *Sensitivity* takes into account that a human source can discriminate between events based on their past social interactions, hypotheses, and expectations.
- *Objectivity* of a source is related to one's culture, motivations, and beliefs.
- *Veracity*, involves the truthfulness of the evidence a source provides which is based on their character and behaviors (e.g. truthful, accurate, integrity).

Bennett, M., and Waltz, E., *Counterdeception Principles and Applications for National Security*, Artech House, 2007.

Challenges

Challenges for SCB modeling for HLIF :

1) ***Situation modeling theory*** (context, environments, and processes) for association management;

2) ***Decision support processes*** (reasoning, inference, and explanation of relationships) from SBC models to answer user's needs;

3) ***Standardized evaluation methods*** (measures of performance/ effectiveness, and empirical case studies) to conduct SCB evaluation separately and within an HLIF system;

4) ***Systems design techniques*** (scenario-based, user-based, and distributed-agent) to provide reasoning capabilities over difference contexts, cultural and social situations, and among different actors; and

5) ***Representations of SCB information*** (semantic, knowledge, and complex) for acquisition, relevancy, and processing of SCB data and information.

Overview

- Data Fusion Information Group Model
  - Proposed additions to the JDL for actors
- Summary of past Fusion Panels
  - SA/TA/IA Assessment
  - Need: Social/Behavioral/Cultural Models
- Issues
  - Use of the SBC models and evaluation of/with for HLIF
  - EX: Human Observer: sensitivity, objectivity, veracity
- Challenges
  - Representations of SCB information using SBC models
  - Decision Support, Info Management, and evaluation

Erik Blasch – SPIE 12 9

backup

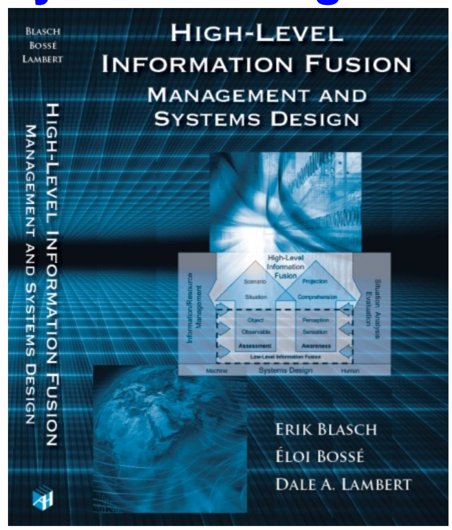
Erik Blasch – SPIE 12

xxx

New Text (Artech 2012)
High-Level Information Fusion Management and Systems Design

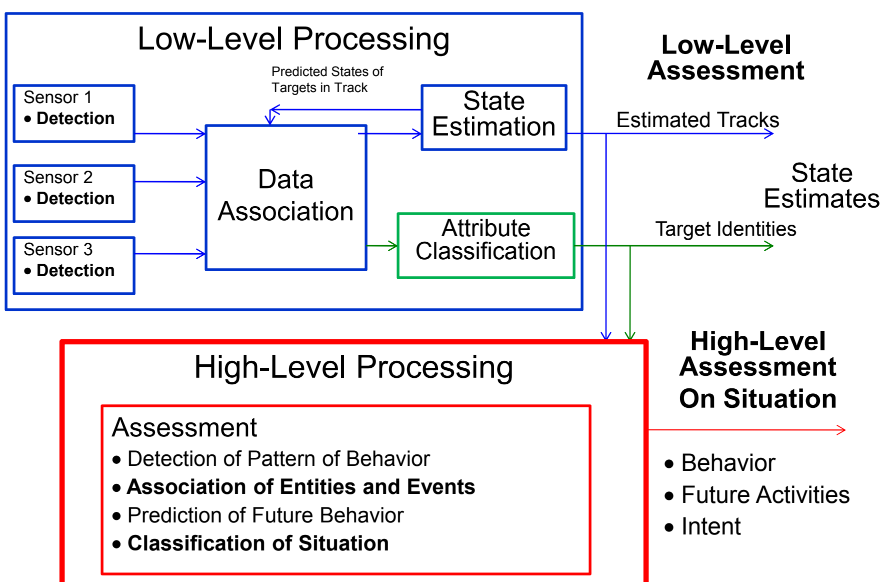
Editors:
Erik Blasch(AFRL),
Dale Lambert (AUS),
Éloi Bossé (Univ. Laval)

Authors:
Adel Guitouni, Anne-Laure Jousselme, Patrick Maupin, Luc Pigeon, Pierre Valin (DRDC),
Mark Linderman, Michael Hinman, John Salerno, George Tadda (AFRL)
Elizabeth K. Bowman (ARL),
Peter D. Houghton (UK),
Elisa Shahbazian (OODA)



High Level Fusion

Adapted from E. Waltz and J. Llinas, *Multisensor Data Fusion,* Artech House, Norwood, MA [1990])



HLIF Grand Challenges – Lambert 2003

Semantic Challenge: What symbols should be used and how do those symbols acquire meaning?

Epistemic Challenge: What information should we represent and how should it be represented and processed within the machine?

Paradigm Challenge: How should the interdependency between the sensor fusion and information fusion paradigms be managed?

Interface Challenge: How do we interface people to complex symbolic information stored within machines?

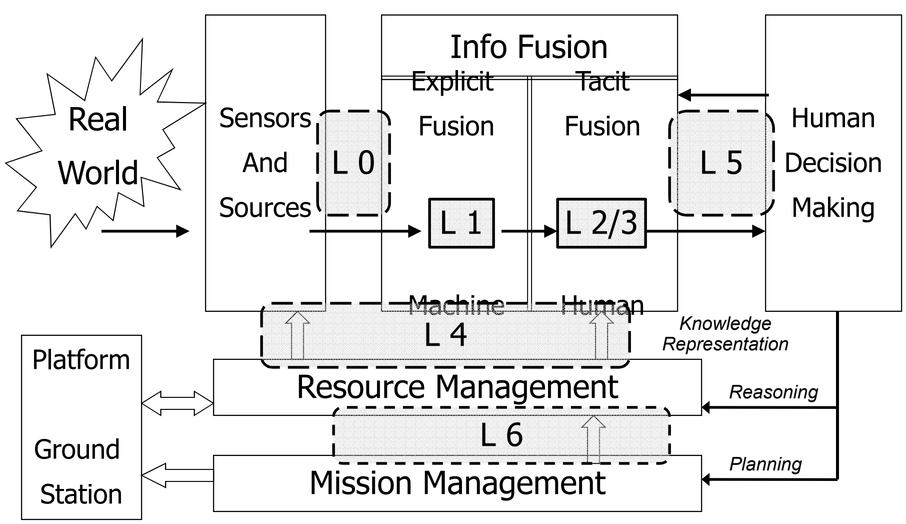
System Challenge: How should we manage data fusion systems formed from combinations of people and machines?

ADDED – 2010

Design Challenge: How should we design information fusion systems formed from combinations of people and machines?

Evaluation Challenge: How should we evaluate the effectiveness of information fusion systems?

HLIF Functions – DFIG Model Erik Blasch (Fusion 2010)

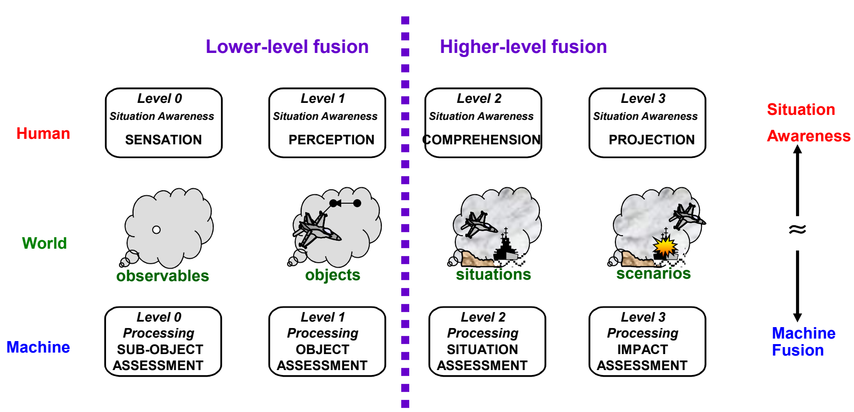


HLIF Functions – DFIG Model
Erik Blasch (Fusion 2010)

- Level 0 – Data Assessment (DA): estimation and prediction of signal/object observable states on the basis of pixel/signal level data association (e.g. information systems collections);
- Level 1 – Object Assessment (OA): estimation and prediction of entity states on the basis of data association, continuous state estimation and discrete state estimation (e.g. data processing);
- Level 2 – Situation Assessment (SA): estimation and prediction of relations among entities, to include force structure and force relations, communications, etc. (e.g. information processing);
- Level 3 – Impact Assessment (IA): estimation and prediction of effects on situations of planned or estimated actions by the participants; to include interactions between action plans of multiple players (e.g. assessing threat /intent actions to planned actions and mission requirements, performance evaluation);
- Level 4 – Process Refinement (PR): (an element of Resource Management): adaptive data acquisition and processing to support sensing objectives (e.g. fusion process control and information systems dissemination).
- Level 5 – User Refinement (UR): (an element of Knowledge Management): adaptive determination of who queries information and who has access to information (e.g. information operations) and adaptive data retrieved and displayed to support cognitive decision making and actions (e.g. human systems integration).
- Level 6 – Mission Management (MM): (an element of Platform Management): adaptive determination of spatial-temporal control of assets (e.g. airspace operations) and route planning and goal determination to support team decision making and actions (e.g. theater operations) under social, economic, and political constraints.

Erik Blasch – SPIE 12

STDF Model – Lambert circa 2006
Dale Lambert (Fusion 2010)

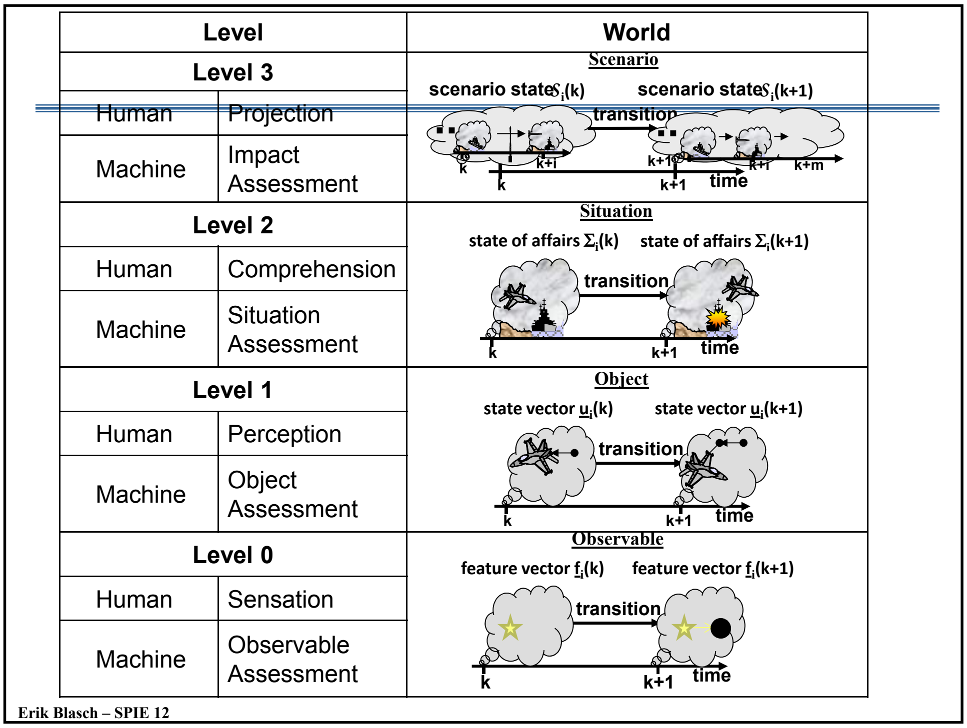


Division of Labour: (Lambert, circa 2001)

- *situation awareness is fusion performed by people*
- *machine fusion is "situation awareness" performed by machines.*

Erik Blasch – SPIE 12

xxxiii


Level
World
Level 3
Scenario
scenario state $S_i(k)$
scenario state $S_i(k+1)$
Human
Projection
transition
Machine
Impact Assessment
k
k+i
k+1
k+i
k+m
time
Level 2
Situation
state of affairs $\Sigma_i(k)$
state of affairs $\Sigma_i(k+1)$
Human
Comprehension
transition
Machine
Situation Assessment
k
k+1
time
Level 1
Object
state vector $\underline{u}_i(k)$
state vector $\underline{u}_i(k+1)$
Human
Perception
transition
Machine
Object Assessment
k
k+1
time
Level 0
Observable
feature vector $\underline{f}_i(k)$
feature vector $\underline{f}_i(k+1)$
Human
Sensation
transition
Machine
Observable Assessment
k
k+1
time
Erik Blasch – SPIE 12

Exploitation of Social-Behavioral Modeling Understanding for Information Fusion

Erik P. Blasch and Guna Seetharaman
Air Force Research Laboratory
Rome, NY 13441

ABSTRACT

For the last decade, the information fusion community has explored methods from human factors (e.g.; situational awareness, decision support), human modeling (e.g.; motion and behavioral models), and human-organization interactions (e.g.; social network and business analysis). For each of these methods, there is a profound contribution they bring to future complex information fusion systems. Roughly, attributes of human interaction, social/cultural/behavioral (SCB) modeling, and operational analysis can be thought of as high-level information fusion (HLIF). Low-level information fusion (LLIF) is concerned with object tracking, identification, associations, cliques, and spanning subsets, whereas HLIF is concerned with situations, impacts, causal-links, and management functions. In this presentation/paper, we discuss the importance of human modeling and cultural interactions as corresponding to the perceptual, situation awareness, and situation assessment developments from the other panelists. We emphasize the link between human SCB modeling in high-level information fusion which has three real-world issues and challenges (1) knowledge representation and modeling, (2) assessment semantics, and (3) SCB model evaluation.

Keywords: DFIG Fusion Model, Situation Awareness, User Refinement, Social/Cultural/Behavioral Modeling

INTRODUCTION: *Knowledge Representation and Modeling*: With the advent of web technology, the real-world challenges and issues to information fusion have progressed from low-level information fusion (object assessment of tracking and classification/identification) to high-level information fusion of situation/impact assessment (SA/IA), situation awareness (SAW), and information management (IM).[1] IM seeks to coordinate the information fusion products with the user's needs [2] such as objects, activities, events, and relationships among them over geospatial, temporal, and semantic properties [3]. Users have an important role to play in an information fusion system design [4] and management and bring contextual, cultural, social, and business understanding to the data, missions, and sensor products.[5] Figure 1 presents the unification of the Joint Director of the Labs (JDL) model and its variants, the Data Fusion Information Group (DFIG) model [6], Endsley's SAW model [7], and Salerno's SA model [8]. The key element of the model for the panel discussion is the notion of "us" and "them" [1]. The coordination of "us" and "them" was evaluated using multiple OODA loops [9]; but there is a concern of what constitutes the model, what knowledge is represented in the models, and the interaction between the "us" and "them" models.

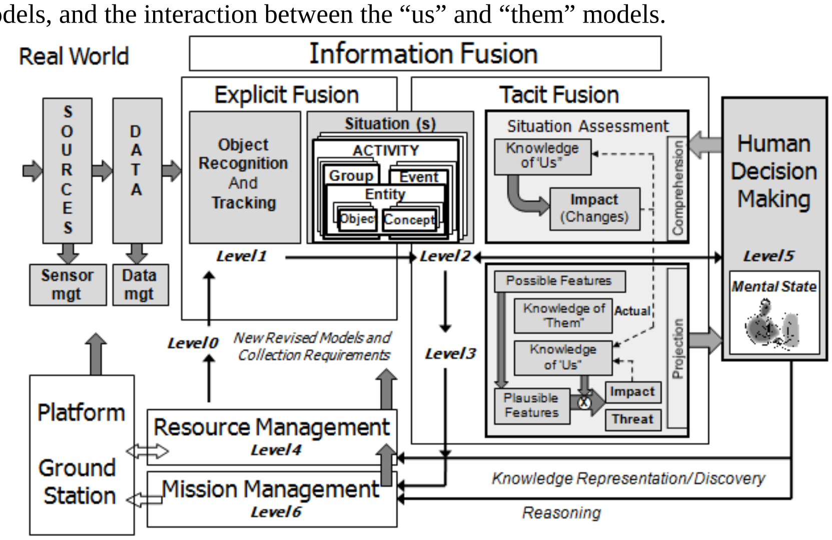


Figure 1. Information Fusion Situation Assessment (IFSA) model.

One of the challenges to understand the "us-them" is to model the social/cultural effects in information fusion. Numerous efforts are supporting research thrusts in knowledge management for HLIF, but there are key issues needed of assessment semantics and evaluation of SCB model development.

HLIF CHALLENGES: *Assessment Semantics*: From the last decade of HLIF panel discussions, papers, and summaries; a notional taxonomy of the results is shown in Table 1. Table 1 demonstrates the pervasive topics that are of interest to the community over the last decade from the Fusion Conference panel discussions. The sustained topics represent challenges for the community that are unresolved or need more attention such as situation/threat/impact assessment (SA/TA/IA), SCB modeling, and common ontologies and semantics, which we group as "assessment semantics."

Table 1: Key Topics from High-Level Information Fusion Panels over the Last Decade

Panel	2000	2001	2004	2005	2006	2007	2008	2009	2009	2010
Topic	Vision	L2-4	HLF	KR-RM	RM	Agent	HLIF	Coalition	TA/IA	HLIF-GC
Reference Model	⊗	⊗	⊗	⊗			⊗	⊗	⊗	⊗
Knowledge Representation				o	o		o	o	o	o
Semantics/Ontologies				o		o	o	o	o	o
SA/TA/IA Assessment	⊗	⊗	⊗	⊗		⊗		⊗	⊗	⊗
Social/Behavioral Model			⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
User/Agent Coordination	⊗	⊗	⊗	⊗	⊗	⊗		⊗	⊗	⊗
Display (Interactive)	x			x				x		x
Common Scenario			x					x	x	x
Performance Eval/Metrics	⊗	⊗	⊗	⊗	⊗			⊗	⊗	⊗
Uncertainty Analysis	o			o	o		o	o		o
Resource Planning	⊗		⊗	⊗	⊗	⊗	⊗	⊗		⊗
Joint Theory of Methods		x	x		x	x				x

Of the many issues in HLIF that have been posed; it would be difficult to say that any have been solved. For example, there is an existing wealth of knowledge in Graph Theory [10], of which scalability or complexity has not been the main barrier impeding progress. The cause is more likely to be the elusive nature of the quantifiable features that can be captured to represent "associations," including suitable representations for multi time-scale graph analysis. As technology and information change, so do the systems that are designed to synthesize the data for users. If we capture the issues from the panels of the past decade, we see consistent themes that are important. We see that the most common discussion was on ***social/cultural/behavioral (SBC) models*** which supports situation modeling theory for threat and impact assessment and awareness [11]. The second most common theme is ***user and agent (machine)*** coordination that incorporates decision support. An emerging theme for modeling is the ***semantics and ontologies*** that also require ***data and knowledge representations***. A common reference model and resource planning are important as system design techniques that facilitate both the operational development and deployment of HLIF systems, respectively. Form Table 1, we see that to enable real-world HLIF situation/threat/impact assessment (SA/TA/IA), common SBC models [12, 13] are needed with an organized set of semantics and ontologies.

While covered in a few panel discussions, we might conclude that the complexity, difficulty, and undefined nature of HLIF all limits the ability to fully capture a joint theory across all levels of information fusion, employment of common scenario of interest to all developers, and research analysis into display technology for the multitude of HLIF designs. The third theme is ***performance evaluation and metrics*** such as uncertainty analysis and common scenarios for standardization of evaluation methods. [14] A real-world issue of evaluation will also include SCB model development, validation, and verification.

SOCIAL/CULTURAL MODELING: *SCB Model Evaluation*: Key elements of real-world issues and challenges in Social/Cultural Modeling could vary over the users (knowledge management), systems (i.e. resources, sensors), and applications (targets and environment). The connection between SBC modeling and HLIF *knowledge management* is based on a semantic information representation framework otology {thing, place, path, action, cause} [15]. Modeling these attributes can be of concept, primitive, feature, or signal layers of information fusion processing. One example is models for language and sentiment analysis BULLBEAR [16] which seeks methods for decision support such as (1) data associations, (2) source pedigree reducing duplicates from derived sources of information, as well as ontologies and analysis. For *systems*, there is a needed for software [17] to be flexible to the needs of SBC modeling. Finally, for targets in the environment, the parameters of the models need to be understood in relation to the contextual factors [18].

