Maritime Domain Awareness via Agent Learning and Collaboration

Dr. Ying Zhao, Research Associate Professor, Naval Postgraduate School
Dr. Douglas J. MacKinnon, Research Associate Professor, Naval Postgraduate School
Dr. Shelley P. Gallup, Research Associate Professor, Naval Postgraduate School
Dr. Charles Zhou, Project Manager, Quantum Intelligence, Inc.

1. Abstract

Maritime security is vital to US security. Enhanced Maritime Domain Awareness (MDA) of potential threats in this dynamic environment can be achieved, yet requires integrated analysis from numerous sources in real time. We will present a learning agent technology that integrates structured and unstructured data and discovers behavior patterns from varied sources such as: Automatic Information Systems (AIS), Coast Guard, and police contextual information including: maritime commercial activities, weather, terrain, environmental conditions, maritime incidents, casualties, and military exercises. These discovered patterns can help correlate warnings and reduce false alarms in support of maritime security. We will show our test results from the Trident Warrior (TW08) exercise.

We will also discuss the agent learning applied to system self-awareness, where we consider that the cognitive interface between decision makers and a complex system may be expressed in a range of terms or “features,” i.e. specific vocabulary to describe a System of Systems (SoS) or so-called *Lexical Link Analysis (LLA)*. MDA is an extremely varied and dynamic SoS, requiring constant collaboration and decision making. We will discuss prototypes of agent learning and collaboration, LLA, and visualization that provide real-time “views” of SoS to support large-scale decision making for MDA technology acquisition, irregular warfare at sea, and intelligence collection with analysis automation.

2. Introduction

Data sources for DOD applications, for example, in the area of intelligence analysis for situation awareness, include disparate real-time sensor and archival sources with multiple dimensions, with very high rates and volumes. The data sources include HUMINT - Human Intelligence, GEOINT - Geospatial Intelligence, IMINT - Imagery Intelligence, MASINT - Measurement and Signature Intelligence, OSINT - Open Source Intelligence, and SIGINT - Signals Intelligence, COMINT - Communications Intelligence, ELINT - Electronic Intelligence, and Special Signals. The data sources could be structured data that are of traditional forms (e.g. stored in relational databases, Excel or XML files with well-defined labels with meta-data). They can be also in unstructured data including free text, word, .pdf, Powerpoint documents, and emails. A large percentage of such data remains unstructured. Retaining logical integrity of the separate data sources, supporting multiple parallel and asynchronous functions of storage, analysis, search, and retrieval, of these data sources while cross-examining all the data to create a full picture of situation awareness remains a daunting task. Analysts need automation tools to facilitate their analysis so they can maintain situation awareness in real-time.

For example, in a recent near-disaster in the Christmas Day attack on a U.S.-bound airliner, the initial analysis from the White House indicated that there is a need for dramatic changes within the U.S. intelligence community to improve its information gathering, dissemination and correlation (White House Report, 2010), specifically,

- U.S. counterterror agency lacks "Google-like" search. Google and other common internet search engines routinely offer alternative spellings for searches, particularly with names. For example, in this attempted attack, Abdulmutallab had been flagged beforehand by U.S. embassy staff in Nigeria, but not under his full name.
- There lacks a particular standard for name-checks once a U.S. visa is granted.
- “The U.S. government at the CIA and the NCTC (National Counterterrorism Center) had sufficient information prior to the attempted December 25 attack to have potentially disrupted the AQAP plot. The problem appears to be more about “connecting the dots” rather than a lack of “information sharing”, as largely considered a problem after 911. The information that was available to analysts was fragmentary and embedded in a large volume of other data.
- Information technology within the counterterrorist community did not sufficiently enable the correlation of all-sources data.
- There was a delayed dissemination of the intelligence reports to *all-source* analysts.

Integrated with database and knowledge management techniques, data and text mining represents an emerging field with a wide range of pattern recognition, visualization and navigation techniques to represent large-scale data as networks of conceptually interrelated nodes. These efforts facilitate search and retrieval, pattern discovery, automated classification and categorization. One of the challenges for mining large-scale data of high rates and large volumes is that the data dimensionality far exceeds the computing and analysis methods available today. Many of the traditional data mining methods, e.g. clustering (Duda et al.,1996), classification (Breiman, 1994), neural networks (Ripley, 1996), decision trees (Breiman, 1994), association rules (Agrawal et al., 1996), are not readily applicable. There are a number of extant tools for mining large-scale text data including advanced search engine (Foltz, 2002; Gerber, 2005), tagging technology (Gerber, 2005), ontology (cf. Tecuci et al., 2007, 2008), etc., yet more powerful and parallel tools are needed.

In summary, we report a system of approaches to address the areas that could improve critical intelligence gathering, analysis and dissemination dramatically:

- It is an architecture not only “Google-like” search but also *anomaly search* and *discovery search* to allow the real-time system self- awareness
- It is an architecture to correlate all-sources data, cross-validate warnings and reduce false alarms.
- The models used for search and learning can be generated from distributed raw data sources with reports from the point of data collection to facilitate timely gathering, analysis and dissemination.

3. Approaches

3.1 Agent learning and collaboration

As illustrated in Figure 1, to automate human cognitive tasks, e.g. to separate and extract information automatically from the documents, we *train* synthetic, learning agents to perform tasks like humans. Modern agent-based modeling and simulation systems originated using concepts such as cellular automation from the game of *Life* invented by John Conway in 1970. This began the development and implementation of genetic algorithms (Goldberg, 1989) and other artificial intelligence techniques to improve the ability of one agent acting alone. From this, agent-based software engineering was invented to facilitate information exchange with other programs and thereby solve problems that cannot be easily solved by a single agent – or human. Synthetic, multi-agent, distributed networks were then developed to provide for an integrated community of heterogeneous software agents, capable of analyzing and categorizing large amounts of information and thus supporting complex decision-making processes. At present, self-managing (Hinchey et al., 2006), self-healing (Dashofy et al., 2002), self-optimizing, and self-configuring, self-adapting, software agents are desirable to be used to automate ongoing human cognitive tasks in a complex network environment.

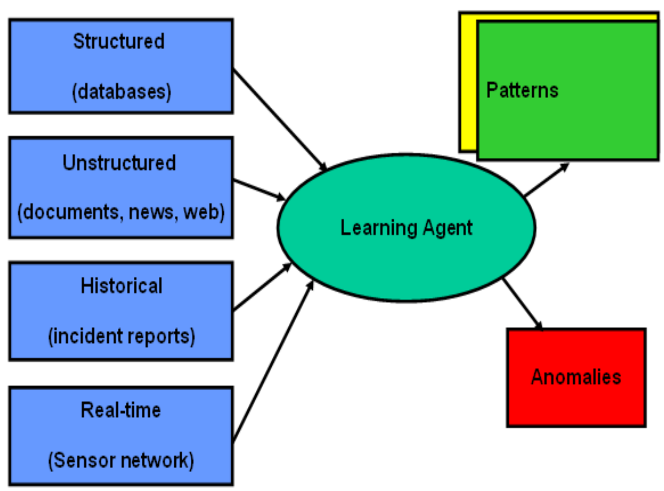

Figure 1: A learning agent ingests structured, unstructured, historical or real-time data and separate patterns and anomalies.

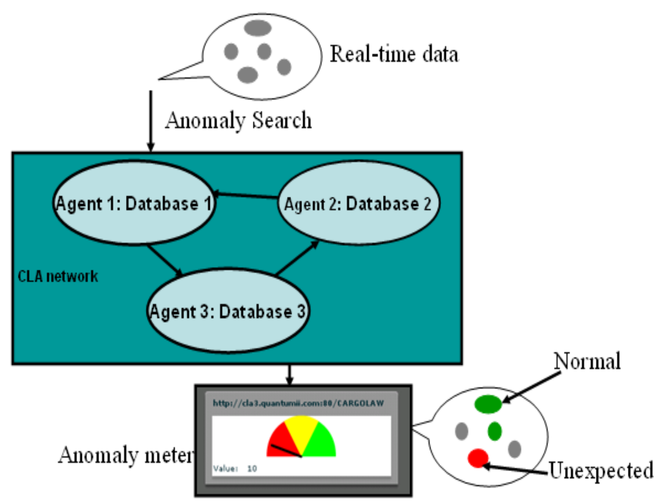

Figure 2: Agent collaboration: multiple agents work together for anomaly search

We believe that employing an alternative strategy of developing an agent learning capability is essential to achieve large-scale automation. Our research creates and develops a computer-based learning agent capable of ingesting and comparing wide range of data sources, while employing a process that separates patterns and anomalies within the data. Multiple agents can work collaboratively in a network as shown in Figure 2. The Collaborative Learning Agents (CLA) were firstly invented and implemented in Quantum Intelligence, Inc. (QI, 2001-2010).

For example, using this architecture, with regards to the Christmas threat, one can apply one type of agent for each of the following information sources

- CIA and NCTC all-source analysis reports

- Pre- and post-US visa databases
- Airport *no fly* list analysis and databases

Each agent indexes and learns the domain information at the point of collection. The raw data do not have to move across network. All the indexes are *shared* across the agents and users. In real-time query such as the report regarding the father's concerns may pass through the three indexes with keywords such person (Abdulmutallab) and location (Yemen). The *hits* in the three agents together for these keywords should indicate there is a real event possibility.

3.2 How the Architecture Improves Information Sharing

Our architecture provides an integrated platform with database and knowledge management data/text mining, pattern recognition, visualization techniques so that large-scale and distributed data are shared as networks of conceptually interrelated nodes to facilitate search and retrieval, pattern discovery, automated classification, and categorization. The characteristics of this architecture can be summarized as follows

- Raw data is stored in logical groups in distributed computer nodes. Raw data can stay in the node where it is collected without moving to a centralized location.
- An agent is installed in each node that is close to and can thereby process the data locally.
- Multiple agents process and index individual and distributed data in parallel and then collaborate to provide a collective view through index sharing.
- This mechanism is designed to search for new and interesting information patterns and anomalies in contrast to the mechanisms used by popular information search engines.

One might think such mechanisms should be already in place, yet they are not - and this task is non-trivial. The data used for indexing and correlation should be as current as possible to address the lesson learned such as "delayed dissemination." In reality, even Google-like search engines need to move the data into a centralized location to be indexed. When it involves many agencies and various government policies, data are not centrally located and therefore, timely processing is not possible. Our agent approach assumes an agent is installed at the data collection point and indexes the data immediately following its collection thus facilitating its indexing in preparation for immediate search and correlation. This mechanism makes the information sharing qualitatively faster and easier at the point of data collection and enables near real-time information sharing.