The SBC paradigm is well explored for systems (machines) monitoring other people; however, it can also apply to humans as sensor inputs. Human actors are also sensors that need to be evaluated as to their influences from social, cultural and behavioral contexts, and their impact on the evidence these actors deliver. For example, evaluating the credibility of a source providing testimonial evidence involves three variables: observational sensitivity, objectivity, and veracity [19]. *Sensitivity* takes into account that a human source can discriminate between events based on their past social interactions, hypotheses, and expectations. The *objectivity* of a source is related to one's culture, motivations, and beliefs. *Veracity*, involves the truthfulness of the evidence a source provides which is based on their character and behaviors (e.g. truthful, accurate, integrity).

Open **issues** for SCB modeling for HLIF are:

Who are the actors and consumers of SBC models in an HLIF design?
What are the ontological discourse (framework) from which they perceive or respond to situations?
What IF methods/algorithms are important to process of social/behavioral cultural issues?
Where do we utilize the social/cultural information in the HLIF design?
When it is necessary to display (visual analytics) the SBC model results to an operator?
How do we create models of the unknown targets (behaviors of people)?

Our brief paper has outlined the need for social/cultural/behavior modeling in the development of future information fusion systems designs. From an analysis of the HLIF discussions of past panel conferences on information fusion, one pervasive issue is that of social/cultural/behavioral modeling. Some analysis revealed that it is less well known, studied, or defined and thus remains a critical challenge and issue in IF designs.

Challenges for SCB modeling for HLIF based on issues from the last decade are:

1) ***Situation modeling theory*** (context, environments, and processes) for association management;
2) ***Decision support processes*** (reasoning, inference, and explanation of relationships) from SBC models to answer user's needs;
3) ***Standardized evaluation methods*** (measures of performance/ effectiveness, and empirical case studies) to conduct SCB evaluation separately and within an HLIF system;
4) ***Systems design techniques*** (scenario-based, user-based, and distributed-agent) to provide reasoning capabilities over difference contexts, cultural and social situations, and among different actors; and
5) ***Representations of SCB information*** (semantic, knowledge, and complex) for acquisition, relevancy, and processing of SCB data and information.

REFERENCES

[1] Blasch, E., Salerno, J. J., and Tadda, G., "Measuring the Worthiness of Situation Assessment," *IEEE Nat. Aero. Electronics Conf.*, 2011.
[2] Blasch, E., Bosse, E., and Lambert, D.A., *High-Level Information Fusion Management and Systems Design*, Artech House, 2012.
[3] Blasch, E., Deignan, Jr, P. B., Dockstader, S. L., Pellechia, M., Palaniappan, K., and Seetharaman, G., "Contemporary Concerns in Geographical/Geospatial Information Systems (GIS) Processing," *Proc. IEEE Nat. Aerospace Electronics Conf.*, 2011.
[4] Blasch, E., "Assembling an Information-fused Human-Computer Cognitive Decision Making Tool," *Int. Conf. on Info Fusion*, 1999.
[5] Blasch, E., Valin, P., Bosse, E., Nilsson, M., Van Laere, J., and Shahbazian, E. "Implication of Culture: User Roles in Information Fusion for Enhanced Situational Understanding," *Int. Conf. on Info Fusion*, 2009.
[6] Blasch, E., Kadar, I., Salerno, J. J., Kokar, M. M., Das, S., Powell, G M., Corkill, D. D., and Ruspini, E.. H., "Issues and Challenges in Situation Assessment (Level 2 Fusion)," *J. of Adv. in Info. Fusion*, Vol. 1, No. 2, pp. 122–139, December 2006.
[7] Endsley, M. R., and D. J. Garland, *Situation Awareness Analysis and Measurement*, Mahwah, NJ: Lawrence Erlbaum, 2000.
[8] Salerno, J. J., Sudit, M., Yang, S. J., Tadda, G. P., Kadar, I., and Holsopple, J., "Issues and Challenges in Higher Level Fusion: Threat/Impact Assessment and Intent Modeling (A Panel Summary)," *Intl. Conf on Info. Fusion*, 2010.
[9] Blasch, E. "Proactive Decision Fusion for Site Security," *Int. Conf. on Info Fusion*, 2005.
[10] Holsopple, J. and Yang, S. J., "FuSIA: Future Situation and Impact Awareness," *International Conf. on Information Fusion*, 2008.
[11] Ferry, J. P. and J. O. Bmgarner, "Community detection and tracking on network from a data fusion perspective," submitted for review.
[12] Social Computing, Behavioral-Cultural Modeling and Prediction Conferences.
[13] Liu, H., Salerno, J. J., and Young, M. J., *Social Computing and Behavior Modeling*, Springer, 2009.
[14] Salerno, J. J., Blasch, E., Hinman, M., and Boulware, D., "Evaluating Algorithmic Techniques in Supporting Situation Awareness," *Proc. SPIE*, Vol. 5813, 2005.
[15] Yin, Y., and Man, H., "Behavior Modeling of Human Objects in Multimedia Content", in *Multimedia Security and Steganography*, Frank Shih Editor, CRC Press, 2012.
[16] OODA technologies - http://www.ooda.ca/BullBear_en.html
[17] Grewe, L., *OpenSocial Network Programming*, Wiley Publishing, 2009.
[18] Fenstermacher, L., Leventhal, T., and Canna, S., *Countering Violent Extremism, Scientific Methods* and *Strategies*, 2011.
[19] Bennett, M., and Waltz, E., *Counterdeception Principles and Applications for National Security*, Artech House, 2007.

The Interest Graph and Beyond--

Social Modeling and Information Fusion

Lynne Grewe,

Professor Computer Science

California State University East Bay

lynne.grewe@csueastbay.edu

Who can get Johnny to buy that lolipop?

The Interest Graph

- modeling of users/organizations/groups interests and their interactivity and influence.
- answers the questions like “who knows X” and “who can influence A about X”.
- Construction uses data understanding and fusion techniques on social and other data

What is the Input

- Social Data examples
  - interests, status, activities, work, education, movies, tv, smoking, drinking, looking for, gender, age, name, addresses (often private), current location, status, social status, kids, languages spoken, religion, ethnicity, living arrangements, sports, birthday, fashion, food, humor, posts, favorite song, movie, heroes, quotes and more.
- Other "Social" (focused) Data –Intelligence Reports
- Physical Data – video, image, maps, audio, …

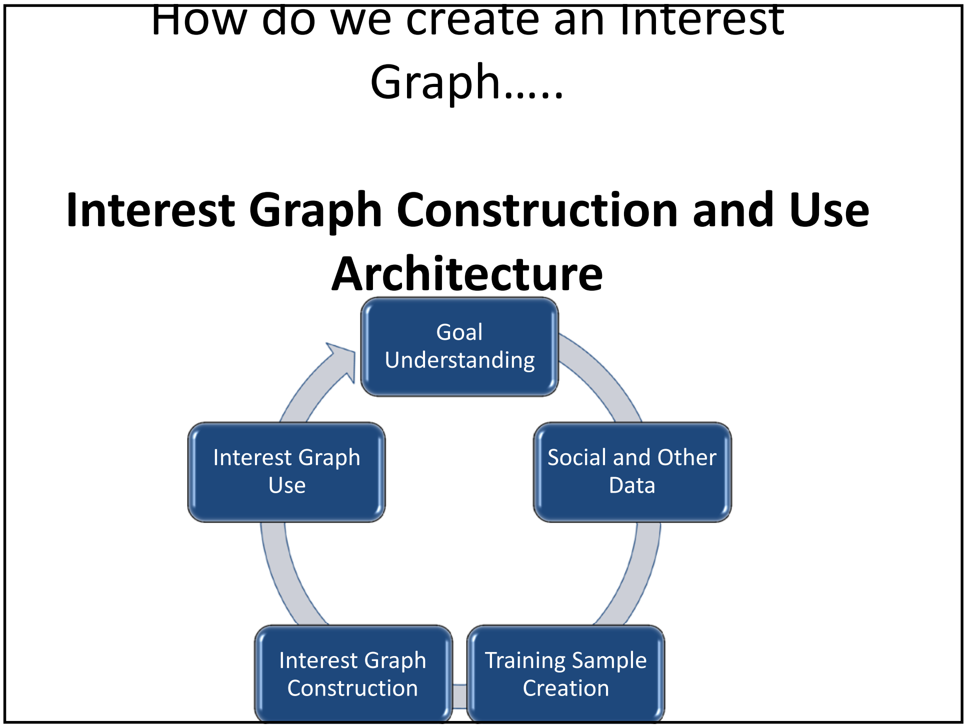
Social Data—we've got problems

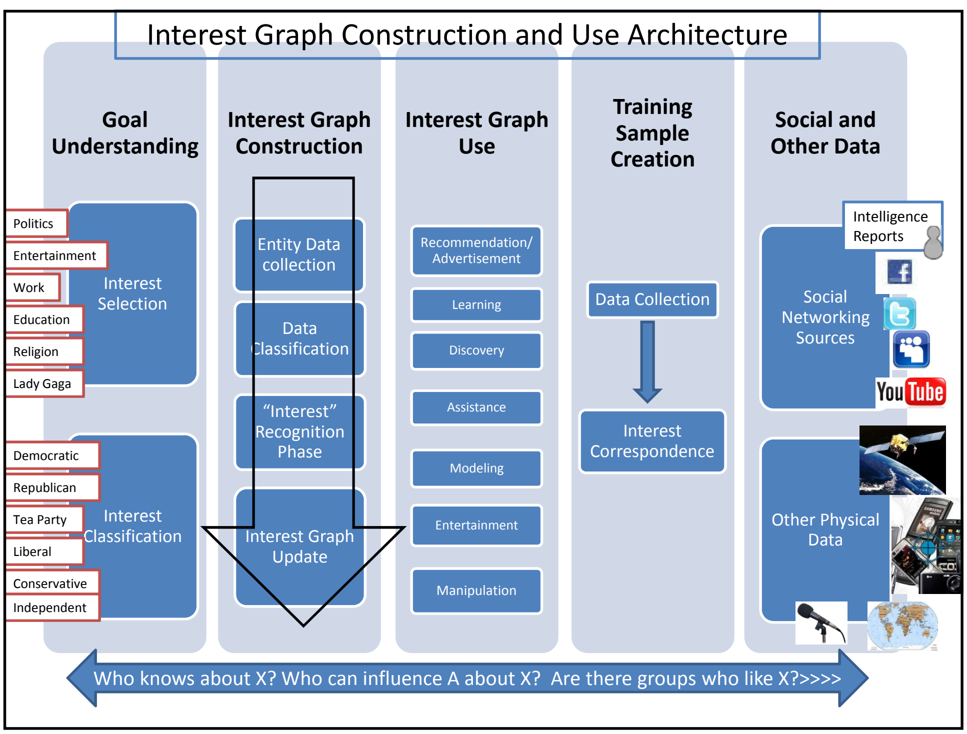
- Veracity

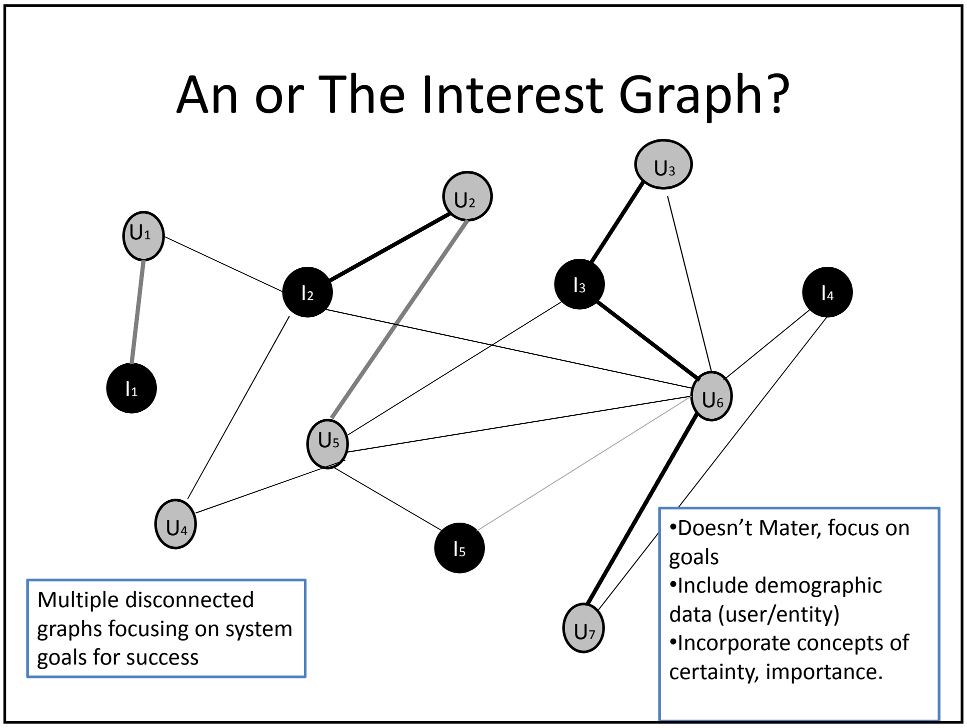
  I invented the equation V=IR
- Intelligibility

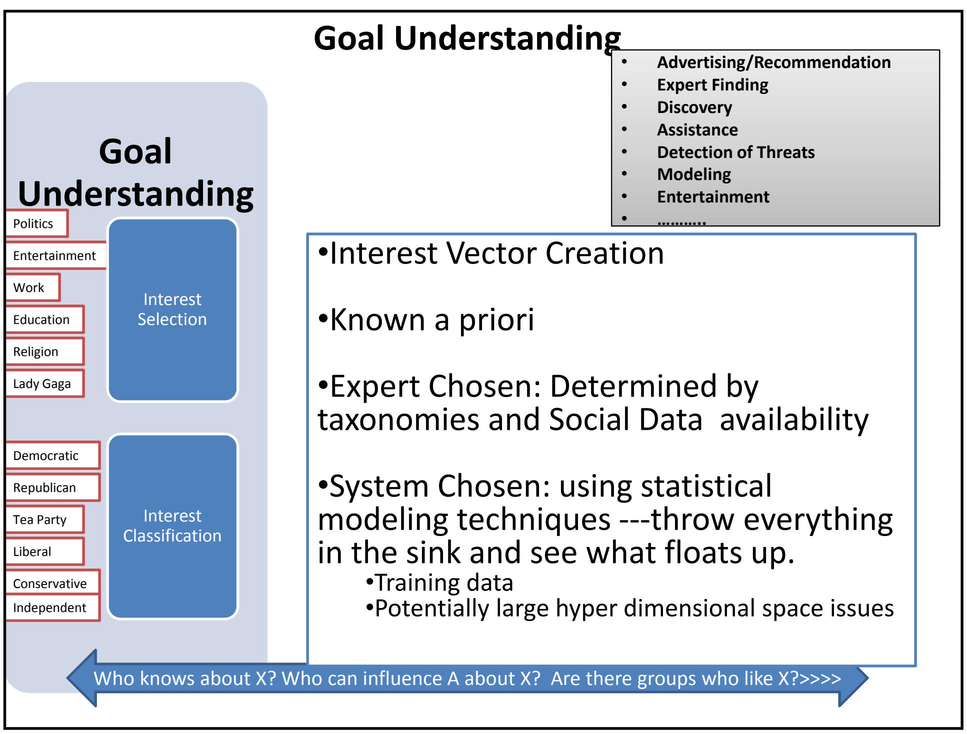
  OMG, you are BFF LoL
- Context/Culture
- Positive/Negative

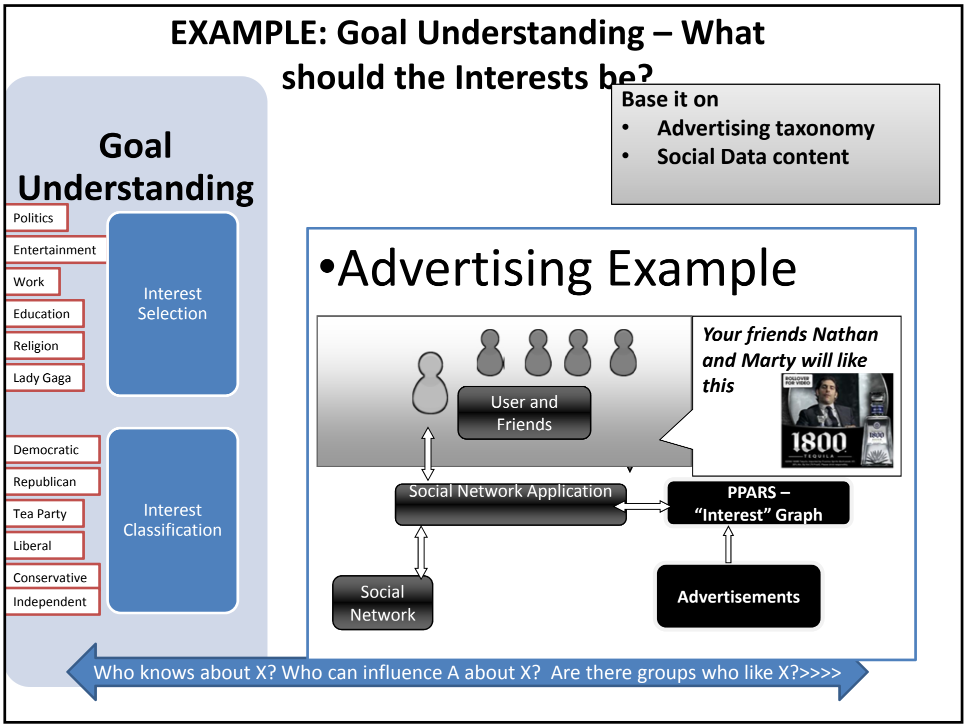
  I really love people who smoke, especially when they blow it in my face.
- Certainty/Importance

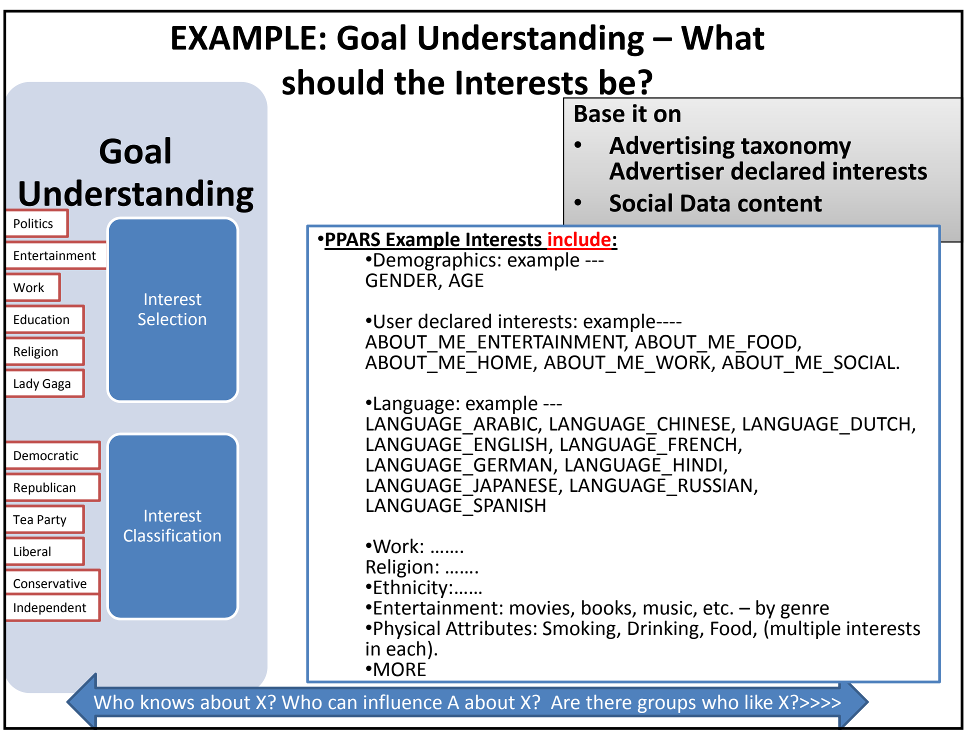

How do we create an Interest Graph…..
Interest Graph Construction and Use Architecture
Goal Understanding
Social and Other Data
Training Sample Creation
Interest Graph Construction
Interest Graph Use


Interest Graph Construction and Use Architecture
Goal Understanding
Interest Graph Construction
Interest Graph Use
Training Sample Creation
Social and Other Data
Politics
Entertainment
Work
Education
Religion
Lady Gaga
Interest Selection
Democratic
Republican
Tea Party
Liberal
Conservative
Independent
Interest Classification
Entity Data collection
Data Classification
"Interest" Recognition Phase
Interest Graph Update
Recommendation/ Advertisement
Learning
Discovery
Assistance
Modeling
Entertainment
Manipulation
Data Collection
Interest Correspondence
Intelligence Reports
Social Networking Sources
Other Physical Data
Who knows about X? Who can influence A about X? Are there groups who like X?>>>>

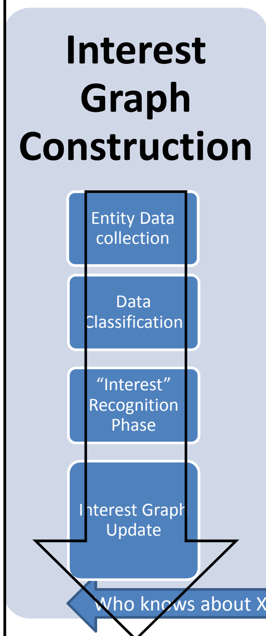

An or The Interest Graph?
U1
U2
U3
I1
I2
I3
I4
U4
U5
U6
I5
U7
Multiple disconnected graphs focusing on system goals for success
•Doesn't Mater, focus on goals
•Include demographic data (user/entity)
•Incorporate concepts of certainty, importance.