3.3 Details

3.3.1 Anomaly Search

Situation awareness requires a full spectrum of fusing real-time, historical, structured, and unstructured information to extract patterns and find new, interesting, and anomalous information that are critical for decision making. Existing fusion methods are often used to deal with structured data such as observations from sensor networks. Analysts, at present, must manually wade through unstructured information. Analysts are trained to recognize patterns and anomalies. For example, "intelligence analysts are specifically trained to factor in deception as part of the analytic process, to look for anomalies and outliers instead of focusing on the central tendencies of distribution" (Johnston, 2003). Our research creates and develops a computer-based learning agent capable of ingesting and comparing data, while employing a process that separates patterns and anomalies within the data. At a high level, patterns are correlated among data that occur frequently while recording anomalies among the data that occur rarely. Anomalies that might be interesting are thus revealed so that human analysts are alerted and can further investigate them.

Currently available search engines (e.g. Google-like search) are based on popularity or *authority* scores, which are proven to be useful in marketing and advertising applications but not in the intelligence applications where finding anomalous information is more critical than finding popular information. Our solution is an *anomaly search* mechanism where search results are sorted according to the degree of *anomalousness*, that favors new and interesting information. Using an anomaly meter shown in Figure 8, a piece of information is classified into one of the three categories:

1) Anomaly (red), i.e. a search input that has low correlation with previously discovered patterns
2) Relevant (green), i.e. an input is highly correlated to the previously discovered patterns
3) Medium Correlation (yellow), i.e. between relevant and anomaly
4) Irrelevant (white), i.e. an input is not related to any of knowledge patterns found before

An anomaly meter is used to show the degree of the correlation in Figure 3. If an input has a high, medium, or low correlation (>66.67%, between 33.33% and 66.67%, <33.33% respectively) with the existing patterns, it is classified into the corresponding categories of: Relevant, Medium Correlation, or Anomaly respectively.

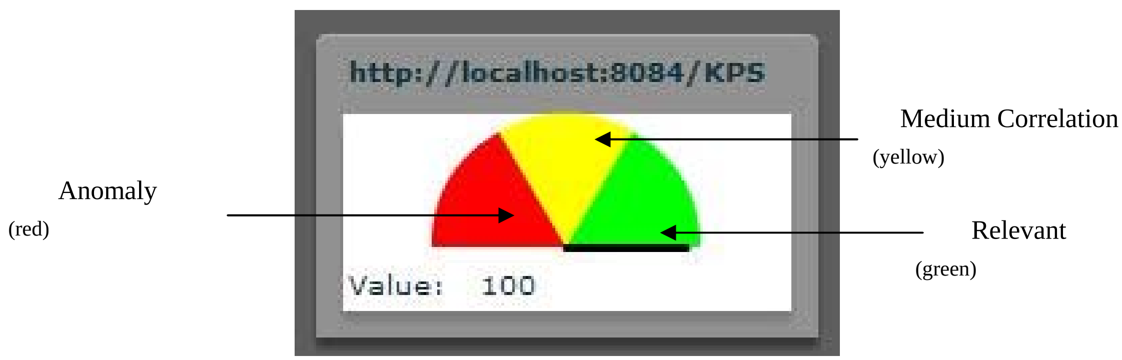

Figure 3: Anomaly Meter for a single agent

A set of agents forms an agent network and performs a collaboration to decide together if an input is expected or anomalous. Each anomaly is classified into one of four categories using the following decision rules:

- An input is Relevant if at least one of the agents decides the input is relevant
- An input is Medium Correlation if the agents cannot decide if it is an anomaly or relevant.
- An input is an Anomaly if all the agents decide the input is anomalous
- An input is Irrelevant if none of the agents decides there is any relevance

3.3.2 All-sources Correlation

How do agents make collaborative decisions using all the data and learned results in the agent network?

A collaborative result of the agents is shown in Figure 4. Critical events are identified – red is an anomaly event and green is a pattern event as a result of fusing all the results from all the agents.

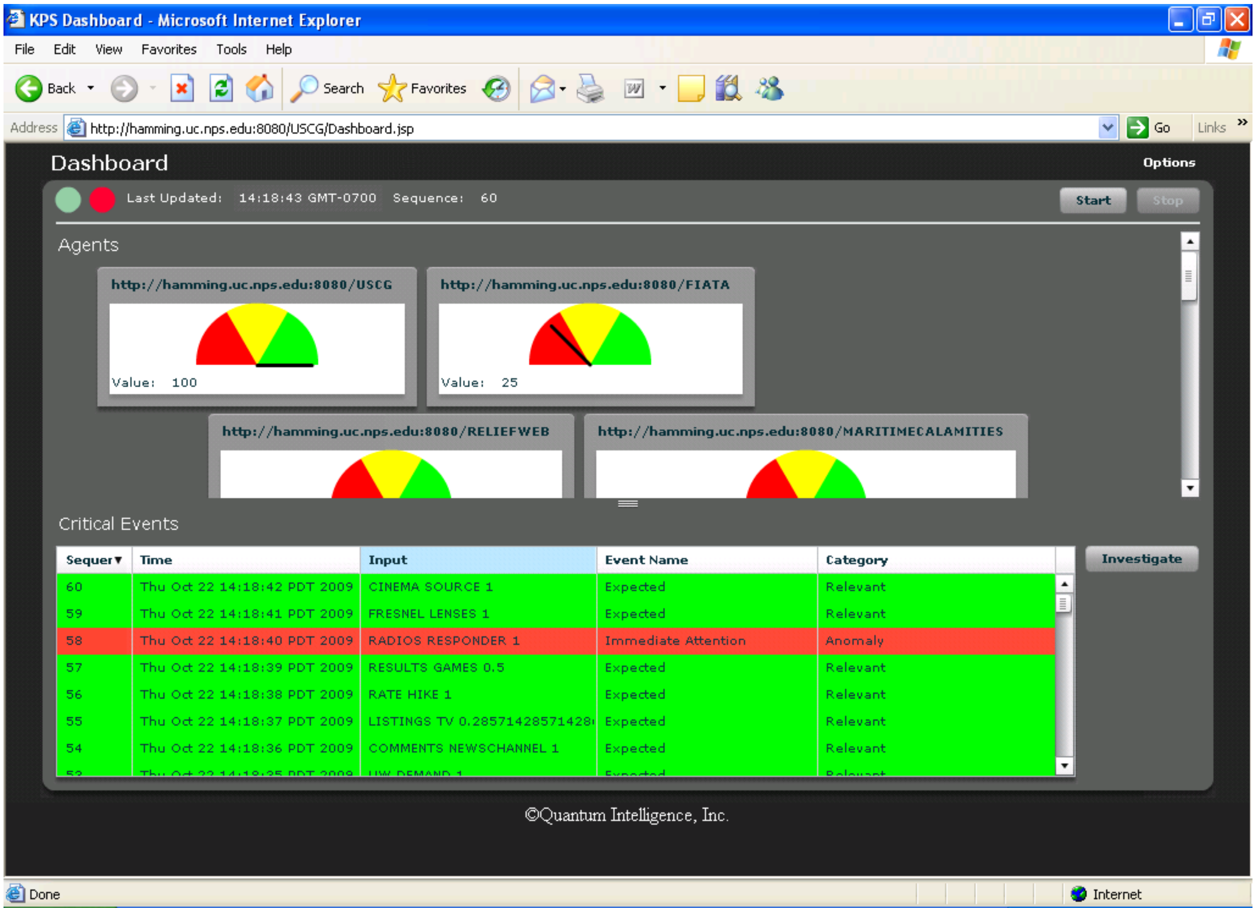

Figure 4: Agent collaboration for anomaly detection.

3.3.3 Discovery Search

As stated before, search terms are required to be input in the current search engines. In contrast, as shown in Figure 4, after the agent learning process in our approach, it outputs a set of word pairs that are discovered from the data. The word pairs are used as input to compare with the learning models (indexes) from all the agents in the network, to produce anomalies (red) and expected (green). The anomalies are discovered terms or concepts driven by data that might be interesting to analysts. In this manner, analysts do not have to come up with terms and concepts for search, the system monitor what is new in the data itself.

3.3.4 Parallel Computation

We have prototyped a multi-agent network of ~10 to 100 agents, that is capable of periodically learning, separating, extracting and visualizing dynamic data. To operationalize this concept, we are working with the NPS (Naval Postgraduate School) High Performance Computing Center (HPCC) to install these agents in the Hamming Linux cluster which provides required supercomputing and visualizations for this project. Servers are also available in the NPS Secure Technology Battle Lab (STBL) for classified data.

4. Trident Warrior Results

Trident Warrior is an annual, Navy FORCEnet Sea Trial exercise to evaluate new technologies that could benefit warfighers. The prototype of a Collaborative Learning Agent (CLA) technology was selected for Trident Warrior 08 (TW08). In the TW08 setup, we used three agents learning patterns from three historical maritime domain information sources. We used one agent to mine a single open-source MDA data: Agent 1 (http://cla1.quantumii.com/FAIRPLAY) for the Lloyd's Register – Fairplay (LRF) news and Agent 2 (http://cla2.quantumii.com/JOC) for the Journal of Commerce. These two sources include information of port, cargo, and vessel activities (e.g. departure/arrival schedules), financial links (e.g. changing owners) and commercial activities (e.g. cargo freight forwarders and custom brokers). Agent 3 (http://cla3.quantumii.com/MPC) is for the Maritime Press Clippings, which consist of worldwide freelance eye-witness descriptions of ships, locations and their activities. The data paints a broader picture of the maritime domain environment that could involve not only people, vessels and cargos, but also the participating partners such as Navy, Police and Coast Guard.

We were able to access the Navy real-time vessel AIS data from SPAWAR DS COI (SPAWAR data sharing, community of interest, https://mda.spawar.navy.mil) as shown in Figure 5. The MDA DS COI has AIS-based track information and associated alerts including data from Navy Organic Sensors aboard Navy ships, the Department of Transportations (DOT), The United States Coast Guard (USCG), Office of Naval Intelligence (ONI) to track merchant shipping. The data is published as the NCES Messaging Service that can be integrated with standard web services. The data show worldwide real-time ship's names and locations.