Goal Understanding
• Advertising/Recommendation
• Expert Finding
• Discovery
• Assistance
• Detection of Threats
• Modeling
• Entertainment
• ....
Goal Understanding
Politics
Entertainment
Work
Education
Religion
Lady Gaga
Interest Selection
Democratic
Republican
Tea Party
Liberal
Conservative
Independent
Interest Classification
•Interest Vector Creation
•Known a priori
•Expert Chosen: Determined by taxonomies and Social Data availability
•System Chosen: using statistical modeling techniques ---throw everything in the sink and see what floats up.
•Training data
•Potentially large hyper dimensional space issues
Who knows about X? Who can influence A about X? Are there groups who like X?>>>>


EXAMPLE: Goal Understanding – What should the Interests be?
Base it on
• Advertising taxonomy
• Social Data content
Goal Understanding
Politics
Entertainment
Work
Education
Religion
Lady Gaga
Interest Selection
Democratic
Republican
Tea Party
Liberal
Conservative
Independent
Interest Classification
•Advertising Example
User and Friends
Your friends Nathan and Marty will like this
1800
Social Network Application
PPARS – "Interest" Graph
Social Network
Advertisements
Who knows about X? Who can influence A about X? Are there groups who like X?>>>>


EXAMPLE: Goal Understanding – What should the Interests be?
Base it on
• Advertising taxonomy
Advertiser declared interests
• Social Data content
Goal Understanding
Politics
Entertainment
Work
Education
Religion
Lady Gaga
Interest Selection
Democratic
Republican
Tea Party
Liberal
Conservative
Independent
Interest Classification
•PPARS Example Interests include:
•Demographics: example --- GENDER, AGE
•User declared interests: example---- ABOUT_ME_ENTERTAINMENT, ABOUT_ME_FOOD, ABOUT_ME_HOME, ABOUT_ME_WORK, ABOUT_ME_SOCIAL.
•Language: example --- LANGUAGE_ARABIC, LANGUAGE_CHINESE, LANGUAGE_DUTCH, LANGUAGE_ENGLISH, LANGUAGE_FRENCH, LANGUAGE_GERMAN, LANGUAGE_HINDI, LANGUAGE_JAPANESE, LANGUAGE_RUSSIAN, LANGUAGE_SPANISH
•Work: …….
Religion: …….
•Ethnicity:……
•Entertainment: movies, books, music, etc. – by genre
•Physical Attributes: Smoking, Drinking, Food, (multiple interests in each).
•MORE
Who knows about X? Who can influence A about X? Are there groups who like X?>>>>

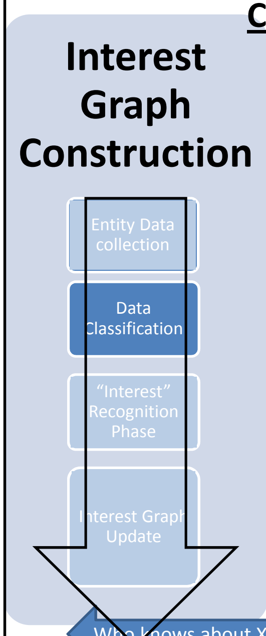
Interest Graph Construction



- Data Classification and Processing
  - Segmentation / Parsing
  - Data Types
- Interest Recognition Phase
  - Techniques based on Data Types
- Interest Graph Update

Who knows about X? Who can influence A about X? Are there groups who like X?>>>>

Interest Graph Construction –<u>Data Classification and Processing</u>



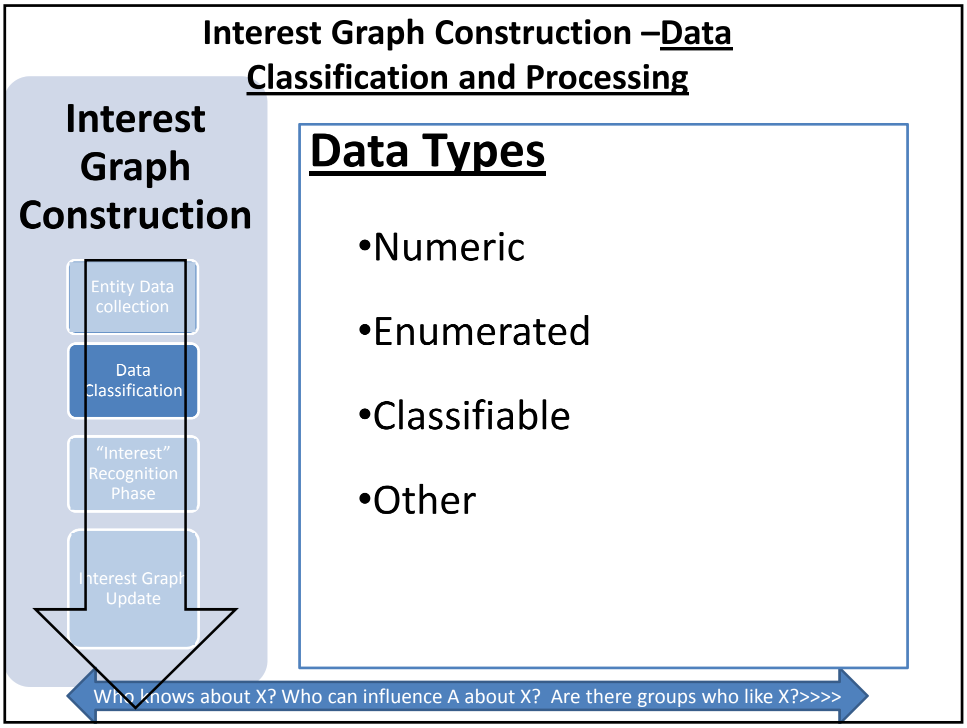
<u>Segmentation / Parsing</u>

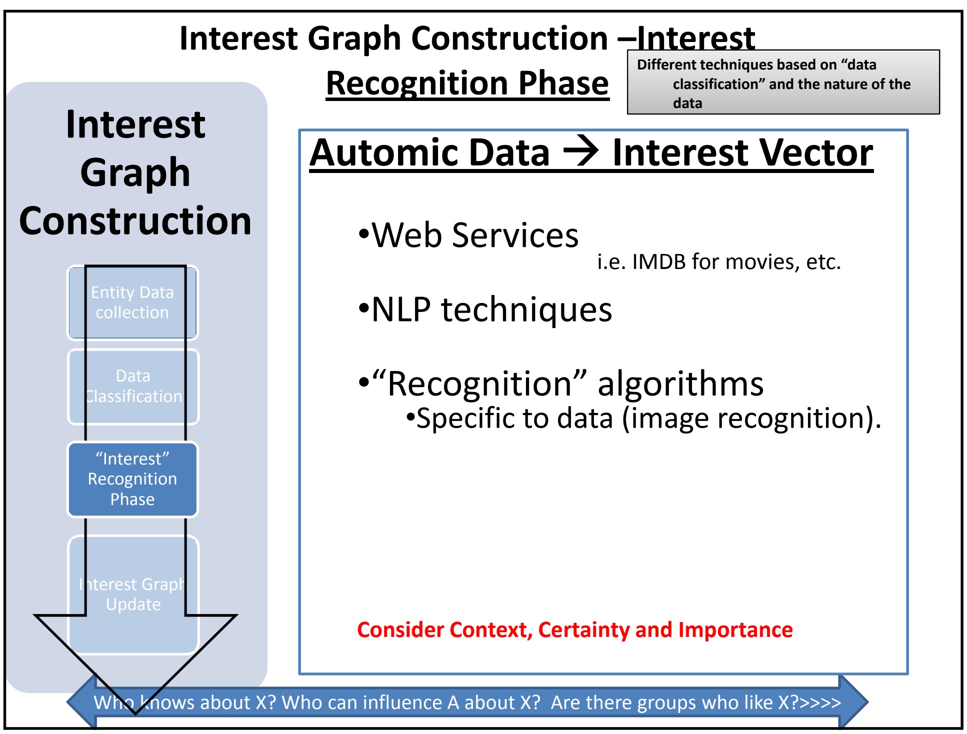
" Purdue University West Lafayette, IN Graduated: 2010 Student status: Alumni Degree: Bachelor's Degree Major: Business/ Management 2002 to 2006 Kokomo High School Kokomo,IN Graduated: 2006 Student status: Alumni Degree: High School Diploma"

Purdue University West Lafayette
IN
Graduated
2010 Student status
Alumni Degree
Bachelor's Degree Major
Business/ Management 2002 to 2006Kokomo High School Kokomo
IN Graduated
2006 Student status
Alumni Degree
High School Diploma

Who knows about X? Who can influence A about X? Are there groups who like X?>>>>

xliii


Interest Graph Construction –Data Classification and Processing
Interest Graph Construction
Entity Data collection
Data Classification
"Interest" Recognition Phase
Interest Graph Update
Data Types
•Numeric
•Enumerated
•Classifiable
•Other
Who knows about X? Who can influence A about X? Are there groups who like X?>>>>


Interest Graph Construction –Interest Recognition Phase
Different techniques based on "data classification" and the nature of the data
Interest Graph Construction
Entity Data collection
Data Classification
"Interest" Recognition Phase
Interest Graph Update
Automic Data → Interest Vector
•Web Services
i.e. IMDB for movies, etc.
•NLP techniques
•"Recognition" algorithms
•Specific to data (image recognition).
Consider Context, Certainty and Importance
Who knows about X? Who can influence A about X? Are there groups who like X?>>>>

Interest Graph Construction –Interest graph Use

Interest Graph Use

Recommendation/ Advertisement

Example PPARS

•Result based on key Interests emphasized by interests in home, gender and age.

You and your friends Medha Singri, Rashmi G will like this

Who knows about X? Who can influence A about X? Are there groups who like X?>>>>

Challenges

- **Context/culture** (twitter example and more0
- **Expanding Recognition** options
- **Accuracy** **-** more (better) data, smaller graphs/more focus, different system algorithms, certainty and importance features, veracity, negative vs positive data, etc.
- **Performance** --- to the cloud and distributed – future mobile peer-to-peer selective interest based computation.

The Interest Graph Architecture-- Social Modeling and Information Fusion

Lynne Grewe[a]
[a]Computer Science, California State University East Bay (CSUEB), 2500 Carlos Bee Blvd, Hayward, CA, USA;

ABSTRACT

Social Networking and Modeling combined with concepts of Data Fusion will bring to fruition the development of "Interest Graph Modeling". This paper reviews issues and challenges in social data modeling and the problems in the development of a "Interest Graph" that goes beyond a simple "Social Graph". A generic architecture for the development of an "Interest Graph" is discussed along with some real world examples.

Keywords: Interest Graph, Social Modeling Network, Data Fusion

1. INTRODUCTION

In Social and Internet Communities, the new desired "model" being discussed is not the "Social Graph" but, now the "Interest Graph"[1]. The Interest Graph can be defined as the modeling of users/organizations/groups interests and their interactivity and influence. An Interest Graph can answer the question of "who knows X" and "who can influence A about X". These are powerful pieces of knowledge and allow a system to utilize what we call the social web for specific purposes, commercial or otherwise. The development of this "Interest Graph" can be done using data understanding and fusion techniques on social and other data.

In this paper, I will present a generic architecture for the creation of the Interest Graph, discuss some of the issues around social data capture, processing and understanding, highlighting the challenges that are faced at different architecture stages. Two systems using "social" data implemented by the author trending towards the "Interest Graph" are discussed.

2. THE OR AN INTEREST GRAPH?

An interesting question, is whether or not there is only one interest graph or many? To make something perform well you need to focus, which means not capturing all the interests that might be ever possible over all time by users across the entire world (wide web), but, instead focusing on creating a great graph that really describes the interests of your system. Fundamental components that any Interest Graph implementation will include are nodes which represent entities (users, groups, organizations) and interests (desires, feelings, skills or even activities) and edges that connect them. Some may say that an Interest graph should only contain information about "interests" and not demographic data – but, I allow for the inclusion of demographic data [1]. A connection between an entity node and an interest node represents the entities expression of this interest. A connection between two entity nodes represents some kind of direct connection in the social graph (i.e.. being friends). A connection between two interest nodes represents correlation between them. This graph can answer questions like: "Who knows about A" and "Who can influence X about A".

3. SOCIAL AND OTHER DATA

The expertise of this author extends the meaning of social data to what is found in social "networks" like Facebook, Twitter and MySpace[2] to the more focused data of user generated reports ("intelligence reports")[3]. Not directly social data such as geographic and physical information are potential inputs for Interest Graph construction. In the field of data fusion, there is a long history of understanding and fusion of such "physical" data. [4]

This paper focuses on social data from social networks which includes: interests, status, activities, work, education, movies, tv, smoking, drinking, looking for, gender, age, name, addresses (often private), current location, status, social status, kids, languages spoken, religion, ethnicity, living arrangements, sports, birthday, fashion, food, humor, posts, favorite song, movie, heroes, quotes and more. In this paper, I will focus on how to derive Interests from such Social data and will not focus on "activity-based" social data --tracking interactions or user engagement..

There are a number of problems with social data that are much less experienced that in "physical" data (like images, video) and include veracity of data, intelligibility of data, contextual/cultural issues and detection of negative interest. Data veracity is a problem with social data as people frankly lie or stretch the truth. Another problem with data is that of intelligibility. "OMG", "LoL"….these are the new language of online chat. It changes, and it can change with context and culture. A famous recent example, are the English tourists detained and deported for having a tweet about "tearing up the US" --- British slang for partying [5]. Bringing context into the picture is important and can be critical. Looking at the field of speech recognition and related natural language understanding, the context-based systems that are most successful are focused, specific in content and goals[6]. Finally, there are the concepts of data certainty and importance as discussed in [3].

4. INTEREST GRAPH ARCHITECTURE

The creation and use of an Interest Graph requires a system that goes through a series of constructive steps arriving at a general architecture that can be seen in Figure Y. Goal Understanding is the part of the Interest Graph Architecture that yields a set of Interests the system will incorporate into the building of its Interest Graph. This is not a simple task and can involve many techniques like statistical density modeling but, can also involve selection or direction by human experts. As discussed in [2], the PPARS system has the goal of advertising and the related taxonomies drive Interest selection. A very different set of interests may present for the goals of advertising compared to either performing disaster relief or the detection of intrusions or terrorists.

The Interest Graph Construction stage uses the Interest vector derived from the Goal Understanding TASK, and can (optional) use the Training Data from the Training Sample Creation TASK. It has a number of components that are discussed in below including data capture, classification, interest recognition phase and finally interest graph update.

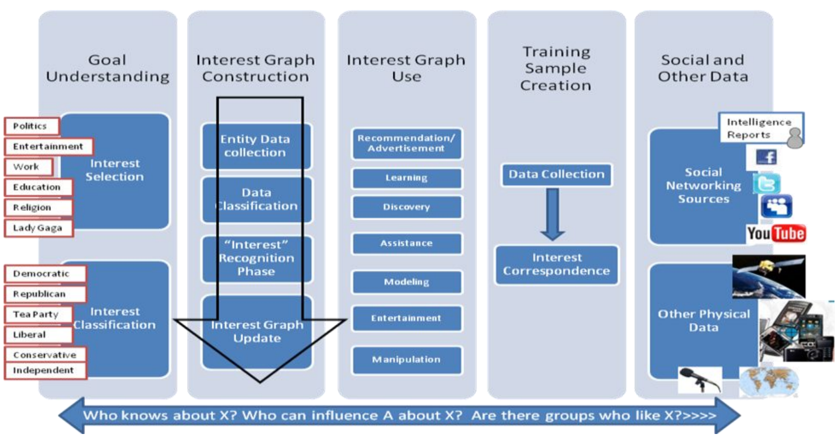


Figure Y. Interest Graph Construction and Use Architecture.

4.1 GOAL Understanding TASK --- Creation of the "Interest Vector"

GOAL Understanding yields the set of Interests our system will use and these may be known apriori by a goal expert as in [2], where the goal is advertisement recommendation and an advertising taxonomy is used with pre-determined social data types to form a set of Interests describing both advertisements and users (entities).

However, it is possible that the Interests are not immediately apparent to the system architect. In this case, it may be possible to come up with a large and broad set of possible Interests, and use a large set of training data to apply statistical metrics for selection of distinctive features in the large hyper-dimensional Interest space. Success using such metrics will correspond to training data quality, quantity and representation.

4.2 Interest Graph Construction: Segmentation, Data Classification and Interest Recognition

Before trying to figure out how social data maps to Interests, a system must first parse and segment the possibly long data set (narrative) into atomic semantically meaningful pieces of data. This is akin to performing image segmentation or parsing of sentences into phrases. This is a critical step and done wrong yields a poor Interest Graph. There is no one way to segment and many believe over-segmentation better than under-segmentation and vice versa [2]. Once a segmentation stage is performed, the (more) atomic information needs to be quantized so it can pass to the Interest Recognition phase. Sometimes the process of quantization and Interest Recognition can be combined as in [2].

As described in [2], the author hypothesizes that social data can be grouped into the classes of "numerical", "enumerated", "categorizable", and "other". Examples of numerical social data are age, zip code, address (can be translated to latitude and longitude). Enumerated data is one in which the number of possible values are stipulated by the social data provider. Examples of this include gender (male, female), smoker (no, yes, often, quit, occasionally), and drinker. Numerical and Enumerated data are easily translated into the numerical language that computer systems understand and that work best with techniques from data fusion to understanding and recognition that can construct our "Interest Graph". The last kind of data, I lump into "Other". This is data that it is hard find a taxonomy or recognition strategy for. Frankly, "Other" data is really data that the Interest Graph architects have chosen not to process due to lack of interest or the difficulty of processing it or reliably processing it.

Another type of social data is "categorizable" which means that there is a natural taxonomy that can be mapped to a set of categories. An example of this is "Movie" social data. A discussed in [2], the social movie data after a parsing (segmentation) process is passed to a movie web service like IMDB that will return search results including Genre information (Drama, Comedy, Action ,etc.). This is the common taxonomy of movies and one that is used for most purposes related to this kind of information –like search, advertisement, recommendation, etc. Other examples of social data that can be "categorizable" are ones in which there is not such a strong taxonomy as in our previous example. Examples of this kind of data include "interest", "work/jobs", "education", "status". While it can be supposed that the language of work/jobs and education are more focused than "interests" and "status", they all have the ability to be relatively unlimited in content. This leads us back to the discussion of being goal driven in the creation of our Interest Graph. For example in [2], the goal is one of advertisement recommendation based on finding groups of entities with shared interests that map to advertisements with similar interest profiles. Here the question is "What user grouping will like X (or knows about X)". As the goal is advertisement driven –we can look into our "interests" or "statuses" social data for information of use to our advertisers. In [2], rather than doing some generic NLP approach, we use a dictionary based approach to look for keywords and phrases of interest in the parsed data atoms. We suggest this approach as one that is sufficiently focused to achieve good results.

Expanding recognition techniques on "loosely" categorizable data is something that can take place. An example might be using image recognition to classify images containing people of certain demographics (age, gender), or you might take an image and try to recognize locations.

5. LARGE SCALE INTRESET GRAPH - PERFORMANCE AND FUTURE

Currently the trends for speed/performance boosts involve cloud and distributed systems. Looking to the future, the author feels that the mobile devices with ever more processing power may yield the way for peer-to-peer systems where not only social data is created but, processing could be done on these mobile nodes. Beyond performance there is the possibility of node goal based processing --- processing with peers with related location or currently known interests.

REFERENCES

[1] How Twitter Is Pairing Its Interest Graph With Ads http://techcrunch.com/2012/03/01/how-twitter-is-pairing-its-interest-graph-with-ads/

[2] L. Grewe and S. Pandey, "Quantization of Social Data for Friend Advertisement Recommendation System", PCDTA (Parallel, Distributed Computing Technologies and Applications), Chennai, India, 596-614, 2010.

[3] L. Grewe, S. Krishnagiri, J.Cristobal, "Metrics of a System for Disaster Relief", IEEE VECIMS08, pp. 45-50, 2008.

[4] R. Brooks and L. Grewe, "Data Registration", Distributed Sensor Networks, 2nd Edition, Chapter 21, 2012

[5] US bars Friends of Twitter Joke, http://www.thesun.co.uk/sol/homepage/news/4095372/Twitter-news-US-bars-friends-over-Twitter-joke.html, 2012.

[6] Nuance Corporation, Speech Recognition solutions, http://www.nuance.com/

Issues and Challenges in Intent Modeling in Social/Cultural Networking Domain

Ivan Kadar
Interlink Systems Sciences, Inc.
Lake Success, NY, USA

24 April 2012

Invited Panel Discussion on "Real-World Issues and Challenges in Social/Cultural Modeling with Applications to Information Fusion"

SPIE Conference 8392 "Signal Processing, Sensor Fusion and Target Recognition XXI", April 23-25, 2012, Baltimore, MD

Problem Statement & Proposed Methods

- Information Fusion/Situation-Threat-Intent-Assessment (SA), (JDL Levels 2/3) and the human role in intent modeling
  - Human aspect is addressed as a key input component to intent modeling in SA
- Given Social/Cultural setting coupled with Social Networks (SNs) derived global real-time information: - How to detect/identify impending intent of populations in various geographical regions of interest to assess the situation?
- Components Addressed:
  - Definitions of Intent and SA
  - Representative Intent models; the need/type/apps of a Cognitive model
  - Social Network Information Exchange, extracted contextual information as input to the intent model
  - Social Network sites
  - Real-world example of Twitter based/extracted information and its apps
  - The cognitive Perceptual Reasoning Machine and its use to detect/identify impending intent
- Challenges: Research, Implementation and Testing of the proposed methods

Definition of Intent, where it belongs in Situation Assessment and how to Model

- *Intent* (and by implication goals) can be viewed as the determination or resolve to do a certain thing, or the state of mind with which something is done. That is, the "notion of *intent* revolves around the ideas of aim, will, goal, target, objective, plan, and *purpose*". *Purpose* is defined as : "an anticipated outcome that is *intended* or that guides ones' planned actions."
- The above definitions clearly show that *"intent"* is *not directly observable:*
  - *Intent* is an intangible concept. It must be inferred from unexpected observations events/processes [1].
- Previous efforts towards the definition of a *model for intent* have included:
  - A military perspective on commander's intent, a belief-desire-intent framework, planning-based models of *intent*, explicit & implicit *intent*, & subject matter experts.
- Having a desire alone does not allow the execution of *intent*.
  - One needs an *opportunity* which is a favorable juncture of circumstances for taking actions. *Opportunities* make it possible to carry out one's intent given sufficient capabilities.
  - That is, an *opportunity* is the presence of an operating environment in which potential targets of an action are present and are susceptible to being acted upon

[1] E. Bosse, J. Roy, and S. Wark, *Concepts, Models, and Tools for Information Fusion,* Artech House, Inc. 2007.

Definition of Intent, where it belongs in Situation Assessment and how to Model

- Per *Steinberg, "Threat Assessment involves assessing situations to determine whether detrimental events are likely to occur"
  - That is, the Level 3 JDL Data Fusion Process which has been broadened to "Impact Assessment"
- One can decompose *threat* into *capability*, *opportunity,* and *intent* as principal factors in predicting (intentional) actions.
  - *Capability* involves an agent's physical means to perform an act;
  - *Opportunity* involves spatial-temporal relationships between the agent and the situation elements to be acted upon;
  - *Intent* involves the will to perform an act.
- Given the definitions of intent, a cognitive-like adaptive learning paradigm appears most suitable to model it.
- Why a cognitive-like model ?
  - One needs to model not only to recognize impending *threat/intent* given an *opportunity*, but also need to identify the potential for unexpected spoofing, misleads and deceptions based on prior learned information

Alan N. Steinberg, "Foundations of Situation and Threat Assessment", Chapter 18 of Handbook of Multisensor Data Fusion,eds. Martin E. Liggins, David L. Hall and James Llinas, CRC Press, London, 2009.

Models of Intent – Issues and Challenges

- The capability to "sense/observe, mine/*access* data, associate, learn, recall, anticipate and predict/act" are key ingredients of ***human perceptual reasoning*.** These attributes are necessary constructs in ***cognitive modeling.***
  The key ingredient is timely information access in modeling intent.
  - ***Cognitive models*** are imbedded in large family of methods called **Predictive Analytics/Modeling** (techniques to predict future entities) including:
    - sensing/collecting, sorting, organizing, aligning, associating, fusing, data mining; and using a-priori and learned, SME based and current data
  - **Predictive models** *using algorithms, such as:* regression, neural and abductive networks, learning, classifiers, distance measures, Bayes-nets/ influence diagrams, logic, decision making under uncertainty,…, *have been used in intent modeling, but did not use a cognition framework, which incorporates the above methods/algorithms,* called the ***cognitive PRM paradigm***:

  - The ***Perceptual Reasoning Machine (PRM)*** [2]: a "meta-level information management system", for adaptive information gathering/assessment, learning, anticipation, and *prediction*.
- **Objectives of models**? : minimize uncertainty and maximize the value of deduced information to *detect/identify potential intent*, and to act in a real-time environment with time constraints

[2] I. Kadar, "Data Fusion by Perceptual Reasoning and Prediction", *Proceedings of the First Tri-Service Data Fusion Symposium*, John Hopkins University, APL, Laurel MD, June 9-11, 1987. Original Paper - Extensions/apps/pubs to 2010.

Cognitive Models of Intent – Issues and Challenges via Social Networking for Information Access

- What is the Role of Social Networking in the PRM intent model framework?
  - Social Networks (SNs) **provide access to real-time information exchange** [derived* **context (e.g., sentiments, emotions**), extracted from cultural/social interactions - messages with location and time stamped data] to be used as input to the model

    Potential Issues and Challenges:

  1. Is the extracted data based on consensus of the population or only from "outliers"? ("outliers" can exert ***influence,* coalesce and become significant intent indicators). Furthermore, how to handle potential data sparsity (per individual) vs. enormity (web) of data; and contextual validity into emotional aspects?
  2. Is information exchange restricted globally by particular entities? (potential direct intent indicators)
  3. How to "*associate*" massive information from multiple SNs as input to PRM ?

*** Note: The preprocessing of linguistic messages to learn, classify and group various context is assumed a given herein.**

****W. Pan, W. Dong, M. Cebrian, T. Kim, J. H. Fowler and A. (Sandy) Pentland, "Modeling Dynamical Influence in Human Interaction", *IEEE Signal Processing Magazine,* March 2012.**

li

Cognitive Models of Intent – Issues and Challenges via Social Networking for Information Access (Cont'd)

- ***Information access*** is crucial as an input both for real-time assessment, prediction and to data bases (learning) & for *message rate "change detection"* impending intent?

  - Example: [3] Ben Zimmer, "Twitterology: A New Science?", *The New York Times,* October 30, 2011.The article illustrates the degree of relevant real-time information that can be derived from social/cultural interactions expressed in Twitter:

  - The extracted Twitter context information from messages can be used as ***input to intent modeling:***
    sentiments, emotions - **moods, opinions* etc. extracted context data** (**including *locations, time, consensus types, groups and number of constituting elements or computed probabilities***) used as **input with other data sources to detect/ID potential intent via the cognitive PRM model.**
  - the article is illustrated, in part, op. cit. in subsequent viewgraphs

***A.Pak, and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, Valletta, Malta.**

Cognitive Models of Intent – Issues and Challenges via Social Networking Sites

Representative Social Networking (SN) Sites:

- Facebook currently is largest SN site - 800 mil active users - It has been used by over several million people during the Egyptian Revolution*.
- Twitter currently is next largest SN site -100 mil+ active users.
  In Twitter the information is publicly accessible and searchable (globally), while the others have a more private system – need to sign up. Twitter has also been used by several million people during the Arab Spring**. Its use and apps is illustrated herein.
- MySpace is another network that was very popular a few years ago but has been abandoned in favor of the aforementioned two.
- LinkedIn - a business-focused site.
- Google Plus - Google's input to social networking.
- There are thousands of sites out there, a lot of them cater to specific interests, others are more general.

***J. A. Vargas, "How an Egyptian Revolution Began on Facebook", *The New York Times*, February 17, 2012**

****B. Zimmer, "Twitterology: A New Science?", *The New York Times,* October 30, 2011**

Cognitive Models of Intent use of Twitterology op.cit [3]

Twitter is many things to many people, but lately it has been a gold mine for scholars in fields like linguistics, sociology and psychology who are looking for real-time language data to analyze.

Twitter's appeal to researchers is its immediacy — and its immensity. Instead of relying on questionnaires and other laborious and time-consuming methods of data collection, social scientists can simply take advantage of Twitter's stream to eavesdrop on a virtually limitless array of language in action.

At the University of Texas, for example, a group of linguists and social psychologists has been monitoring Twitter to track on-the-ground sentiment over the course of the Arab Spring, particularly in Egypt and Libya. *After the death of Colonel Qaddafi, the linguist David Beaver and his assistants quickly summoned thousands of Arabic-language tweets before and after the event. They zeroed in on messages known to be from Libya by using Twitter's system of geocoding. (Posts from cellphones, for instance, very often encode the user's geographic coordinates.) The tweets were then automatically translated from Arabic to English and fed into a text-analysis computer program.*

The researchers were able to create a dynamic portrait of Libya's Twitter traffic. The overall traffic skyrocketed in the hours after Colonel Qaddafi's death was announced, as did terms related to positive sentiment like "good" and "wonderful." Religious sentiment was also on display, with a significant increase in the frequency of words like "Allah," "sacrifice" and "gospel."

Cognitive Models of Intent – Use of Twitterology op.cit [3]

In this burgeoning field of Twitterology, moods are also being gauged on a more global level. Two sociologists at Cornell University, Scott A. Golder and Michael W. Macy, recently published a study in the journal Science that looked at how **emotions** may relate to the rhythms of daily life, across many English-speaking countries. They observed a gradual falloff in positive terms from the beginning of the workday, bottoming out in the late afternoon.

One criticism of "sentiment analysis," as such research is known, is that it takes a naïve view of emotional states, assuming that personal moods can simply be divined from word selection. This might seem particularly perilous on a medium like Twitter, where sarcasm and other playful uses of language often subvert the surface meaning.

James W. Pennebaker, a social psychologist at the University of Texas who pioneered the text-analysis program often used in this kind of research, warns that positive and negative emotion words are the "low-hanging fruit" in such studies, and that deeper linguistic analysis should be explored to provide a "richer, more nuanced view" of how people present themselves to the world.

But even if we can't expect Twitter to be an unerring emotional barometer, it is proving extremely valuable for understanding how language varies among different demographic groups. A team of computational linguists at Carnegie Mellon University led by Jacob Eisenstein and Brendan O'Connor has used geocoded tweets to build maps of regional language use across the United States. The amount of data available for analysis is many orders of magnitude bigger than what could be collected with traditional dialect surveys.

Cognitive Models of Intent – Using Twitterology op.cit [3]

From these mountains of data can be gleaned hidden patterns of informal English, like the profusion of *hella,* as a form of emphasis in Northern California, as in, "It's hella cold out there." Slangy phonetic spellings also show distinct patterns of distribution, with New Yorkers preferring *suttin* to *sumthin* (for *something*) and Californians writing *koo* or *coo* for *cool*. Even emoticons differ from region to region.

This study attracted negative attention in 2011 from Senator Tom Coburn of Oklahoma, who listed it as one of the "questionable" projects financed by the NSF in a report challenging the foundation's budget for the social sciences. **But the research was vigorously defended by Randal E. Bryant, dean of Carnegie Mellon's School of Computer Science, *who pointed to its real-world applications. "The key finding was that seemingly meaningless slang and jargon can reveal important properties of the author's identity, a point of interest for both corporations and the intelligence community,"* Mr. Bryant said.**

Still, the Twitterologists will continue to have a tough row to hoe in justifying their research to those who think that Twitter is a trivial form of communication. No less a figure than Noam Chomsky has taken Twitter to task recently for its "superficiality."

"It is not a medium of a serious interchange," Mr. Chomsky said, a blanket charge that ignores the diversity of voices to be found on Twitter. Regardless of how unserious Twitter exchanges may appear on the surface, many of Mr. Chomsky's fellow linguists are discovering that Twitter can help uncover truths about our social interactions that are quite serious indeed.

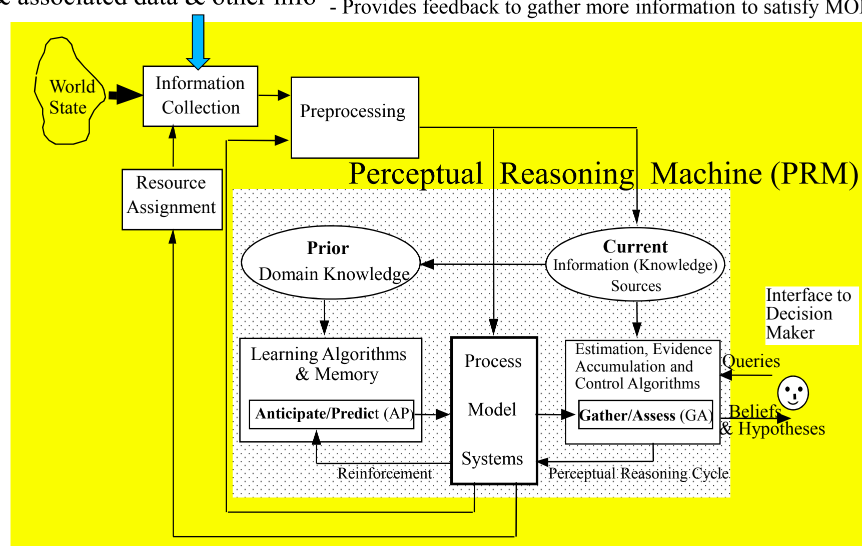
[3] Ben Zimmer, "Twitterology: A New Science?",*The New York Times,* October 30, 2011

Generic Information Process Model via PRM

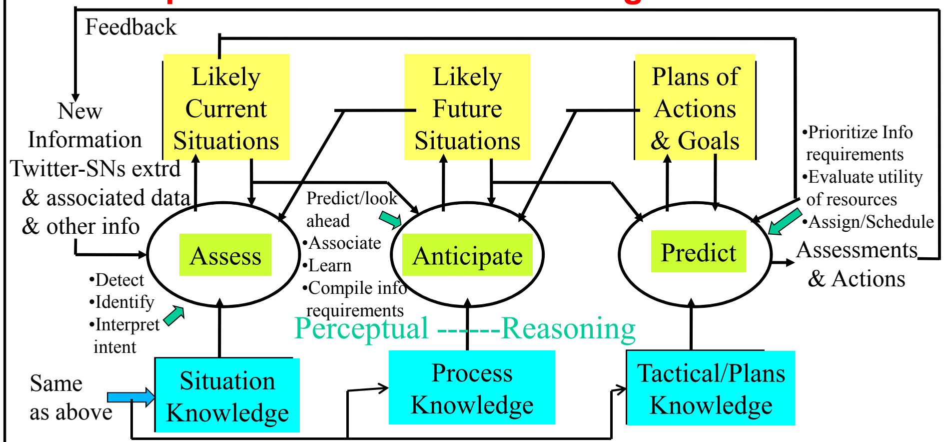
Issues addressed

- Use of *Knowledge* – Priori, Learned and Current
- Use of *Process* – gather facts - anticipate/predict
- User queries and Fusion System presents Beliefs
- Provides feedback to gather more information to satisfy MOMs

Twitter extracted & SNs extrtd & associated data & other info



Example: Information Flow Among PRM Elements



The "**Assess**/**(gather)"** module responding to "dynamically managed" received multisource information uses additional information from its **KB** and from the **"anticipate**" module to form a database of **likely current situations** which include potential detected intent & threats.

The "**assess**" module provides **likely current situations** information to the **"Predict/(preplan/act)"** module, which along with its **KB** and **likely future situations** information from the "**anticipate"** module, issues assessments, identifies potential intents/threats, provides **plans of actions and goals**, and as needed, request actions for additional information to confirm or negate conflicting hypotheses - **"closing the loop".**

The **"Anticipate**" module provides information to the "**predict module"** on expected **likely future situations/intent** for short and long-duration planning based on **likely current situations/intent** from the "**assess**" module which is updated with prior, adaptively learned, process, intent/tactical knowledge and associated likely hypotheses.

Summary

- Questions and Comments?
- There are many issues and challenges remaining requiring research, implementation, testing to validate the proposed methods.

  - However, PRM-based models have been shown to converge faster to a solution, than non-cognitive models. Therefore, the PRM model is expected to perform well, and can also be used in many other apps.
  Addressed:
- Definition of Intent
- The role of Intent in Situation Assessment
- Models of Intent
- The need for Cognitive Models of Intent - Issues and Challenges
- The Importance of Information Exchange as Input to Cognitive Models, viz., Perceptual Reasoning Machine (PRM)
- Social Networking Sites providing information exchange
- Type of Social Networking Sites
- The role of Twitter – "Twitterology" inputs derived for use in PRM
- The Generic Information Process Model via PRM
- The Cognitive Perceptual Reasoning Machine Paradigm Information Flow

Issues and Challenges in Intent Modeling in Social/Cultural Networking Domain

Ivan Kadar
Interlink Systems Sciences, Inc.
1979 Marcus Avenue, Lake Success, NY 11042

1. INTRODUCTION

Problem Statement: Information Fusion /Situation-Threat-Intent Assessment (SA), (JDL Levels 2/3), and the human role in intent modeling: the human aspect is addressed as a key input component to intent modeling in (SA). Issues and Challenges: Given Social/Cultural setting coupled with Social Networks (SNs) derived global real-time information, how to detect/identify impending intent of populations in various geographical regions of interest to assess the situation.

This succinct position paper, coupled with the associated viewgraphs, highlights issues and challenges of the complexity of the meaning (i.e., definition) of intent and its modeling, specifically in the (SA) setting in the social/cultural networking domain. New methods are introduced to model the potential of intent using Social Networks (SNs) Web extracted information (messages). The use of SNs is introduced with real-world application examples showing the potential to extract contextual data, such as sentiments and emotions, tagged with location and time information of the user population under particular situations of interest. Specific example of the apps of Twitter is illustrated. Contextual information is the key ingredient used to infer potential intent, which can be reinforced by associated other information (e.g., message traffic change, sensors and human observers). Given the above depicted data sets, the contextual information can be further divided into groups to examine whether or not the data represents a large percentage of the population or only "outliers". Methods to handle these data sets are described.

Next the definition and potential models of intent are reviewed. The focus is to use cognitive models, not because one is dealing with a social cultural setting, rather because cognitive models emulate human perceptual reasoning, which has the capability to use: incoming information, prior knowledge and recall, learn, reinforce prior knowledge in a positive or negative reinforcement sense based on incoming and learned information, anticipate and predict. Therefore, this type of model is expected to converge faster to a solution, for given hypotheses, than non-cognitive models.

The cognitive Perceptual Reasoning Machine (PRM) paradigm [1-10] is introduced as a method for potential intent modeling. The functions of the PRM are introduced within the description of the General Process Model system based on the PRM. The information flow among the PRM elements is described.

2. DEFINITIONS OF INTENT, ROLE IN SITUATION ASSESSMENT AND MODELS

As one can gather from the complex definitions of *Intent*, which also implies *goals*, that it can be viewed as the determination or resolve to do a certain thing, or the state of mind with which something is done [10]. That is, the "notion of *intent* revolves around the ideas of aim, will, *goal*, target, objective, plan, and *purpose*". The concept of *intent* has always been at the root of some of tort law's most basic categories [11]. *Purpose* is defined as: "an anticipated outcome that is *intended* or that guides ones' planned actions, viz., the commander's *intent* was to execute immediate action with a conscious aim...". The above definitions clearly show that *"intent"* is *not directly observable:* Reference [10, 12, 13] notes: *Intent* is an intangible concept that cannot be directly observed by sensors. It must rather be inferred from other data that becomes *indicators of intent*, e.g., unexpected observations events/processes. Having a desire alone does not allow the execution of *intent*. One needs an *opportunity*. *Opportunities* make it possible to carry out one's intent given sufficient capabilities. That is, an *opportunity* is the presence of an operating environment in which potential targets of an action are present and are susceptible to being acted upon [13].