Figure 5: SPAWAR data sharing, community of interest, https://mda.spawar.navy.mil

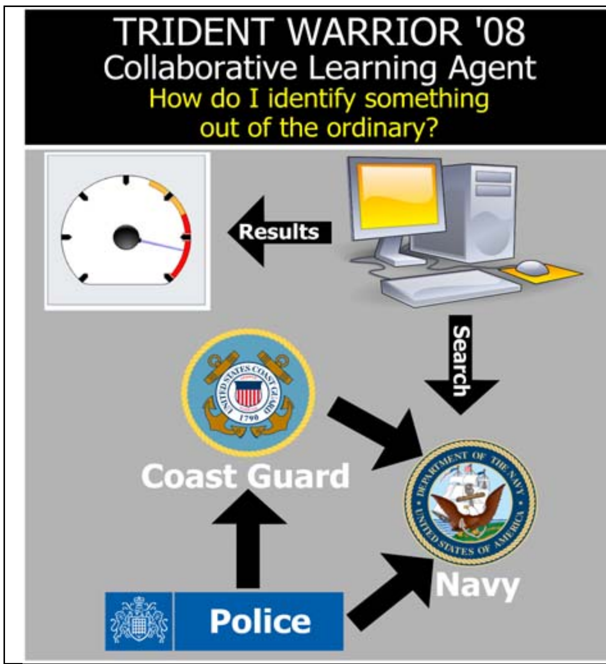

Figure 6: The CLA concept in TW08

In the TW08, a test responder observed the test process from real-time inputs to assess the relevance and accuracy of the Collaborative Learning Agent (CLA) during input events. Each input (sequence) represents a vessel's name or real-time location from the SPAWAR MDA DS COI. The input was checked against the patterns produced in the CLA network to see if context patterns involving people, places and events are of interest or relevance to the current vessel or its location at alert. This process is summarized in Figure 6.The TW08 results show the overall accuracy for the CLA prediction of anomalies is 72% (Zhou,et al., 2009).

5. Related Work to Others

Agent learning is related to a reinforcement learning framework (Sutton 1998) that assumes a system is characterized by a set of states and transitions between them. Each model includes one or more decision makers, sometimes referred to as an agent, controller, or player(s) when appropriate. Each can exert control by affecting system transitions, Bayesian belief networks (Pearl, 1986; Ben-Gal, 2007), Hidden Markov Models (Huang 1990), which is an important agent learning paradigm where only partial information can be observed and the rest has to be predicted. Agent learning is also closely related to the wealth of research on information extraction and integration (Arasu, et al., 2003), where patterns are extracted as structured data, and derived relationship discovery (Roth et al., 2002). The process is conceptually linked to a full text indexing in the traditional information retrieval area where inverted files are computed for an original text (Frakes et al., 1992). The advantage of the algorithm over the traditional methods is that it captures the cognitive level of understanding of text observations using a few key concepts. This algorithm also makes it possible to separate patterns and anomalies for unstructured data and thus is closely related to search index methodologies.

Agent collaboration is related to distributed knowledge management architecture (Bonifacio, M., et al., 2002) consisting of *knowledge nodes* (peers that own local knowledge) that cooperate through a set of services. Each node can act as a *seeker* (allowing the user to search for information and forwarding queries), as a *provider* (accepting and resolving queries and returning results) or both. Searching can be either semantic (based on a context matching algorithm) or textual (based on keywords). Groups of nodes have the ability to form *federations*, similar to social aggregations, by agreeing to be considered as a sole entity by other peers when they request certain semantic search services. This collaborative infrastructure is a peer-based system, where agent-like applications are distributed among a grid of computers. Each application is considered itself to be a peer or node among a network of similar applications. The infrastructure is "fault-tolerant", "distributed", and "self-scalable" (Ahmad, H. F., et al., 2005). Yet, among the peer-based systems, there lacks a full-text analysis capability to discover new structures, patterns, and relationships.

Agent collaboration is also related to social network research. The results of social network research routinely identify communication networks, knowledge clusters (Nissen, 2006), and shared cognition in organizations (Moreland, 1999). Social network analysis (Hoff, 2002) is widely used to analyze relational information among interacting units. This framework has many applications in recent years in the social and behavioral sciences including, the behavior of epidemics and dynamics associated with terrorist networks. Social network research is also related to information retrieval and text analysis. For example, the search engine Google uses the interconnectedness of the World Wide Web for page ranking (Brin et al., 1998). The Google page ranker is link-based by incorporating the human confirmation of search results as "authority score" of a specific page. This is a successful application of a collective social network effect leveraging collaborative feedback. In our case, the agent peers can be physically connected as in a traditional peer-based system or logically connected through shared concepts.

Our solution uniquely couples agent learning and collaboration to generate a data driven, dynamic ontology or semantic markup for unstructured data that can significantly increase the automation. This enables the formation of a real-time network to achieve so-called optimized “swarm intelligence” (Bonabeau et al., 1999) with desired collective behavior in a decentralized, self-organized environment.