In the context of SA, associated with intent is "Threat Assessment [10, 14], which involves assessing situations to determine whether detrimental events are likely to occur". That is, the Level 3 JDL Data Fusion Process [15] which has been broadened to "Impact Assessment". One can decompose *threat* into *capability, opportunity and intent* as principal factors in predicting (intentional) actions (see details in viewgraphs). Given the definitions of intent, cognitive-like adaptive learning paradigm appears most suitable to model it.

Why a cognitive-like model? One needs to model not only to recognize impending t*threat/intent* given an *opportunity*, but also need to identify the potential for unexpected spoofing, misleads and deceptions based on prior learned information [1-10].

The capability to "sense/observe, mine/*access*" data, associate, learn, recall, anticipate and predict/act" are key ingredients of *human perceptual reasoning*. These attributes are necessary constructs in *cognitive modeling. The key ingredient is timely information access in modeling intent.* *Cognitive models* are imbedded in large family of methods called Predictive Analytics/*Modeling* (techniques to predict future entities) including: sensing/collecting, sorting, organizing, aligning, associating, fusing, data mining; and using a-priori and learned, SME based and current data. *Predictive models have been used in intent modeling, but did not use a cognition framework, which includes many well known algorithms (please see viewgraphs for additional details)* including the *cognitive PRM paradigm*: the cognitive *Perceptual Reasoning Machine (PRM)* [1-10]: a "meta-level information management system", for adaptive information gathering/assessment, learning, anticipation, and *prediction*. Objectives of models are to minimize uncertainty and maximize the value of deduced information to *identify potential intent*, and to act in a real-time environment with time constraints.

3. SOCIAL NETWORKS AND USE OF COGNITIVE INTENT MODEL

Social Networks (SNs) provide basis for information exchange, in a social-cultural setting, allowing exchange/expression of information and enabling extraction of context, such as: ideas, concerns, sentiments, emotions, and opinions. The extracted context information can be used in intent modeling to assess the potential of impending intent collected in real-time via the Web. As depicted in the accompanying viewgraphs [16], before SNs, "global" information exchange was not possible. The new medium appeals to researchers because its immediacy — and its immensity [16, 17]. For example, as depicted in the viewgraphs, instead of relying on questionnaires and other laborious and time-consuming methods of data collection, social scientists can simply take advantage of Twitter's stream to eavesdrop on a virtually limitless array of language in action. This directly implies the relevance of SNs, as illustrated from results using Twitter [16], to use the information exchange of social/cultural messages to extract context data as input to intent modeling and associated prediction in given settings/on-going processing as noted before. As a matter of fact, by continually monitoring areas of interest one can use "change detection" in the rate of message traffic, along with the associated extracted context, to assess potential for impending intent. The viewgraphs describe differences among representative SNs.

What is the Role of Social Networking in the PRM intent model framework? Social Networks (SNs) provide access to information exchange [derived* context (e.g., sentiments, emotions), extracted from cultural/social interactions - messages with location and time stamped data] to be used as input to the model.
Potential Issues and Challenges: (1) Is the extracted data based on consensus of the population or only from "outliers"? ("Outliers" can exert *influence* [18], coalesce and become significant intent indicators). Furthermore, how to handle potential data sparsity (per individual) vs. enormity (web) of data; and contextual validity into emotional aspects? (2) Is information exchange restricted globally by particular entities? - (represents potential intent); (3) How to "*associate"* massive information from multiple SNs as input to PRM?
Information access is crucial as an input both for real-time assessment, prediction and to data bases (learning) & for *message rate "change detection"* - impending intent?
An example: [16] Ben Zimmer, "Twitterology: A New Science?", *The New York Times*, October 30, 2011.The article illustrates the degree of relevant real-time information that can be derived from social/cultural interactions expressed in Twitter from several point of view (see partial op.cit of the article in the viewgraphs).That is, Twitter extracted information from messages can be used as *input to intent modeling:* - such as sentiments, emotions – moods , opinions [19], etc. based extracted data (including *locations, time, consensus types, groups and number of constituting elements or computed probabilities)* used as input with other data sources to assess potential intent via the cognitive PRM model.
*Note: The preprocessing of linguistic messages to learn, classify and group various context is assumed a given herein.

4. THE PERCEPTUAL REASONING MACHINE PARADIGM

Viewed as a "meta-level information management system", PRM consists of a feedback planning/resource management system whose interacting elements are: "assess", "anticipate" and "preplan/act" [1-10]. That is:

- *Gather/Assess* current, *Anticipate* future (hypotheses by learning), and *Preplan/Act* (predict) on information requirements as well as likely intent and threats,
- *Anticipate/Predict (Plan)* the allocation of information/sensor/system resources and acquisition of data through the control of a specific distributed multisource sensors/systems resource manager (RM),
- *Interpret and Act* (shared by above functions) on acquired (sensor, spatial and contextual) data in light of the overall situation by interpreting conflicting/misleading information to either identify or rule out the potential or existence of intent.

The elements of the fundamental PRM construct are shown in Figure 1, below, depicting the interrelations among the constituting elements described above, providing adaptive information gathering (e.g., fusion) learning, anticipation, assessment, prediction and control.

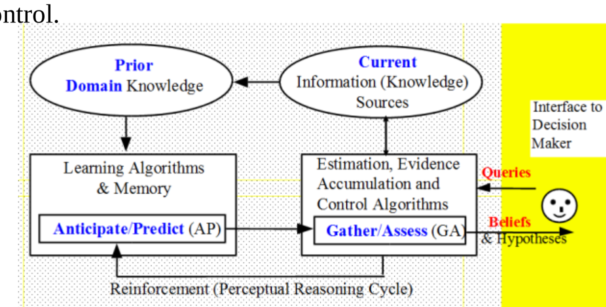


Figure 1: The Perceptual Reasoning Machine

It should be noted that the *current information* noted in Figure 1 can be derived from processed information collection which is can be controlled by a systems/sensors resource manager by feedback from the PRM [1-10]. This function is illustrated in the viewgraph, depicting the Generic Information Process Model system application of the PRM (the shaded part in the viewgraph is the PRM) [1-10]. Process modeling is defined as a set of procedures and algorithms that capture the functional and required (temporal and spatial) dependency relationships of tasks (e. g., needed for intent/threat assessment) and/or processes, which are being modeled.

The PRM information flow management elements and their relationships are depicted in the viewgraph entitled "Information Flow Among the PRM Elements" along with the knowledge requirements for each PRM function. Referring to the slide, the "assess module," responding to dynamically managed and received multisource information, uses additional information from its associated knowledge base and from the "anticipate module" to form a database of "likely current situations" which include potential intents/threats. The "anticipate module" provides information on "likely future situations" that are used for short- and long-duration planning. This planning is based on the "likely current situations" from the "assess" module; prior, learned, process and tactical/planned knowledge and associated hypotheses. The "likely current situation" information is fed back to the "predict module", which provides "plans of actions and goals". The "assess module" also provides current situations information to the "predict module" which, along with its knowledge base and likely future situations information from the "anticipate module", (based in part on associated process knowledge), issues assessments, identifies potential intents/threats, and as needed, request actions from the resource manager for additional information to confirm or negate conflicting hypotheses thus closing the outer loop via the systems/sensors manager.

SUMMARY

Methods were presented along with proposed solutions of models for intent detection/identification at fusion Levels 2/3 "Information Fusion/ Situation-Threat-Intent Assessment" (SA) baseline addressing the human role in intent modeling. The human role and associated information/sentiment was identified as a key input component to intent modeling in SA. Specifically, given the Social/Cultural setting coupled with Social Networks (SNs) derived global real-time information was addressed, i.e., how to detect/indentify impending intent of populations in various geographical regions of interest to assess the situation. The key element identified is the use of the cognitive Perceptual Reasoning Machine based intent model, using in part SNs (e.g., Twitter) message traffic information exchange extracted context data as input to the model. The PRM-base intent model is expected to provide enhanced assessment of intent/threat as it emulates human perceptual reasoning, which has the capability to use: incoming information, prior knowledge and recall, learn, reinforce prior knowledge in a positive or negative reinforcement sense based on incoming, learned information, anticipate and predict. There are many issues and challenges remaining requiring research, implementation and testing of the proposed methods.

REFERENCES

[1] I. Kadar, "Data Fusion by Perceptual Reasoning and Prediction", *Proceedings of the First Tri-Service Data Fusion Symposium*, Johns Hopkins University, APL, Laurel, Md., June 9-11, 1987.

[2] I. Kadar, "Perceptual Reasoning Managed Situation Assessment and Adaptive Fusion Processing," *Proceedings of the Signal Processing, Sensor Fusion and Target Recognition Conference*, Ivan Kadar, Editor, Proc. SPIE Vol. 4380, Orlando, FL, 2001.

[3] I. Kadar, "Perceptual Reasoning in Adaptive Fusion" *Proceedings of the Signal Processing, Sensor Fusion and Target Recognition Conference,* Ivan Kadar, Editor, Proc. SPIE Vol. 4729, Orlando, FL 2002.

[4] I. Kadar, "Knowledge Representation Issues in Perceptual Reasoning Managed Situation Assessment", Invited Panel Session on "Issues and Challenges of Knowledge Representation and Reasoning in Situation Assessment (Level 2 Fusion), Organizer: Ivan Kadar; Moderators: Ivan Kadar and James Llinas; *Proceedings of the 8th International Conference on Information Fusion*, July 25-29, 2005, Philadelphia, PA

[5] I. Kadar, "Issues in Adaptive and Automated Information Fusion Resource Management", Invited Panel Session on "Issues and Challenges in Resource Management (and its Interaction with Levels 2/3 Fusion) with Applications to Real World Problems", Organizer: Ivan Kadar; Moderators: Ivan Kadar and John Salerno, *Proceedings of the 9th International Conference on Information Fusion,* 10-13 July 2006, Florence, Italy.

[6] E. P. Blasch, I. Kadar, J. J. Salerno, M. Kokar, S. K. Das, D. Corkill, G. M. Powell, E. H. Ruspini, "Issues and Challenges in Knowledge Representation and Reasoning Methods in Situation Assessment (Level 2 Fusion)", *Proceedings Signal Processing, Sensor Fusion and Target Recognition XV,* Ivan Kadar Editor, Proc. SPIE Vol.6235, April 2006.

[7] E. P. Blasch, I. Kadar, J. Salerno, M. M. Kokar, S. Das, G. M. Powell, D. D. Corkill, E. H. Ruspini, "Issues and Challenges in Situation Assessment (Level 2 Fusion)", *Journal on Advances in Information Fusion (JAIF),* Vol. 1, No. 2, December 2006.

[8] I. Kadar, **"**Results from Levels 2/3 Fusion Implementations: Issues, Challenges, Retrospectives and Perspectives for the Future - An Annotated View**",** Invited Panel Session on "Results from Levels 2/3 Fusion Implementations: Issues, Challenges, Retrospectives and Perspectives for the Future**",** Organizer: Ivan Kadar; Moderators: Ivan Kadar and John Salerno, *Proceedings of the 10th International Conference on Information Fusion*, 9-12 July 2007, Quebec City, Canada.

[9] E. Blasch, I. Kadar, J. Salerno, K. Hintz, J. Biermann, C. Chong, S. Das, "Resource Management Coordination with Level 2/3 Fusion Issues and Challenges", *IEEE A & E Systems Magazine,* March 2008.

[10] J. Salerno, S. J. Yang, I. Kadar, M. Sudit, G. P. Tadda, J. Holsopple, "Issues and Challenges in Higher Level Fusion: Threat/Impact Assessment and Intent Modeling (A Panel Summary)" *13th International Conference on Information Fusion*, Endinburgh, Scotland, 26-29 July 2010.

[11] A. J. Sebok, "Purpose, Belief & Recklessness: Pruning the *Restatement (Third)'s* Definition of Intent*", Vanderbilt Law Review*, Vol, 54:3:1165

[12] E. Bosse, J. Roy, and S. Wark, *Concepts, Models, and Tools for Information Fusion,* Artech House, Inc.2007.

[13] J. Roy, "A View on Threat Analysis Concepts", *Proceedings of the 12th International Conference on Information Fusion (Fusion 2009),* Seattle, WA.

[14] A. Steinberg, "Foundations of Situation and Threat Assessment", *Chapter 18 of Handbook of Multisensor Data Fusion,* eds. Martin E. Liggins, David L. Hall and James Llinas, CRC Press, London, 2009.

[15] A. Steinberg, C. Bowman, and F. White. Revisions to the JDL Data Fusion Model, presented at the Joint NATO/IRIS Conference, Quebec. October 1998

[16] B. Zimmer, "Twitterology: A New Science?", *The New York Times,* October 30, 2011

[17] J. A. Vargas, "How an Egyptian Revolution Began on Facebook", *The New York Times*, February 17, 2012

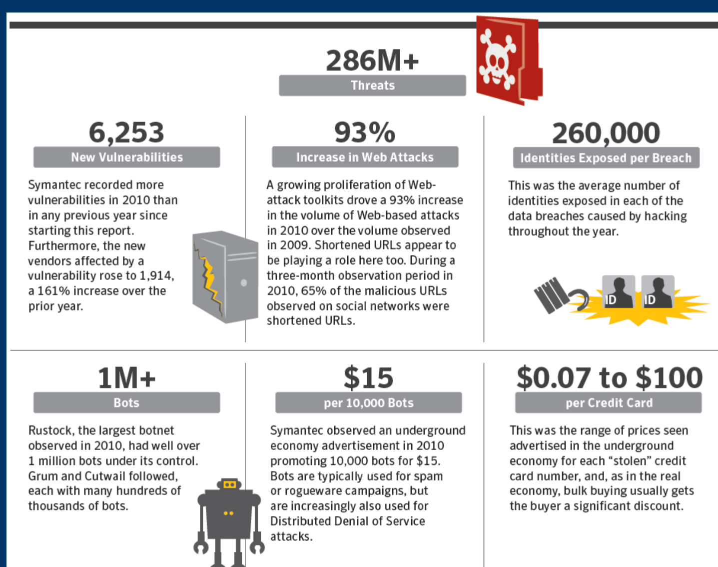
[18] W. Pan, W. Dong, M. Cebrian, T. Kim, J. H. Fowler and A. (Sandy) Pentland, "Modeling Dynamical Influence in Human Interaction", *IEEE Signal Processing Magazine,* March 2012.

[19] A. Pak, and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010),* Valletta, Malta.

Characterizing Large-scale Attack Behavior for Predictive Cyber Situation Awareness

S. Jay Yang
Associate Professor & Department Head
Department of Computer Engineering
Rochester Institute of Technology
Email: jay.yang@rit.edu
http://www.ce.rit.edu/people/yang

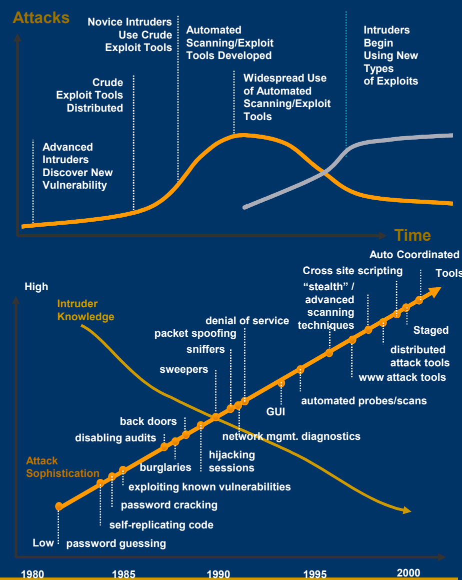
Cyber Attacks 2010 in Numbers [Fossl2010]



Understanding the Hackers

[Fuchsberger'05]

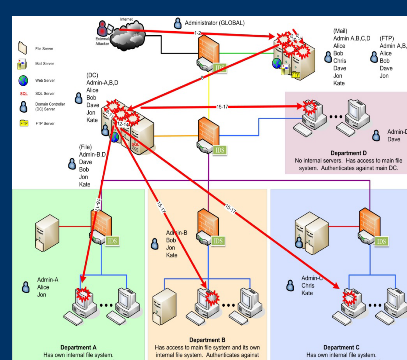
- Cyber attack life cycle
- Sophistication of world-class hackers
- Volume of attacks increases drastically due to easy access of hacking tools and well educated coders
- Hardening security and intrusion detection are almost always steps behind
- Advanced persistent, multistage and coordinated attacks



Cyber Fusion for Large-Scale Cyber Attacks

Deducing complex cyber attack relationship from drowning data

- Alert Correlation
  - Aggregation
  - Mapping to attack scenarios
- Attack Characterization
  - Impact Assessment
  - Threat prediction
- Attack Clustering
  - Identify similarly behaving attacks
  - Treat coordinated/colluding attacks

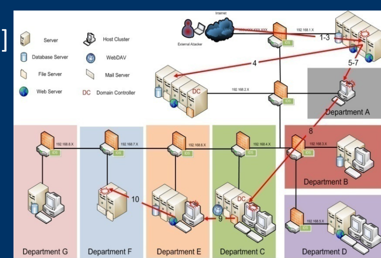


- Cyber fusion to extract features of adversary behavior
  - Spatial: where and what target (e.g., IP and service)
  - Temporal: when and in what order attack actions are executed

From Intrusion Detection to Projection

With overwhelming alerts…

- Alert aggregator/correlator forms attack tracks
  - E.g., [Cuppens2002] [Xu2004] [Valeur2004] [Stotz2007]
- Estimate impact of observed attack tracks
  - Rule/scenario based [Ning02] [Porras02] [Valeur04]
  - Network based [Argauer08]
- Project attack actions
  - Matching attack plans [Qin04] [Arnes06]
  - Recommendation systems [Soldo11]



TBM Fused Capability & Opportunity

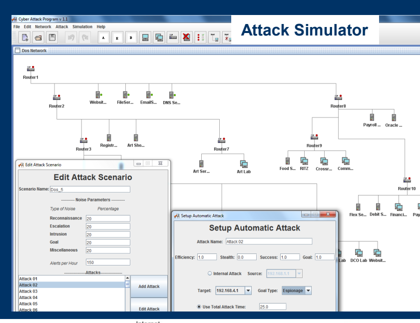
- Capability:
  - Aims at estimating adversary ability to exploit vulnerabilities
  - Assumptions:
    - An attacker is capable of attacking same services he/she has attacked, even by exploiting different vulnerabilities
  - Statistical profiling
- Opportunity:
  - Aims at finding what opportunities are available to "red" given blue's estimate of red's progression on the operation environment
  - Breath-first search of neighbors "exposed" from compromised machines – subject to firewall and routing rules
- Need to combine the estimates
  - Hacker needs capability to explore opportunity
  - Transferrable Belief Model vs. Context-Specific Fusion

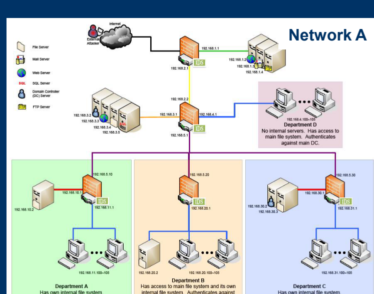
F-VLMM to Capture Behavior Trend

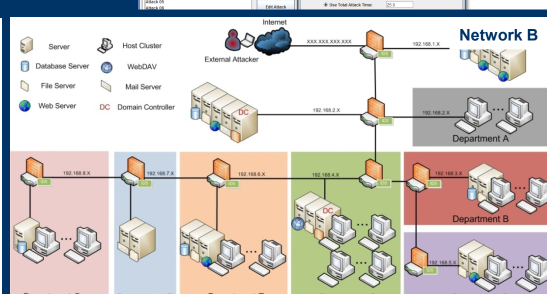
- Aims at finding adversary patterns due to
  - Routines, habits, human preference
  - Uses of toolkits, …
- Approaches:
  - Statistical profiling
  - Adaptive Bayesian Network, e.g., [Qin04] [Arnes06]
    - Model structure needs to be defined (o.w. model space too large)
  - Recommendation systems [Soldo11]
    - Borrowed ideas from movie ranking and online shopping
  - Variable Length Markov Model (VLMM)
    - Effective graphical model to combine various orders of Markov Models – from text compression community
    - Fuzzy system to fuse VLMM predictions based on different attack attributes, e.g., target IP and attack method

Experiment Setup

- Network configurations
  - NetA – specialized servers
  - NetB – larger & more duplicate services
- Attack generation
  - Simulator [Kuhl07]
  - Efficiency
  - Stealthy

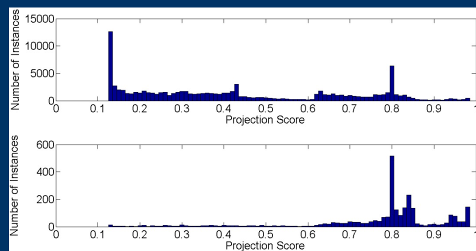






Overall Results

- Histogram of all targets vs. attacked targets via F-VLMM



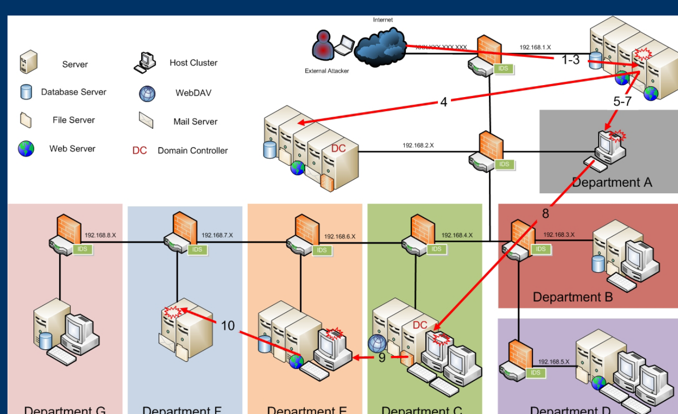
- Overall performance: TBM-CO vs. F-VLMM

	Net A	Net B		
		Eff=0.8	Eff=0.5	Eff=0.3
TBM-CO	[75%,80%]	[59%,79%]	[55%,77%]	[54%,74%]
F-VLMM	[81%,88%]	[85%,88%]	[84%,88%]	[81%,84%]

- TBM-CO compensates F-VLMM when patterns is insufficient.

High Efficiency Attack Example (F-VLMM)

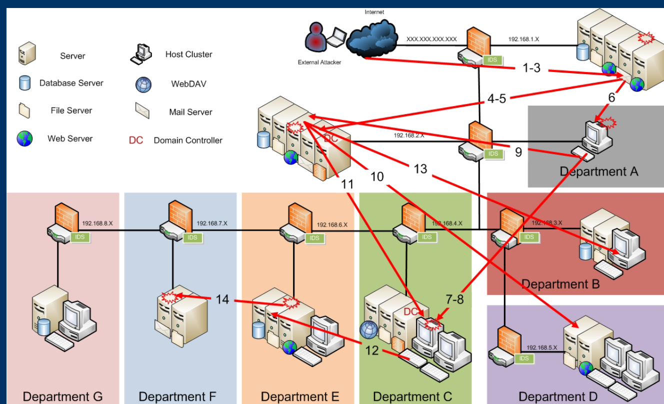
- Direct attack penetrating through the network
  - Critical and can be predicted relatively well



Step	Percentile Rank
1	N/A
2	[96.78%,100%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[96.88%,100%]
6	[93.75%,96.88%]
7	[93.75%,96.88%]
8	[96.75%,96.88%]
9	[90.63%,93.75%]
10	[93.75%,96.88%]

Low Efficiency Attack Example (F-VLMM)

- Random movements spreading all over the network
  - Noisy with some less predictable movements
  - TBM-CO helps in some of these cases



Step	Percentile Rank
1	N/A
2	[90.63%,93.75%]
3	[93.75%,96.88%]
4	[96.88%,100%]
5	[50.00%,53.13%]
6	[84.38%,87.50%]
7	[81.25%,84.38%]
8	[96.88%,100%]
9	[37.5%,40.63%]
10	[40.63%,43.75%]
11	[84.38%,87.5%]
12	[84.38%,87.5%]
13	[18.75%,21.88%]
14	[3.13%,6.25%]

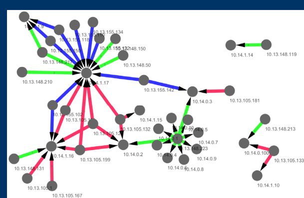
Overall Premise and Extension

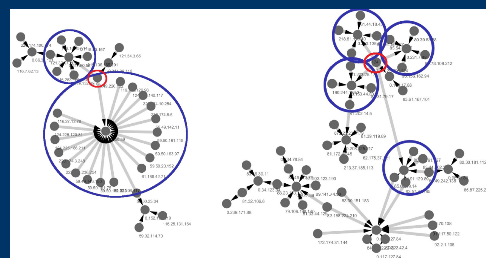
- Predictive Cyber Situation Awareness & Actionable Information
  - Joint spatial and temporal attribution of coordinated cyber attacks
  - Enterprise networks, sensor networks, intelligent networks
- Temporal Attribution: treat each sequence of observables as a Virtual Track in cyber space
  - Capability: estimates the service vulnerabilities and exploits each track is more capable doing, based on its own and similar tracks
  - Opportunity: estimates the exposed hosts and services for each track currently given the virtual terrain of the network
  - Behavior Trend: captures the attack patterns in semi-real-time manner using Variable Length Markov Model
  - Ensemble approach: combines the estimates from different approaches

Spatial Attribution of Cyber Attacks

Explore common target space across cyber attack tracks

- Attack Social Graph (ASG): defined to reveal coordinated / colluding attacks
- Utilizes social network analysis, e.g., centrality and clustering to find
  - Similarly behaving tracks
  - Complementary roles in coordinated attacks

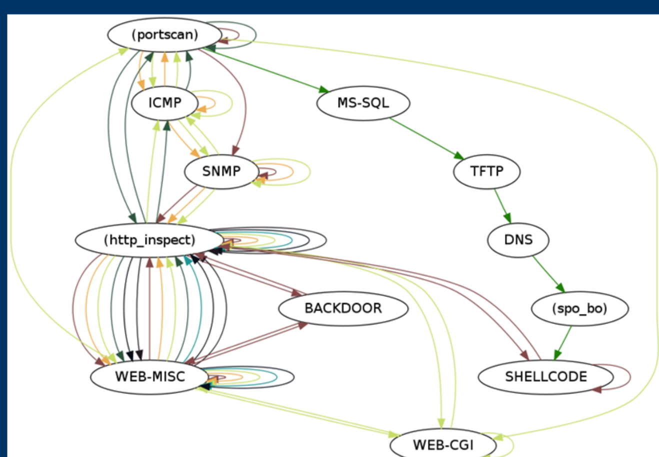




Next Step:
Integrated temporal and spatial analysis

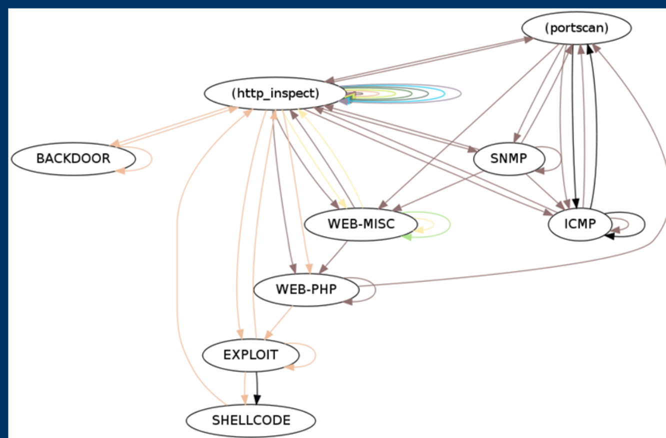
Spatial Attribution of Cyber Attacks

- Attack Social Graph (ASG) in another view, showing attacking strategies of coordinated team – data from iCTF 2008.



Spatial Attribution of Cyber Attacks

- Attack Social Graph (ASG) in another view, showing attacking strategies of coordinated team – data from iCTF 2008.



Questions & Comments?

- Questions and comments?
- Expertise areas needed for Cyber Fusion
  - Security, networks, machine learning, data mining, social computing, graph theory, information theory, belief combination, …
- Contributing students / alumni
  - Jared Holsopple, Daniel Fava, Brian Argauer, Stephen Byers, Chris Murphy, Jordan Bean, Daniel Liu, Jon Szymaniak, Corey Beres
  - Haitao Du, Biru Cui, Steven Strapp, Neil Wong
- NetIP Lab @ RIT
  - S. Jay Yang
  - http://www.ce.rit.edu/people/yang
  - http://netip.ce.rit.edu

Characterizing Large-Scale Attack Behavior for Predictive Cyber Situation Awareness

Shanchieh Jay Yang*[a]
[a]Department of Computer Engineering, Rochester Institute of Technology
83 Lomb Memorial Drive, Rochester, NY, USA 14623-5603

ABSTRACT

Previous works in the area of computer network security have emphasized the creation of Intrusion Detection Systems (IDSs) to flag malicious network traffic and computer usage. Raw IDS data may be correlated and form attack tracks, each of which consists of an ordered collection of alerts belonging to a single hypothesized attack. Assessing an attack track in its early stage may reveal the attacker's capability and behavior trends, leading to projections of future intrusion activities. Behavior trends can be captured via Variable Length Markov Models (VLMM) without predetermined attack plans. Attacker's capability can be inferred by correlating services that have been exploited by same attack tracks. Extending from these techniques, which process observables of individual attack tracks, this work will also discuss challenges in characterizing colluding attacks where deriving of spatial and temporal relationship across tracks is needed.

Keywords: information fusion, attack attribution, threat projection

1. INTRODUCTION

Prevalent computing devices with networking capabilities have become critical cyber infrastructure for government, industry, academia and every day life. As its value rises, the motivation driving cyber attacks on this infrastructure has shifted from the pursuit of notoriety to the pursuit of profit [1,2]. At the same time, the attacking strategy is also becoming more and more sophisticated, with multistage, coordinated attacks mixed with other large-scale malicious as well as mis-configured traffic. Botnets are being advertised in underground market for $15 per 10,000 bots [1]. The ease of entry and continuous emergence of new vulnerabilities have made the protection of cyber infrastructure challenging. Simply detecting the existence of malicious incidents is not sufficient because they happen everywhere and all the time. To this end, the research community calls for the need for higher-level fusion that estimates adversary strategies for enhanced, predictive situation awareness.

Analyzing adversary attack strategy is not new and falls under the notion of threat and impact assessment. Many data fusion reference models [3-7] have included this idea. Endsley [4] defines situation awareness as a "state of knowledge that results from a process." Salerno [6] suggests that this process is "situation assessment." In other words, situation awareness is a cognitive process, from perception to comprehension and anticipation, that can be aided by a process called situation assessment. Using these definitions (which should be noted, are slightly different with the commonly referenced JDL model [3]), situation assessment encompasses both threat and impact assessment, with the goal to enhance the situation awareness of a decision-maker or analyst, enabling effective and educated decisions.

While the notion of analyzing adversary behavior is not new, limited success has been shown to characterize cyber attack behavior. For over a decade, much work has devoted to alert correlation, *e.g.,* [8-16], aiming at aggregating intrusion detection system (IDS) alerts to create hypothesized cyber attack tracks for higher-level analyses. Most alert correlation work depends largely upon attack plans or pre- and post-conditions that are developed based on a priori knowledge. These a priori models represent specific attack progression but could not catch up with the diverse and evolving nature of sophisticated cyber attacks. This work advocates for scalable and adaptive computational techniques that can capture temporal and spatial behaviors of multistage and potentially coordinated cyber attacks.

2. CYBER ATTACK CHARACTERIZATION

Large-scale cyber attacks can take the traditional form of a botnet, from which a large number of hosts perform similar actions for, *e.g.,* DDoS or distributed stealthy scans [2]; they can also consist of a team of colluding sources dividing up tasks, interleaving the actions over time and dispersing over the IP and port spaces to conceal their overall strategy. It is not uncommon for an enterprise or global network to face multiple coordinated attack teams simultaneously along with other large-scale malicious activities.

To estimate cyber attack strategies, one can draw analogy from the threat assessment framework in traditional warfare [17]. Particularly, we argue that one needs to assess the following perspectives of cyber adversaries.

- **Capability**: The intrusion methods the attacker has use is indicative to the types of vulnerabilities he is capable of exploiting. Examples of computational techniques that analyze attack capabilities include graph-based estimation [18], statistical estimation [19] and the use of Recommendation Systems [20].
- **Opportunity**: Given the already compromised entities or privileges by a given attack, originally hidden entities or vulnerabilities may be exposed and give opportunities to the attacker. The computational techniques that analyze attacker opportunities are primarily graph-based estimation [18,21,19].
- **Intent**: The intent of a cyber attacker can be quite diverse, perhaps making it impossible to estimate. Instead of assessing the true intent, cyber intrusion projecting could examine the criticality of network entities and operations to determine the worst-case intent of the attacker. To the author's knowledge, there isn't any scalable, adaptive computational technique that successfully estimates intent of cyber attacks.
- **Behavior trend**: The patterns exhibited in the observed cyber attack actions can be indicative to future targets or actions. The pattern may exist in attack methods, types of services or OS attacked, subnets visited, protocols exploited, etc. Examples of computational techniques that analyze attack patterns include Variable Length Markov Models (VLMM) [22,19] and Recommendation Systems [20].

The above characterization aspects need to be combined to create an ensemble technique so that the prediction is robust against various stealthy and concealing attack strategies. Du *et al*. [19] has utilized Transferable Belief Model (TBM) to combine estimates from Capability and Opportunity algorithms, and Fuzzy combination to fuse estimates produced by VLMM with respect to different alert attributes. The experimental results shown that the behavior trend based analysis is effective for cyber attacks that are directly progressing toward the final target. Adversaries use direct attacks if they want to minimize the number of observables or time to reach the ultimate target in a multistage attack. If the attack explores a wide range of systems in the network, either intentionally or unintentionally, Capability and Opportunity based techniques will be more robust to the less relevant yet still malicious actions.

3. OUTLOOK: INTEGRATION WITH SPATIAL ATTRIBUTION

The techniques described in Section 2 are primarily based upon analyzing the sequence of attack actions as exhibited by IDS alerts. As a result, the temporal behavior is being extracted, but much spatial relationship between attacking sources has not been explored. Drawing analogies from social network analysis, we hypothesize that large-scale cyber attacks can and should be analyzed to determine the colluding behavior [23]. By using properly defined Attack Social Graphs, one may develop efficient algorithms to extract the dependencies between attack sources in their role of a coordinated attack. The various centrality measure and community prediction techniques could be the bases to decode sophisticated attacking strategies in a timely manner and provide enhanced predictive situation awareness.

REFERENCES

[1] M. Fossl *et al*., "Internet Security Threat Report," Symantec Corp., Technical Report, vol. 16, 2010.

[2] C. Zhou, C. Leckie, and S. Karunasekera, "A survey of coordinated attacks and collaborative intrusion detection," Computers & Security, vol. 29, no. 1, pp. 124–140, Feb. 2010.

[3] M. Liggins, D. Hall and J. Llinas, [Handbook of Multisensor Fusion: Theory and Practice], Second Edition, CRC Press, 2008.

[4] M. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," Human Factors: The Journal of the Human Factors and Ergonomics Society, 37, 32-64, 1995.
[5] D. Lambert, "Grand challenges of information fusion," in Proceedings of the Sixth International Conference on Information Fusion, vol.1, pp.213-220, 2003.
[6] J. Salerno, "Measuring situation assessment performance through the activities of interest score," in Proceedings of the 11th International Conference on Information Fusion, July 2008.
[7] A. Stotz, "Advancements in Situation Assessment Definition and Computation with a Suite of Graph Based Technologies," Ph.D. Dissertation, University at Buffalo, Oct 2008.
[8] A. Valdes and K. Skinner, "Probabilistic alert correlation," in Recent Advances in Intrusion Detection (RAID). Springer, 2001, pp. 54–68.
[9] O. Dain and R. K. Cunningham, "Fusing a heterogeneous alert stream into scenarios," in Proceedings of ACM Worksop on Data Mining and Security, December 2001.
[10] H. Debar and A. Wespi, "Aggregation and correlation of intrusion-detection alerts," in Recent Advances in Intrusion Detection (RAID). Springer, 2001, pp. 85–103.
[11] F. Cuppens and A. Miege, "Alert correlation in a cooperative intrusion detection framework," in Proceedings of IEEE Symposium on Security and Privacy, pp. 202–215, 2002.
[12] S. Cheung, U. Lindqvist, and M. W. Fong, "Modeling multistep cyber attacks for scenario recognition," in Proceedings of DARPA Information Survivability Conference and Exposition, vol. 1, April 2003, pp. 284–292.
[13] F. Valeur, G. Vigna, C. Kruegel, and R. Kemmerer, "A comprehensive approach to intrusion detection alert correlation," IEEE Transactions on dependable and secure computing, 1(3), 146–169, 2004.
[14] P. Ning, D. Xu, C. G. Healey, and R. S. Amant, "Building attack scenarios through integration of complementary alert correlation methods," in Proceedings of the 11th Annual Network and Distributed System Security Symposium (NDSS04), 2004, pp. 97–111.
[15] A. Arnes, F. Valeur, and R. Kemmerer, "Using hidden markov models to evaluate the risk of intrusions," in Proceedings of the International Symposium of the Recent Advances in Intrusion Detection (RAID), Hamburg, Germany, 2006.
[16] A. Stotz and M. Sudit, "INformation fusion engine for real-time decision-making (INFERD): A perceptual system for cyber attack tracking," in Proceedings of 10th International Conference on Information Fusion, July 2007.
[17] A. Steinberg. "Open interaction network model for recognizing and predicting threat events," in Proceedings of Information, Decision and Control (IDC) '07, pages 285–290, Feb. 2007.
[18] J. Holsopple, S. J. Yang, and M. Sudit, "TANDI: Threat assessment for networked data and information," in Proceedings of SPIE Defense and Security Symposium: Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications Conference, vol. 6242, April 2006.
[19] H. Du, D. Liu, J. Holsopple, and S. J. Yang, "Toward Ensemble Characterization and Projection of Multistage Cyber Attacks," in Proceedings of IEEE ICCCN'10, Zurich, Switzerland, August 2-5, 2010.
[20] F. Soldo, A. Le, and A. Markopoulou, "Blacklisting Recommendation System: Using Spatio-Temporal Patterns to Predict Future Attacks," IEEE Journal on Selected Areas in Communications, 29(7), 1423–1437, Aug. 2011.
[21] J. Holsopple and S. J. Yang, "FuSIA: Future situation and impact awareness," in Proceedings of ISIF/IEEE International Conference on Information Fusion, July 2008.
[22] D. Fava, S. Byers, and S. Yang, "Projecting cyberattacks through variable-length markov models," IEEE Transactions on Information Forensics and Security, 3(3), 359–369, Sept. 2008.
[23] H. Du and S. Yang, "Discovering collaborative cyber attack patterns using social network analysis," in Proceeding of International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. Springer, 2011, pp. 129–136.

Presentation to SPIE
Baltimore — April 2012

SA Technologies

Building Robust Situation Models for Higher Level Fusion & Social-Cultural Modeling

Mica R. Endsley
SA Technologies

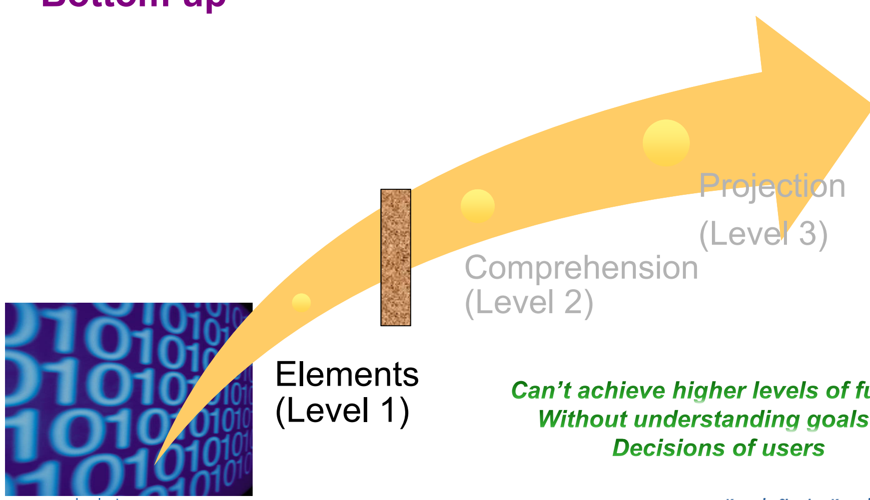
www.satechnologies.com

Know the Situation. Know the Solution.