The Linux cluster is similar to the Cluster Exploratory (CluE) program, which made available for data-intensive computing projects a massively scaled highly distributed computing resource supported by Google and IBM and a similar resource at the University of Illinois in partnership with Hewlett-Packard, Intel, and Yahoo!. Our results will show what the potential benefits of similar cluster technology may have for science and engineering research as well as to applications that may benefit society more broadly.

6. Conclusion

The agent learning and collaboration architecture demonstrated in this paper address the deficiencies in the current intelligence analysis with the key characteristics and innovations such as indexing and learning of raw information as both data and index at the point of collection. The agent learning provides unique anomaly and discovery search mechanisms as the foundation for large-scale automation of timely digesting and analyses wide variety of intelligence. The agent collaboration provides the base for all-sources intelligence correlation that is critically needed to improve intelligence gathering, correlation and dissemination in *irregular warfare*. The architecture is shown to be extended to Lexical Link Analysis (LLA) which can be used in many MDA applications.

7. Future (Near Term) Efforts

7.1 MDA all open sources correlation

We extend the unstructured information sources in MDA in *irregular warfare* at sea and intelligence efforts to obtain open-source data from news, blogs, journal of commerce abstracts, freight forwarder associations/custom brokers websites with regarding to maritime commercial activities, arrival schedules, weather, terrain, environmental conditions, maritime incidents, casualties, and military exercises. The following Table 1 is a list of sample open sources (about 100 sources) for the MDA application. Each source will be monitored by 10 agents for one week period. The agent learning will occur daily and total data size is around 1G per day.

Table 1: MDA open sources list

Vessel ID, Location, Images http://82.146.41.123/ships.htm http://www.vesseltracker.com/en/VesselArchive.html http://www.digital-seas.com/start.html http://www.marinetraffic.com/ais/default.aspx?centerx=30&centery=25&zoom=2&level1=140 Piracy Reporting http://www.lloydslist.com/ll/media/presentation.htm http://www.icc-ccs.org/index.php?option=com_fabrik&view=visualization&controller=visualization.googlemap&Itemid=219	http://www.lloydslist.com/content/rss/lloydslist/piracy_security.xml http://www.lloydslist.com/content/rss/lloydslist/ports_terminals.xml http://www.lloydslist.com/content/rss/lloydslist/print_edition.xml http://www.lloydslist.com/content/rss/lloydslist/ship_management.xml http://www.lloydslist.com/content/rss/lloydslist/shipbuilding_repair.xml http://www.lloydslist.com/content/rss/lloydsl

http://www.imo.org/Circulars/mainframe.asp?topic_id=334&offset=0
Port Operations
http://www.portvision.com/Public/index.aspx
http://www.pier2pier.com/
Container tracking & security
http://www.track-trace.com/container#
http://www.transportsecurity.com/company.php
Weather
http://www.sailwx.info/shiptrack/index.html
http://news.bbc.co.uk/weather/
http://earth.esa.int/ers/eeo4.10075/atsr_med.html
http://www.mediterraneanweather.com/satimages.htm
http://oceancolor.gsfc.nasa.gov/
Shipping Schedules and Lines
https://www.oceanschedules.com/schedules/search.do
http://www.howdydave.com/maritime/shipping.html
Distance Measurement tool
http://jan.ucc.nau.edu/~cvm/latlongdist.html
Marine Services directories
http://www.infomarine.gr/index.php
http://www.madmariner.com/?gclid=CLSgsbrAg50CFVtB5godckpvbw
http://seann.org/Directories/introNew2.asp
http://www.infomarine.gr/greece/
http://www.best-maritime.info/index.php/l/en/mod/companies
http://www.m-i-link.com/directory/profile.asp?bz=21&id=12446&cat=Ship+Manager+%26+Owner
Shipwreck Database and Casualty Reports
http://www.shipwreckregistry.com/
http://www.cargolaw.com/presentations_casualties.php
Other
http://www.lloydslist.com/content/rss/lloydslist/all_headlines.xml
http://www.lloydslist.com/content/rss/lloydslist/breaking_news.xml
http://www.lloydslist.com/content/rss/lloydslist/classification.xml
http://www.lloydslist.com/content/rss/lloydslist/containers.xml
http://www.lloydslist.com/content/rss/lloydslist/cruise_ferry.xml
http://www.lloydslist.com/content/rss/lloydslist/dry_cargo.xml
http://www.lloydslist.com/content/rss/lloydslist/environment.xml
http://www.lloydslist.com/content/rss/lloydslist/insurance.xml
http://www.lloydslist.com/content/rss/lloydslist/lng_lpg.xml
http://www.lloydslist.com/content/rss/lloydslist/logistics.xml

ist/tankers.xml
http://www.lloydslist.com/content/rss/lloydslist/towage_salvage.xml
http://www1.apan-info.net/
http://www.cargobusinessnews.com/
http://www.cargolaw.com/
http://www.uscg.mil/
http://www.piersystem.com/
http://feeds.feedburner.com/CoastGuardNews
http://www.fairplay.co.uk/feed.aspx
http://www.marinelink.com/Story/
http://www.shipspotting.com/modules/altern8news/?
http://www.ifw-net.com
http://maritimecalamities.blogspot.com/
http://www.maritimematters.com/shipnews.html
http://www.maritimematters.com/shipnewspics.html
http://www.fiata.com/index.php?id=95
http://www.airbus.com/en/
http://www.asycuda.org/
http://www.biac.org/
http://www.bimco.org/
http://www.cen.eu/cenorm/homepage.htm
http://www.boeing.com/
http://www.ecac-ceac.org/index.php
http://www.clecat.org/
http://www.efta.int/
http://www.fao.org/
http://www.fidi.com/index.html?page=40&lang=en&
http://www.gfptt.org/Entities/NewsList.aspx?list=all
http://www.iaphworldports.org/
http://www.iccwbo.org/
http://www.ifcba.org/modules/news/index.php
http://www.imf.org/external/index.htm
http://www.imo.org/
http://www.tiaca.org/
http://www.ttclub.com/ttclub/public.nsf/html/index?OpenDocument
http://www.unescap.org/
http://www.ncbfaasa.org/NEWSLETTER.pdf
http://www.aapa-ports.org/Press/PRMemberList.cfm?navItemNumber=537
http://www.reliefweb.int
…

7.2 Lexical Link Analysis

We generalize and operationalize the notion of "awareness" and determine the "self-awareness" of a highly complex system as the collective and integrated understanding of large-scale data. The concept of system self-awareness (Gallup et al., 2008) allows decision makers and analysts to sift through large-scale data sources with greater immediacy, possibly in real-time by training synthetic, computer agents to automate the task of recognizing, separating and visualizing patterns, in an effort to reduce the workload of decision makers and analysts who are otherwise performing the task manually. We propose to extend this approach to employ agent learning, collaboration, Lexical Link Analysis (LLA) (Gallup, et al., 2009) and visualization in a massive parallel fashion, to provide real-time "views" of situation awareness for irregular warfare intelligence collection and analysis.

Lexical analysis is a form of data/text mining, in which word meanings are developed from the context in which they are derived (LA wiki, 2009). Lexical analysis can also be used in a learning mode, where such word and context associations are constantly being "learned," updated and improved as more data become available - or changes are made to information already in use. Lexical Link Analysis (LLA) is an extended lexical analysis combined with learning and mining from real data and data relations.

After the agent learning process, the system produces a set of word pairs as the features of a System of Systems (SoS) being considered. LLA is applied to use the features to link the similarities and gaps among requirements and technologies for MDA acquisition communities (Zhao, et al., 2009). We currently apply the LLA to different priority lists and acquisition targets, for example, CENTCOM/NAVCENT warfighting gap and priority lists, data from OSD DIMHRS, which, due to classification cannot be discussed in detail here.

We also currently apply the LLA method to The Joint Defeat IED program. The Lexical Analysis piece of this is very central, but it is most useful when combined with the distribution and C2 of information within intelligence and tactical communications. In the near future we will integrate the knowledge gained from LLA (Lexical Link Analysis) into other intelligence fusion and decision making tools/processes.

We also seek to apply our technique to Defense Analysis to re-construct social networks and predict terrorist links that had been previously unknown.

8. References

1. Agrawal, R., Imielinski, T., Swamim, A., 1993. Mining Associations between Sets of Items in Massive Databases. Proc. of the ACM SIGMOD Int'l Conference on Management of Data, Washington D.C., 207-216.
2. Arasu, A. and Garcia-Molina, H., 2003 Extracting structured data from Web pages. In SIGMOD-03. http://infolab.stanford.edu/~arvind/papers/extract-sigmod03.pdf
3. Ben-Gal, I.,2007. Bayesian Networks, in F. Ruggeri, R. Kenett, and F. Faltin (editors), Encyclopedia of Statistics in Quality and Reliability, John Wiley & Sons.
4. Bonabeau, E., Dorigo M., Theraulaz, G. 1999. Swarm Intelligence: From Natural to Artificial Systems, SBN 0-19-513159-2
5. Bonifacio, M., et al., 2002. A peer-to-peer architecture for distributed knowledge management. in the 3rd International Symposium on Multi-Agent Systems, Large Complex Systems, and E-Businesses.