Approaches to Fusion

SA Technologies

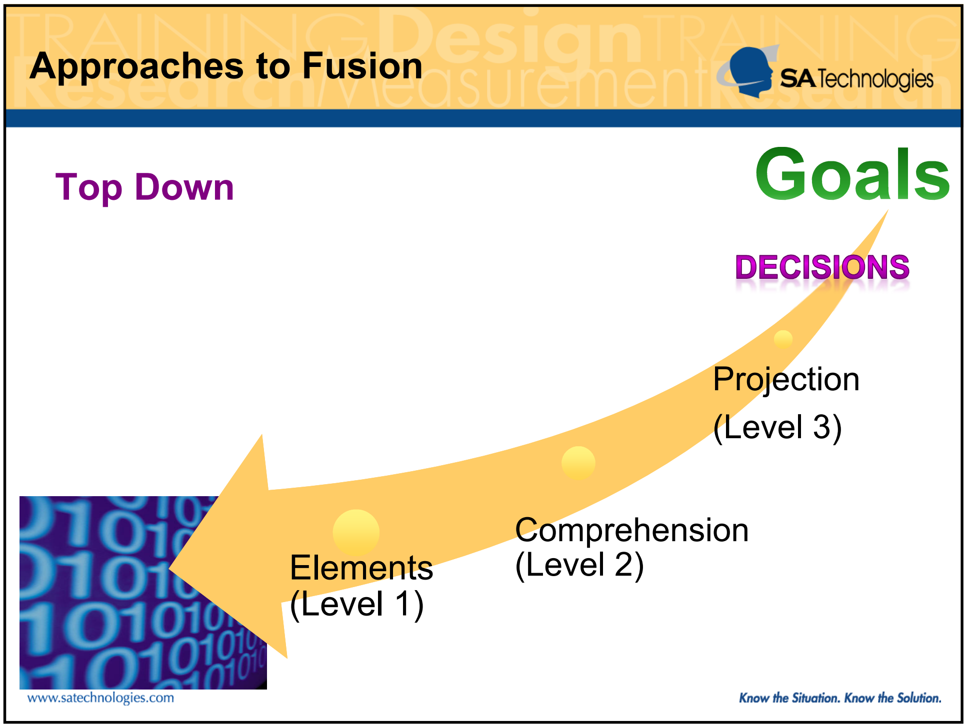
Bottom up

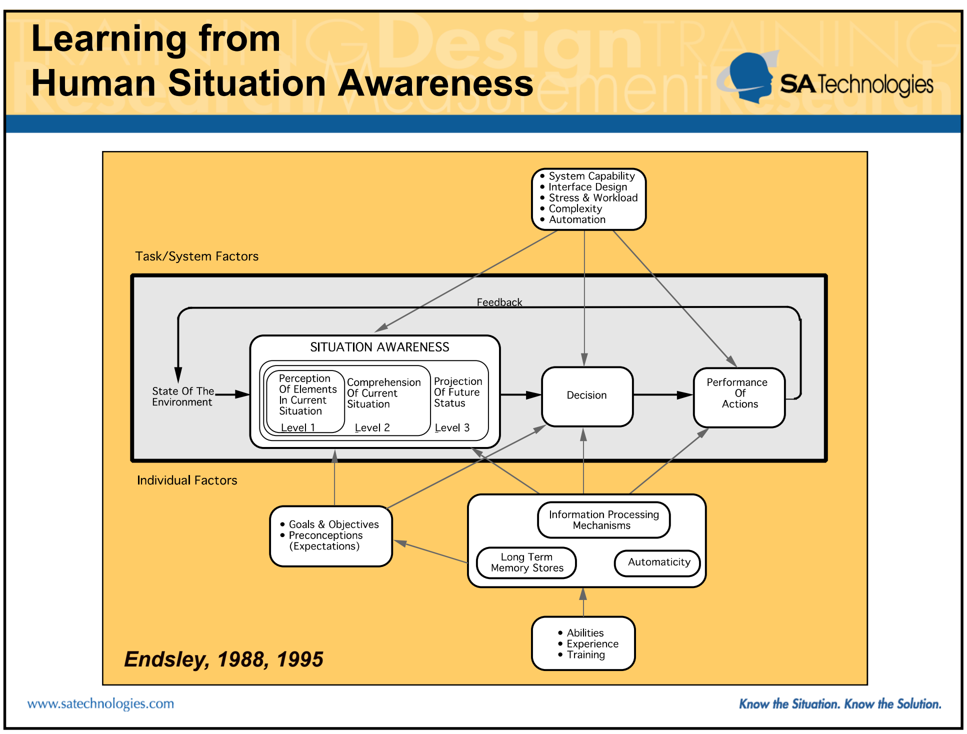


Can't achieve higher levels of fusion Without understanding goals & Decisions of users

www.satechnologies.com

Know the Situation. Know the Solution.


Approaches to Fusion
SA Technologies
Top Down
Goals
DECISIONS
Projection
(Level 3)
Comprehension
(Level 2)
Elements
(Level 1)
www.satechnologies.com
Know the Situation. Know the Solution.


Learning from
Human Situation Awareness
SA Technologies
• System Capability
• Interface Design
• Stress & Workload
• Complexity
• Automation
Task/System Factors
Feedback
SITUATION AWARENESS
State Of The Environment
Perception Of Elements In Current Situation
Level 1
Comprehension Of Current Situation
Level 2
Projection Of Future Status
Level 3
Decision
Performance Of Actions
Individual Factors
• Goals & Objectives
• Preconceptions (Expectations)
Information Processing Mechanisms
Long Term Memory Stores
Automaticity
• Abilities
• Experience
• Training
Endsley, 1988, 1995
www.satechnologies.com
Know the Situation. Know the Solution.

Computer Models of SA Will Need:

SA Technologies

- An internal "Mental Model" of the system/environment
  - Defines "Relevant"
  - Provides Dynamic Information Prioritization
  - Provides Mechanism for Dynamic Integration of Data Creating Meaning
    - Comprehension required
    - Projections required
  - Active Learning and Model Refinement
  - Links to Schema and Scripts for Prototypical Situations
- Goals
  - Pre-requisite to Relevance and Meaning
  - There are frequently Multiple Goals which Vary in Priority
  - Need a Mechanism for Goal Prioritization
    - Critical Cues to Trigger Goal Priorities

www.satechnologies.com

Know the Situation. Know the Solution.

Social-Cultural Information

SA Technologies

- Social, cultural and organizational information is often needed for SA in many domains

www.satechnologies.com

Know the Situation. Know the Solution.

Need to make sure such information is relevant and actionable

SA Technologies

- Multi-cultural teams
- Working with indigenous civilian populations
- Understanding enemy operations & predicting behaviors
- Non-kinetic missions
  - SASO
  - Rescue

www.satechnologies.com

Know the Situation. Know the Solution.

Examples of Relevant Social-Cultural Information needed for SA

SA Technologies

- Projected civilian behavior
- Disposition
- Location
- Number
- Refugee flow
- Known terrorists
- Media
- NGO/IGO
- Living conditions
- Clans present
- Ethnicities
- Culture
- Languages spoken
- Level of organization
- Mood of crowd
- Religious/political beliefs
- Agitators present
- Threatening actions
- Weapons
- Morales/commitment
- Training/skills
- Intent
- Politics
- Potential terrorists

www.satechnologies.com

Know the Situation. Know the Solution.

Goal Directed Task Analysis

SA Technologies

- Goals
  - Subgoals
    - Decisions
      - Projection Requirements
      - Comprehension Requirements
      - Perception Requirements

Establishes Foundation for Higher Level Fusion

www.satechnologies.com

Know the Situation. Know the Solution.

GDTA Identifies key factors needed for situation model

SA Technologies

1.0 Assess mission / operational environment

Goal
Captures user's high-level goals

1.4 Assess impact of human, social, culture, and behavior considerations on mission success

Sub-Goal
Goals are decomposed into sub-goals for higher fidelity

D2: How to minimize disruption to civilian activity?

Decision
For optimal decision-making support, user's decisions are explicitly documented

Information Requirements
SA requirements are listed hierarchically

- Projected effect of significant events on mission
  - Impact of religious and civil days, holidays, cultural events, celebrations
  - Impact of protests
    - location, crowd size, timing, traffic control),
  - Impact of harvest / planting cycles
    - Crop type
  - (…)
  - Impact to socio-cultural terrain
    - Social norms, religious customs, traditions, behaviors, taboos, psychological state
    - Consideration of past history
    - previous missions, mission impact (damage, significant incidents), casualties (direct or indirect)
    - (…)

GDTA identifies:

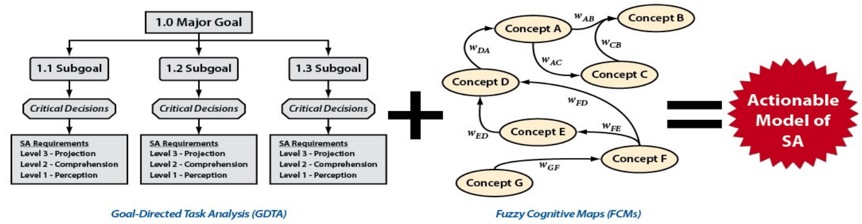
- Goal structure
- Mapping of goals to decisions
- What data is pertinent for each key decision
  - Hard data
  - Social/cultural factors
- How data is combined to reach each decision
- Higher levels of fusion
  - Comprehension
  - Projections

www.satechnologies.com

Know the Situation. Know the Solution.

Fuzzy Cognitive Modeling of SA (SA-FCM)

SA Technologies



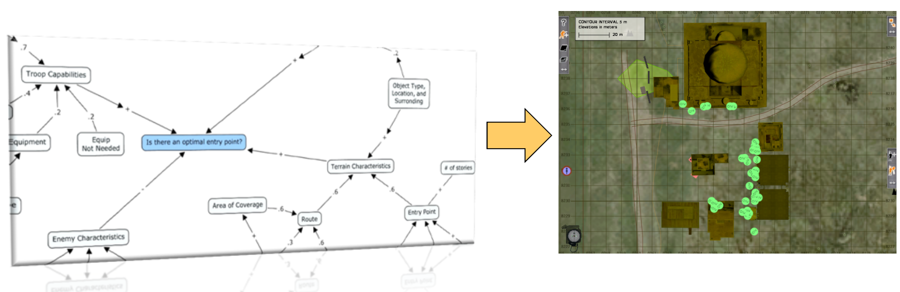
It generally will be necessary to combine socio-cultural type information with other "hard data" to derive meaningful higher level SA

www.satechnologies.com

Know the Situation. Know the Solution.

SA-FCM for Infantry operations

SA Technologies



- **Created SA-FCM for understanding Platoon Leader SA**
- **Passed Turing test in simulated mission in VBS2**

www.satechnologies.com

Know the Situation. Know the Solution.

Summary

SA Technologies

- Computer models of SA are a major goal for intelligent systems
- Foundation is laid for attributes and mechanisms needed for robust computer situation models
- The challenges of modeling socio-cultural information into meaningful and actionable models are significant
- Efforts will need to stay focused on the actual goals of the users to stay relevant
- Existing research and tools provide a starting point for these efforts

www.satechnologies.com

Know the Situation. Know the Solution.

Building Robust Situation Models for Higher Level Information Fusion & Social-Cultural Modeling

Mica R. Endsley*
SA Technologies, 3750 Palladian Village Drive, Building 600, Marietta, GA 30066

ABSTRACT

The development of effective computer models for understanding and interpreting data to form effective decisions is dependent on their ability to accurately characterize key features of the environment, including social and cultural aspects, in order to make accurate and robust situation assessments. Building upon a detailed cognitive model of how human decision makers build situation awareness and the mechanisms they rely on, requirements for computer models of SA are presented. In addition, existing cognitive engineering tools can be leveraged for aiding in the development of situation models.

Keywords: situation awareness, information fusion, situation models, social, cultural

I. INTRODUCTION

A great deal of interest has developed in creating computer models that can form an analog of human situation awareness (SA). These tools would be useful for automating some aspects of situation management, or for aiding human operators in the challenge of maintaining situation awareness across large and complex operations. In order to create effective fusion approaches for creating SA out of basic low level data, it is necessary to understand the mechanisms involved in human SA, and to apply analogs of these mechanisms within the architecture to achieve similar functionality. Such an approach has much to offer as a means of overcoming the limits of many traditional algorithmic approaches that fall far short of the goals of the information fusion community in its quest for level 2 and 3 fusion. Key aspects of human SA need to be incorporated into computer models of SA. In addition, approaches from the field of cognitive engineering can be leveraged for creating the needed computer models.

2. REQUIREMENTS FOR COMPUTER MODELING OF SITUATION AWARENESS

Based on how humans develop good situation awareness in complex and dynamic environments[1], the following characteristics are needed for a computer model of situations that attempts to achieve this goal [2]. First it needs to include a model of the system and environment that the situational model is about, including relevant elements of each, and how they relate to each other to form meaningful states. This model defines what is "relevant" regarding system and environmental information, provides for dynamic information prioritization, and provides a mechanism for integration of low level data to create meaning (e.g. an understanding of the significance or importance of low level data and projections of possible and likely future situation states). A process of active learning is needed to maintain such a model and to refine it as new things about the system are learned. Humans do this through a process of Q-morphisms [1]. In addition, where recognized classes of situations exist (e.g. case based reasoning), these need to be linked to the model (likely requiring a hybrid model), for rapid processing of well defined situations. The more extensive system model can be used in circumstances where there is not a good fit with known cases.

Secondly, to be successful, these models need to capture an understanding of the goals that are relevant for the purposes of the model. Without goals, sensed data has no independent meaning, making strictly bottom-up fusion nearly impossible. Goals define the relevance of information (separating signal from noise), and allow for meaning to be established regarding that information. Most human roles have multiple goals which people are trying to accomplish and that dictate the types of decisions they need to make, and thus what information they need to attend to and how they process it to make those decisions. Thus the higher levels of situation awareness (comprehension and projection) are largely dictated by the goals to which low-level data are being applied. While much data fusion work has attempted to be primarily bottom-up (trying to generate meaning from low level data alone), research on human decision making shows that top-down processing – with goals dictating the processing of the data – is a heavily used process for achieving meaning from low level cues. Finally, as there can be multiple, and sometimes competing goals, the computer model will need to include a mechanism for goal prioritization, along with knowledge of which data states are pertinent for trigger goal priorities. The challenges of creating a robust computer model of situations are not easy, but many of these capabilities do exist in existing computer science approaches and can be combined into a successful model.

3. THE INTEGRATION OF SOCIAL & CULTURAL INFORMATION INTO FUSION

In both asymmetrical warfare and many homeland defense environments, there is need to understand the potential impact of cultural factors, social concerns, and communication skills have on properly understanding the situation and thus the ability to meet tactical objectives. This kind of understanding is typically gained through human situation awareness as decision makers put together data gathered with their understanding of how the enemy operates, cultural practices and norms of civilians, and even social aspects regarding friendly troops (e.g. the effects of fatigue, morale, etc…) to properly interpret data and understand its significance. As human decision makers improve upon their abilities to perceive the critical cues in the environment and to understand how belief and actions interact with those of others, they are better able to assess the impact of their cross-cultural interactions on mission goals.

Culture typically refers to origins, values, beliefs, and the mores and norms of society and social interactions of a particular group of peoples. It is the patterns of human activity and communal structures that are the foundation of a society and give various actions and behaviors meaning and importance. McFarland (2005) states that soldiers must develop a clear understanding and awareness of the sociocultural landscape and human behavior as they relate to mission goals in order to promote effective decision-making and evade potentially catastrophic misunderstandings. This includes the ability to use finesse, diplomacy, and strategic communication to achieve mission goals.

Further, there is a need for tools and techniques to achieve higher cultural SA and, therefore, better decision-making. Social and cultural terrain and the projected influence of cultural factors (e.g., history, religion, domestic and global perceptions, family and social structure, beliefs and norms) on tactical goals need to be incorporated into fusion models in these environments in order to improve their performance and chances of successes. Incorporating a clear understanding and awareness of culture, social-economic landscape, and human behavior within these models, *as they relate to mission goals*, is needed to promote effective decision-making and evade potentially catastrophic mistakes. As the number and aspects of human culture and social interactions is almost limitless, at least so far as creating effective models is concerned, it is critical that modelers understand just which aspects of the social-cultural environment they need to capture and how that links to other data and pertinent decisions. This question must be guided by an understanding of goals and the decisions that need to be made.

A key requirement for progressing in the formation of good fusion models is the identification of the actual social and cultural factors that impact decision making in specific operational contexts. That is, the fusion models need to incorporate not only hard data from sensors regarding objects, but also the "soft data" associated with critical social, cultural and behavioral factors that influence how people in the environment can be expected to act. These factors form the SA requirements for a given decision maker (the things the person needs to be ale to perceive, comprehend and project). Typically, SA requirements analyses have been conducted using a form of cognitive task analysis known as a goal-directed task analysis (GDTA) [3, 4]. The GDTA involves in-depth knowledge elicitation with domain experts in order to identify the major goals of a particular job class and to define the subgoals for meeting each higher goal. Associated with each sub-goal are the major decisions that need to be addressed during task performance. These decisions are identified along with the specific SA elements (related to perception, comprehension, and projection) needed for making the decisions and carrying-out the sub-goals. The SA requirements focus on what data is needed, as well as on how the data should be integrated or combined to make decisions.

An example of some of the output from a GDTA is shown in Figure 1. Each goal and subgoal for the identified role are delineated, with key decisions and SA requirements determined for those goals. The GDTA provides a useful tool that is available identifying the social and cultural factors that are relevant to the mission, and how they are relevant. This methodology has been used in many contexts including aviation, intelligence operations and military command and control. A key advantage of this approach is that it provides a systematic means for identifying not only the hard data needed for SA but also the soft data factors that are pertinent to the decision being made, including the social, cultural and behavioral indicators. Thus it provides a ready foundation for not only the knowledge engineering challenge in this area, but also for how that information needs to be used and combined with other data towards pertinent decisions.

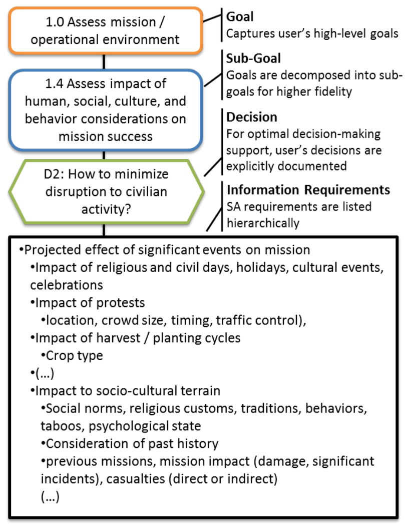


Figure 1. An excerpt from a GDTA identifying relevant factors for decision making

Building on the GDTA, a variety of fusion methods can be applied to create level 2/3 fusion corresponding to the higher levels of SA needed for creating sufficient understanding of the meaning of data obtained from the environment. We have been using Fuzzy Cognitive Models (SA-FCM) to create level 2/3 fusion corresponding to the levels of understanding and projections needed in the operational domain [5]. As an advantage, these models readily combine soft data and hard data dynamically to update the situation model in the fusion engine. The soft data can come from default values in the model (e.g. a member of a particular group as certain general characteristics), but can also be updated in real time when more detailed data is available. In this presentation we will provide examples of GDTAs illustrating how both soft and hard data are combined to form the situation model and examples of the SA-FCM approach for fusing this data.

REFERENCES

[1] Endsley, M.R., "Toward a theory of situation awareness in dynamic systems", Human Factors, 37(1), p. 32-64 (1995).

[2] Endsley, M.R., "Bringing cognitive engineering to the information fusion problem: Creating systems that understand situations", Presentation to Fusion 2011, International Society of Information Fusion: Chicago, IL. (2011).

[3] Endsley, M.R., "A survey of situation awareness requirements in air-to-air combat fighters", International Journal of Aviation Psychology, 3(2), p. 157-168 (1993).

[4] Endsley, M.R., and D.G. Jones, [Designing for situation awareness: An approach to human-centered design], 2nd ed., Taylor & Francis, London, (2012).

[5] Jones, R.E.T., Connors, E. S., Mossey, M.E., Hyatt, J.R., Hansen, N.J., Endsley, M. R. "Modeling situation awareness for army infantry platoon leaders using fuzzy cognitive mapping techniques", Proceedings of Behavior Representation in Modeling and Simulation (BRIMS), Charleston, SC. (2010).

Modeling Human Behavior within the Fusion Process

John Salerno, George Tadda

Integrity ✯ Service ✯ Excellence

John Salerno, PhD
AFRL/RIED
john.salerno@rl.af.mil
315-330-3667

SPIE 2012 **This Briefing is Unclassified** **24 Apr 2012**

Goal of Presentation

Attempt to bring together a number of the concepts discussed in the previous briefs into an overall approach/architecture

Humans is the “system” of interest -> are
Humans as part of the “system” -> involved
Humans within the “system” of interest -> observe

Background

(Why A Reference Model and Fusion)

- Original Purpose of JDL Model was as a taxonomy to describe various activities/programs – Not as an ARCHITECTURE
  - Over time, JDL Model adopted as an international reference model
- Early Years, work was primarily in Level 0/1, Object ID and Tracking, and Level 4, Resource Management
  - Tracking of Physical Objects displayed on maps
- For the Past 7 years, work has increased in Level 2/3 under "Situation Awareness"
  - Why?: Higher fusion levels can address data glut problem by aggregating dots into more complex objects (units, divisions, etc.)
- Over time have seen many attempt to use the JDL Model as an architecture, and on problems beyond the tactical battlefield.

…but

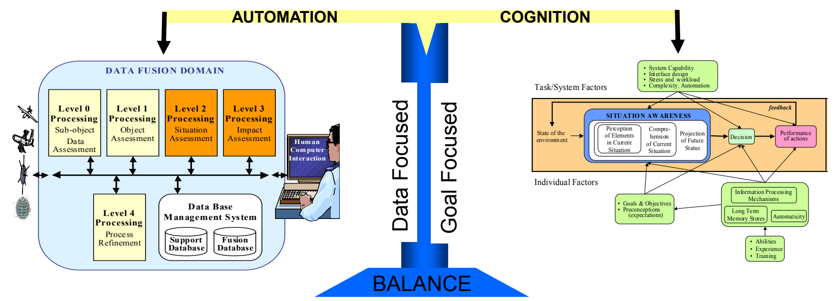
- Where do groups, events, activities fit in?
  - Can we not track a group, an activity (Why only Objects?)
  - Is a group or activity only a complex object?
- What is a Situation? Is there more than one?
- Where does Knowledge Discovery exist?
- What is Situation Assessment?
- Threat Assessment only future – what about today?
  - Can't we also have a Threat within the current situation?
- What about forecasting or projecting the "future" state?

…and where do humans fit in?

A Review of the Popular Models

FUSION - TACTICAL

SITUATION AWARENESS

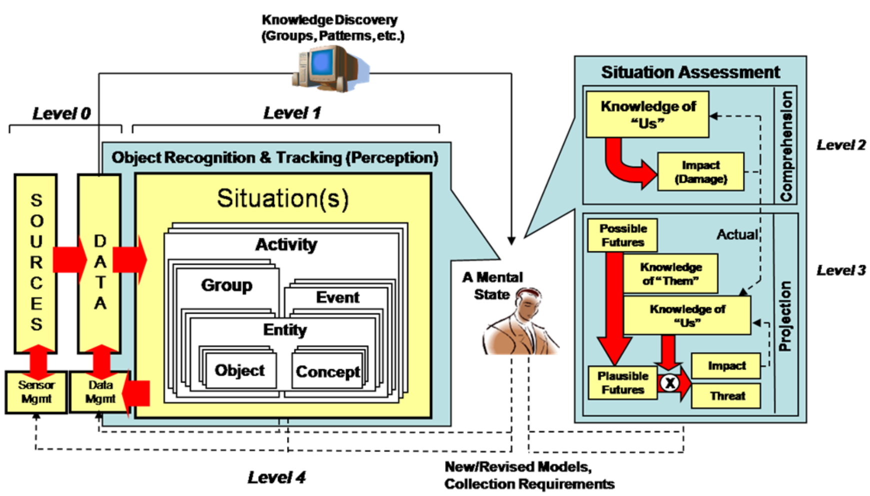


- Most popular is the Joint Director's of Laboratory (JDL) Model (Sensor-based)
- Functional Model
- 5 Levels (Level 0,1, 2, 3, 4)
- Published By Llinas, Hall, White (1992)
- Most work concentrated on Level 0/1/4 (Dots on Map)
- Little definition of Level 2/3 (What do they mean?)
- Bottom-up, Data Driven

- Receiving Much Attention Today from the Cognitive Community
- Mental Model
- 3 Levels: Perception, Comprehension, Projection
- Developed by: M. Endsley (1995)
- Extended by McGuinness and Foy for Resolution
- Top Down, Goal Driven

THERE IS A NEED FOR BOTH

Revised Situational Awareness Reference Model

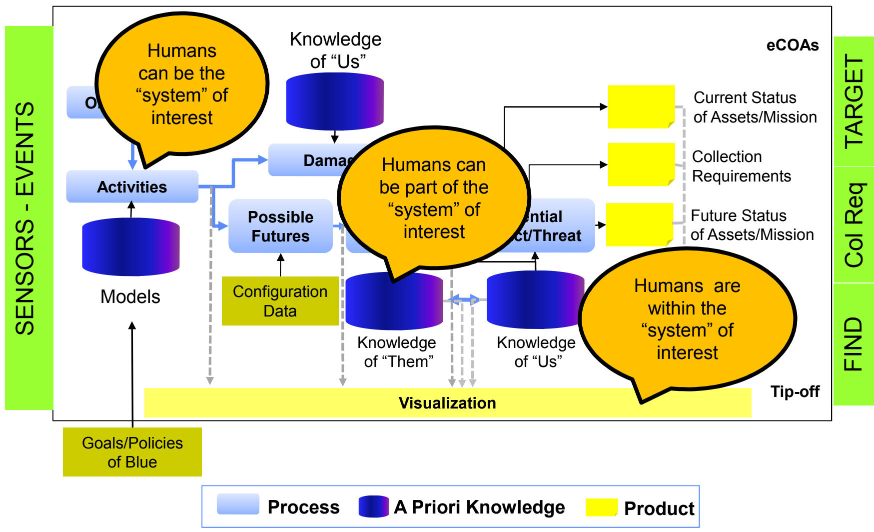


Ranking Activities Based on Their Impact and Threat, ISIF 2009

Capturing the Process

(Our Reference Model - Process Flow Version)

From Observables to Threats (At Time, t)



Humans can be the "system" of interest

(Activities of Interest)

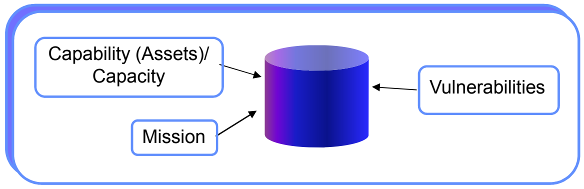
An **activity** is "something done as an action or a movement". Activities are composed of entities/groups related by one or more events over time and/or space. Thus, by definition an event, group or activity can be considered a complex entity (or in terms of the JDL, an object) and can be tracked and identified as such.

- Relates events, objects, groups together
- Groups can be managed via Social Networks

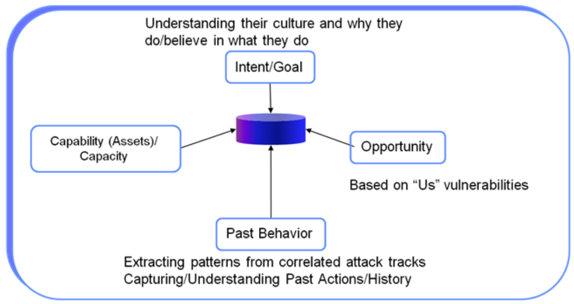
http:\\www.wikipedia.org

Humans can be part of the “system” of interest

Knowledge of “Us”



Knowledge of “Them”



Humans are within the “system” of interest

- What is Available?
  - Visualization of Data, Data, and Data!
  - Limited work looking at the “Big” picture, i.e., a situational map, a view of the status of my assets, what does my current threat picture look at, etc.
- An Attempt…
  - Need to understand the goal or objectives that we are trying to satisfy
  - Base concept on Endsley’s work: Past, Current, Future

Past – Drill Down Evidence

Current Status of Assets/Missions

Projected Threats Collection Requirements

Wrap Up

- Described a process for "quasi-real" time understanding as to what is going on
  - Evidence-based
  - Uses models to support projection and threat
  - Supports Situation Awareness
- Other modeling approaches allow for investigation of possible futures
  - Historical data based
  - What if Analysis
- Used the Proposed Model to bring together the human aspects
- However there are many Open Issues
  - Data Currency/Collection (The need for information NOT Data)
  - Understanding and working with Humans
    - -- Developing/Inferring Intent
    - -- Past Behavior NOT necessarily a good predictor of future

Modeling Human Behavior within the Fusion Process

John J. Salerno, George P. Tadda
Air Force Research Laboratory
Rome NY 13441

ABSTRACT

To date, there has been little research, let alone discussion, within the fusion community concerning how human behavior is modeled and where in the fusion process behaviors should be considered. In this short paper, we introduce some ideas about how to incorporate human behaviors into the fusion process and where in the process their beliefs, intentions, goals, and abilities must be considered.

Keywords: JDL Fusion Model, Situation Awareness, Knowledge of "Them", Knowledge of "Us", Understanding Behaviors

The fusion community, in general, has paid little attention to where and how human behavior is included as part of a system used for data fusion. Blasch and Hall [1, 2] discussed techniques for including a human component but more from the human-machine interface perspective (proposed JDL Level 5) than as an integral part of the fusion process. Waltz introduced in a number of his papers [3], culture and understanding the human (and their environment) as one aspect. But, there has been little or no discussion about where and how human behavior applies or is part of traditional data fusion models such as the Joint Directors' of Laboratories (JDL) Data Fusion Model. In [4], Salerno, et al presented a process that integrated two popular models, the JDL Data Fusion Model [6, 7] and Endsley's Situation Awareness Model [5]. There were two reasons for bringing these two disparate models together. The first was to provide context to the data being fused. This context is based on a model defined *a priori* that attempts to describe a human decision-makers "mental model". In other words, we model the activity of interest that the decision maker desires awareness of within a particular environment. The second reason to integrate these two models was to enable the inclusion of additional human aspects within the fusion process (included by what we termed as the Knowledge of "Us" and the Knowledge of "Them" as detailed in [4, 8]). We also provide a set of definitions and a combined reference model based on almost a decade of research into fusion and situation awareness. Finally, [4, 8], describes a modification of the ideas behind JDL Levels 1 and 2, namely that Situation Identification could be considered a JDL Level 1 process and Situation Assessment (JDL Level 2) as addressing the assessment of the current impact. Then, using Endsley's concept of projection, Threat Assessment (JDL Level 3) is the assessment of the future or projected situation.

So how is human behavior included within the fusion process? Basically, human behavior will influence and affect the entire environment being monitored. Decision makers act and react based on their perceived understanding (SA) about what is impacting them or of particular interest to them. Even physical entities within the environment are in the most part human-controlled and modeling and understanding the human behaviors behind that control must be considered. Particularly, human capabilities, capacities, beliefs and intentions within the domain or environment are necessary aspects for consideration in a fusion process. The combination of the models described above is our Situation Awareness (SA) Reference Model and it primarily captures concepts and definitions. Figure 1 expands upon the reference model and looks at it as a process at an instance in time. Observables are the input to the process and they provide a view of what is occurring in the world (primitive elements of the environment or perception). It is assumed that any attributes associated with the observables have been normalized, cleansed, and transformed into a form that can be used by the subsequent processes. The observables are cues into the activities that a decision maker needs or is interested in (and thus we refer to these as Activities of Interest, AOI – the context) as a way to gain or maintain awareness. The AOI are based on missions, goals, policies, or in general the "things" of interest to one or more decision makers. We define the set of AOIs at an instance in time as the current situation. As observables enter the process, they are categorized and (1) associated with a new stage or step within an existing, ongoing activity; (2) associated with no existing activity and hence become the start of a new activity; or (3) can be a trigger leading to the combination, merging, or removal of existing activities. The aggregation process is similar to tracking individual objects (as defined by JDL Level 1) and why we consider this part of the process, even though dealing with events, still JDL Level 1. Objects are no longer just a physical entity like a tank or tank track but can also be a conceptual (or semantic) entity – a collection of events and observables. The classical tracking problem of association still comes into play using this technique when associating an observable to a step of an activity. At any given time, say *t*, we have a set of ongoing activities (defined earlier as the current situation). As these activities are identified and managed, we are now interested in analyzing the meaning of these activities. This is considered to be Situation Assessment (JDL Level 2). The overall objective of Situation Assessment is to determine if any of the ongoing activities have an impact to 'us' or if they can have a future impact to 'us' (future impact is considered to be Threat Assessment or JDL Level 3). The former looks at the current activities and assessing the impact that the activities have had (follows the upper path of Figure 1).

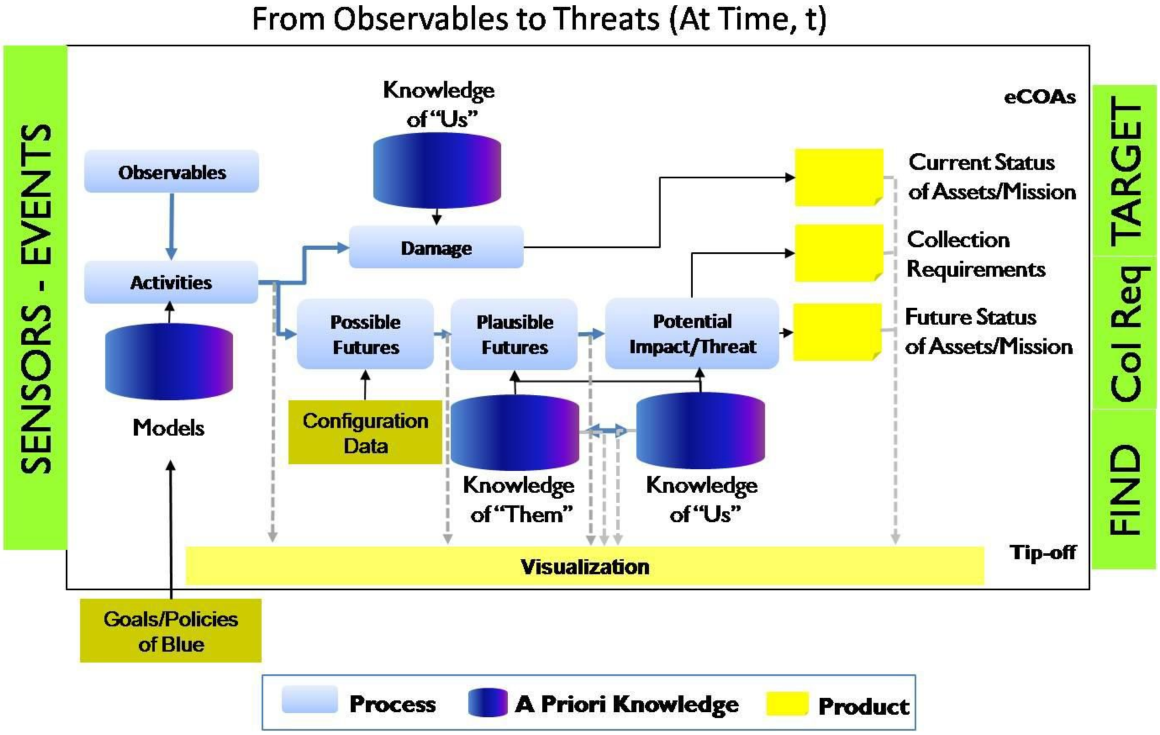


Figure 1 - Situation Awareness Process Model

Since these activities have already happened, Situation Assessment could also be considered to be "Damage" Assessment, i.e., has any of the identified activities caused a current impact and specifically has it caused harm that requires development of a recovery plan to resolve the effect(s) that the activity has caused. In order to accomplish this type of assessment, one not only needs the current, known activities, but also what each activity means to "us" (i.e., Does the given activity impact us in some way?). The data needed to perform this analysis is part of what we refer to as "Knowledge of Us". Thus, this part of the process identifies to the decision maker whether there is a current impact to any capability or asset and whether there is an impact on performing any ongoing or planned mission.

Additionally, a decision maker may be interested in a view of what the adversary (or competitor) is doing or may possibly do. This has generally been described as "getting inside the adversary's OODA (Observe, Orientate, Decide and Act) loop". The sooner we understand what the adversary can/might do, the sooner more options become available to the decision maker. This is addressed by the lower path of Figure 1. The first step of the lower path is to take each identified activity of interest in the current situation and project it forward based on *a priori* knowledge included in the model. For these projections, time isn't considered because we are projecting based on the next step (or stage) in the activity. In some cases it could take milliseconds to go from one stage to another and in other cases it could be days, or longer or it could be that multiple activities being identified at once have differing time scales. The number of stages that we look forward is defined under "Configuration Data". Based solely on the models themselves, we have projected each current activity one or more steps forward; however, these projected or possible futures do not take into account whether they are plausible. In order to determine plausibility, we need to consider additional knowledge. We need both the "Knowledge of Them" and "Knowledge of Us". Specifically, we need to know if the adversary has the capability, capacity, the intent/goal, and have they exhibited similar past behavior consistent with the projections. We also need to know whether they have the opportunity to accomplish the intent(s)/goal(s). Opportunity, in many cases, is based on the vulnerabilities of 'us' (provided as part of the "Knowledge of Us"). Thus, starting with the list of possible futures, we use the “Knowledge of Them” and “Knowledge of Us” to constrain the possible into the plausible for each activity of interest. But what do these plausible futures mean to a decision maker? To answer this question, we again use the “Knowledge of Us” to identify potential impacts and threats to meeting our mission objective(s) but now we assess impacts of the projected plausible. From this portion of the process we get not only future potential impacts/threats but we can also use this knowledge to determine future collection requirements. Using each of the plausible futures, we can identify the key differentiating events that will assist us in distinguishing between which plausible future may actually be unfolding. The key differentiating events can then drive the collection requirements needed to increase the certainty in identifying whether a plausible future is occurring. All this information is needed to constrain the projected space of ‘what is possible’ into the space of ‘what is plausible’. Plausible futures then are a subset of possible futures that have been reduced by eliminating those futures can’t happen or are extremely unlikely to happen give the ‘Knowledge of Them and ‘Knowledge of Us’.

A remaining question not yet addressed are specifics about what are ‘Knowledge of Them’ and ‘Knowledge of Us’ and how they incorporate human behavior within the fusion process described so far. Additional detail is found in [8] about both types of knowledge so this paper will only include the behavioral portions. ‘Knowledge of Us’ contains the least amount of human behavior because it’s primarily focused on describing our environment or the environment in which the activities of interest reside. Typically using a ‘terrain’ (used here to mean anything from an actual physical terrain to a virtual terrain of computers and networks) to capture the environment, ‘Knowledge of Us’ establishes most of the context for the situation and can capture such information as; what am I vulnerable to, what assets are available to me, how am I using those assets to accomplish a specific task or mission, how does the use of particular assets change over time, etc. Thus, ‘Knowledge of Us’ is less about behavior then it is about establishing the environment in which actors are trying to cause effects. Counter to this is ‘Knowledge of Them’ which is largely a description of ‘their’ behavior. They or them is used here rather than specifically saying they are an adversary because we believe these techniques could apply to business processes as well as military operations. ‘Knowledge of Them’ incorporates behaviors and capabilities to assist with the constraining of possible into plausible futures. What might be possible given the ‘Knowledge of Us’ may not be plausible given that an adversary or competitor doesn’t have the capability or capacity to take certain actions. Or, if a competitor’s intent/goal could be identified, some paths through the environment aren’t plausible given that intent/goal. Maybe you’re not vulnerable to a demonstrated capacity of an adversary and thus they wouldn’t have the opportunity to pursue that stage of an activity, eliminating it as a possibility.

This short paper has described a paradigm for implementing a fusion process and how that process captures human behavior. Our greatest challenge in researching and developing technologies that builds a ‘Knowledge of Them’ within a defined ‘Knowledge of Us’ is to assist decision makers in understanding the impacts of the current situation, potential future threats and the impacts of those plausible threats.

REFERENCES

[1] E. Blasch and Plano, S. “JDL Level 5 Fusion Model: User Refinement Issues and Applications in Group Tracking,” SPIE Vol. 4729, *Aerosense,* 2002, pp. 270-279, 2002.

[2] M. J. Hall, S. A. Hall, and T. Tate, “Removing the HCI Bottleneck: How the Human Computer Interface (HCI) Affects the Performance of Data Fusion Systems,” *Proceedings of the 2000 MSS National Symposium on Sensor and Data Fusion*, San Diego, CA, June, 2000, pp. 89-104.

[3] Waltz, Ed, Intelligence Analysis and Processing for Non-Physical Target Systems, IDGA Intelligence Analysis and Processing Conference 2007

[4] J. Salerno, Measuring Situation Assessment Performance through the Activities of Interest Score, Proceedings of the 11th International Conference on Information Fusion, Cologne GE, June 30 – July 3, 2008.

[5] M. Endsley, March 1995. *Toward a Theory of Situation Awareness in Dynamic Systems*. In Human Factors Journal, Volume 37(1), pages 32-64, March 1995.

[6] U.S. Department of Defense, Data Fusion Subpanel for the Joint directors of Laboratories, Technical Panel for C3, “Data Fusion Lexicon,” 1991.

[7] A. Steinberg, C. Bowman, and F. White. Revisions to the JDL Data Fusion Model, presented at the Joint NATO/IRIS Conference, Quebec. October 1998.

[8] J Salerno, S. Yang, I. Kadar, M. Sudit, G. Tadda, J. Holsopple, Issues and Challenges in Higher Level Fusion: Threat/Impact Assessment and Intent Modeling (A Panel Summary), International Society on Information Fusion, July 2010.