6. Breiman, L., et. Al., 1984, Classification and Regression Trees, Wadsworth, Belmont, 1984
7. Chief of Naval Operations, 2009. Establishment of the Navy Maritime Domain Awareness Office. Administrative Message 181837Z MAR 09. Washington, DC.
8. Dashofy, E.M., van der Hoek, A., Taylor, R.N., 2002. Toward Architecture-based Self-healing Systems. The ACM Digital Library, Proceedings of the first workshop on Self-healing systems.
9. Department of Homeland Security, 2005. National Plan to Achieve Maritime Domain Awareness for The National Strategy for Maritime Security. Washington, DC.
10. Doan A. and Halevy A., 2005. Semantic Integration Research in the Database Community: A Brief Survey. AI Magazine, Special Issue on Semantic Integration, Spring 2005.
11. Duda, R. O., Hart, P. E., Stork, D.G., 1996. Pattern Classification and Scene Analysis, John Wiley & Sons, NY, 2nd edition.
12. Dumais, S.T., et al., Using Latent Semantic Analysis to Improve Information Retrieval. In Proceedings of CHI'88: Conference on Human Factors in Computing, 1988: p. 281-285.
13. Ferrucci, D., Lally, A., 2004. UIMA: an architectural approach to unstructured information processing in the corporate research environment, Natural Language Engineering, Vol. 10 , Issue 3-4.
14. Foltz, P.W., 2002. Quantitative Cognitive Models of Text and Discourse Processing. In The Handbook of Discourse Processes. Mahwah, NJ: Lawrence Erlbaum Publishing
15. Frakes, B. and Baeza-Yates, R. eds., 1992. Information Retrieval Data Structures & Algorithms. Prentice-Hall: Englewood Cliffs, New Jersey.
16. Ahmad, H. F., et al., 2005. "Scalable Fault Tolerant Agent Grooming Environment - SAGE Agent Platform", 4th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS), pp.125-126, 25 - 29 July 2005, Netherlands
17. Gallup, S., MacKinnon, D. J., 2008, Status Assessment of Maritime Domain Awareness Capability Development. Monterey, CA: Naval Postgraduate School.
18. Gallup, S. P., MacKinnon, D. J., Zhao, Y. Robey, J. Odell, C. Facilitating Decision Making, Re-use and Collaboration: A Knowledge Management Approach for System Self-Awareness. *International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management (IC3K)*, Madeira Portugal, 6-8 October, 2009.
19. Gerber, C., 2005. Smart Searching, New technology is helping defense intelligence analysts sort through huge volumes of data. In Military Information Technology, 9(9).
20. Girvan, M. and Newman, M. E. J. 2002. Community structure in social and biological networks. http://www.pnas.org/content/99/12/7821.full.pdf.
21. Goldberg, D. E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, Massachusetts, 1989.
22. Hinchey, M.G.; Sterritt, R.; 2006. Self-Managing Software, Computer, Volume 39, Issue 2
23. Hoff, P.D., Raftery, A.E. and Handcock, M.S., 2002. Latent Space Approaches to Social Network Analysis. Journal of the American Statistical Association, 97. http://elvis.slis.indiana.edu/fetched/article/1129.htm
24. Hoff, P.D., Raftery, A.E. and Handcock, M.S., 2002. Latent Space Approaches to Social Network Analysis. Journal of the American Statistical Association, 97. http://elvis.slis.indiana.edu/fetched/article/1129.htm
25. Huang, X. D., Ariki, Y., Jack, M. A., 1990. Hidden Markov Models for Speech Recognition. Edinburgh: Edinburgh University Press.
26. LA wiki, 2009. http://en.wikipedia.org/wiki/Lexical_analysis

27. Moreland, E.L., 1999. Transactive memory: learning who knows what in work groups and organizations. Shared cognition in organizations: the management of knowledge Mahwah, ed. J.M.L. L.L. Thompson, & D.M. Messick. NJ: Lawrence Erlbaum Associates.
28. Nissen, M.E. 2006. Harnessing Knowledge Dynamics: Principled Organizational Knowing and Learning, Hershey, PA: IRM Press.
29. Nadeau, D., Turney, P. D. and Matwin, S., 2006. Unsupervised Named-Entity Recognition: Generating Gazetteers and Resolving Ambiguity. Canadian Conference on Artificial Intelligence.
30. NER wiki 2009. Named Entity Recognition , http://en.wikipedia.org/wiki/Named_entity_recognition
31. Pearl J.,1986 Fusion, propagation, and structuring in belief networks. Artificial Intelligence 29(3):241–288.
32. QI, 2002-2010, http://www.quantumii.com
33. Ripley, B.D., 1996. Pattern Recognition and Neural Networks, Cambridge University Press.
34. Robey, J., Odell, C. 2009. Facilitating Decision-making, Re-use and Collaboration: a Knowledge Management Approach to Acquisition Program Self-awareness. 6th Annual Acquisition Research Symposium Presentations and Papers.
35. Schensul, J. J., Schensul, S. L., and LeCompte, M. D., 1999. Essential ethnographic methods: observations, interviews and questionnaires, Rowman Altamira.
36. Sutton, R.S., Barto, A. 1998. Reinforcement Learning: An Introduction, MIT Press, 1998.
37. Tecuci G., Boicu M., Marcu D., Boicu C., Barbulescu M., Ayers C., Cammons D., 2007. Cognitive Assistants for Analysts, in John Auger, William Wimbish (eds.), Joint publication of the National Intelligence University, Office of the Director of National Intelligence, and US Army War College Center for Strategic Leadership.
38. Toutanova, K., Klein, D., Manning, C. and Singer, Y. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.
39. WordNet (dictionary), 2009. http://wordnet.princeton.edu/
40. White House Report, 2010, http://www.whitehouse.gov/sites/default/files/summary_of_wh_review_12-25-09.pdf
41. Zhao, Y. and C.G. Atkeson, Implementing Projection Pursuit. The IEEE Transactions on Neural Networks, 1996. 7(2): p. 362-373.
42. Zhou, C., Zhao, Y., and Kotak, C., 2009. The Collaborative Learning Agent (CLA) in Trident Warrior 08 Exercise. *International Conference on Knowledge Discovery and Information Retrieval*, Madeira Portugal (KDIR), October, 2009
43. ZHAO, Y. GALLUP, S. P. MACKINNON, D. J. AND KENDALL, A. “REAL-TIME PROGRAM SELF-AWARENESS FOR AGILE ACQUISITION OF C4I SYSTEMS", GMU - AFCEA Symposium 2010, Fairfax, Virginia, 18-19 May 2010