OSINT Analysis using Adaptive Resonance Theory for Conterterrorism Warnings.

Conference Paper · January 2005

Source: DBLP

CITATIONS
2

READS
113

1 author:

Jami M. Carroll
Capitol Technology University
2 PUBLICATIONS **2** CITATIONS

SEE PROFILE

OSINT ANALYSIS USING ADAPTIVE RESONANCE THEORY FOR COUNTERTERRORISM WARNINGS

Jami M. Carroll
Nova Southeastern University
3301 College Avenue
Fort Lauderdale, FL 33314-7796
USA
cjami@nova.edu

ABSTRACT

Open Source Intelligence (OSINT) is an extremely valuable source of data for intelligence analysts in identifying and analyzing potential terrorism warnings and indicators [1]. A key problem is making sense of this large amount of data in time to prevent a catastrophic situation like what occurred on September 11, 2001. These events might been mitigated had U.S. intelligence agencies had better information technology tools for analyzing the situation according to the report of Congress's Joint Inquiry into the events leading up to the Sept. 11 attacks [2]. Improvements to Information and Communications Technologies (ICTs) are necessary to provide the Homeland Security with the proper support for their missions [3].

Artificial Neural Networks (ANNs) have been around for over half a century and their biologically-inspired capability allows functionality similar to the human brain via simulated neurons that can make near-human choices. ANNs are in their genesis with future applications include finance, marketing, medicine and security in data mining [4]. Data mining enables a large amount of data to be sifted and provide avenues to learn or generalize information about that data using feature extraction [5]. Adaptive Resonance Theory (ART) may provide another tool for this analysis.

KEY WORDS
Terrorism, OSINT, feature extraction, and neural networks

1. Introduction

Data mining of open source, Internet web data provides a large amount of information for the identification and analysis of potential terrorism warnings and indicators. Even with web bots or web spiders, the data harvest process would be extremely slow if performed through typical, sequential computer processing. This level of "feature extraction" is looking for specific keywords through clustered datasets. This requires a significant increase in the speed of processing to be effective. Through the employment of neural networks, parallel processing of the data could improve the rate of processing over typical serial, sequential methods. ART is a form of a neural network used by the e-commerce community to help consumers select other desirable options with choices like, "if you like product "A" you may like product "B" [6]. This same capability in ART could be used to more rapidly process potential terrorism warnings and indicators by looking for similar trends with similar keywords. Quicker analysis should result in quicker processing of potential terrorism warnings and indicators before they occur. ART could expand the searching capability to more effectively prosecute variations in potential keywords under search.

2. Why ART?

ART offers a lot of different advantages. It allows the opportunity to use data mining from a wide range of sources. ART is not required to operate linearly, therefore it is more adaptable to real-world, non-linear analysis. ART is relatively simple in its operation.

2.1 Data Mining

Data mining via the web provides a wealth of information. The restrictions of working solely within a database are expanded to structured and unstructured data. The web as a conduit provides access to email, web sites, multimedia, and video all with different data attributes. Visualization of this data through data warehousing techniques becomes even more complicated with varying types. Business Intelligence is already being applied in this area to improve e-commerce as new opportunities become available via the semantic web, search engines, and the global connectivity within a social network that becomes greater everyday [7].

2.2 Artificial Neural Networks

The concepts of Artificial Neural Networks (ANNs) have been around since the 1950s and are biologically-inspired

from the view of the human brain as a processor using interconnected neurons. The ANN is connected like the brain with artificial neurons that are interconnected and adaptive to the output of other connected nodes which have modifiable parameters. The brain is very complex, non-linear, and capable of parallel processing and performs pattern recognition, perception, and motor control much faster than most computers and the computing power of an ANN is derived from the parallel, distributed structure of its architecture and the ability to learn or generalize data [8]. Some of the advantages an ANN include:

- **Non-linear** – Many real-world entities are non-linear and the ANN is capable of acting on data that is linear or non-linear.
- **Learning through training** – By tuning the parameters of ANNs, they can be adapted to many different models.
- **Adaptive** – The ANN can be assimilated to a wide-range of environments.
- **Evidential Response** – ANNs can provide classification or feature extraction of data.
- **Fault Tolerant** – The parallel nature of the neurons in the ANN lend to a certain degree of robustness.
- **Uniform Analysis and Design** – Similar approaches with how ANNs are discussed, designed, and integrated.

2.3 Adaptive Resonance Theory

The seminal paper by Stephen Grossberg and Gail Carpenter [9] created adaptive resonance theory (ART) which lead to a family of algorithms consisting of ART-1, ART-2, ART-3, ARTMAP (supervised), Fuzzy ARTMAP, and Multichannel Fuzzy ART. ART1 is a very simple, binary data, unsupervised learning ANN that performs data clustering and has the capability of personalization, sometimes referred to as a recommender system.

Clustering algorithms take a data set and groups the data into smaller sets of similar and sometimes dissimilar data. Once it is clustered into groups, it can be separated into groups for classification and feature extraction. Clustering of data is similar in how we learn – we separate the things we know and relate new concepts with known concepts. If we run across new concepts that we do not have any way of relating the data to – we create a new structure for the new data. Some ANNs have encountered the problem termed by Grossberg and Carpenter as the "stability-plasticity dilemma" where they learn new things; they sometimes destroy some of the old data clusters from "old learning". The ART-1 has features that prevent much of this from occurring.

In ART-1, objects are represented as feature vectors represented in binary format. Many Internet shopping kiosks, like Amazon and Barnes & Noble, will use feature extraction of customers' purchases to provide another customer with new potential choices. They are often labeled, "If you liked book "A", you may also like book "B". ART becomes a powerful tool when a new customer is about to purchase a book. Accompanying *AI Application Programming* is "C language code" that illustrates this capability with multiple purchasers purchasing from eleven types of articles from three main groupings: tools, desk items, and snacks [6]. The eleven possible purchased items with two customers is shown in Table 1.

Table 1. Consumer Purchases

Customer	Hammer	Paper	Snickers©	Screwdriver	Pen	Kit-Kat©	Wrench	Pencil	Heath Bar©	Tape Measure	Binder
1	0	0	0	0	0	0	1	0	0	0	0
0	1	0	0	0	0	0	0	1	0	0	1

Sageman discusses the current state of terrorism and the impact al Qaeda has made to the world. Applying another excellent tool, Social Network Analysis, he illustrates some of the key social network structure attributes related to 11 September 2001 [10]. From these attributes, I have made minor modifications to the code provided with the *AI Application Programming* book by Jones in accordance with the author's prescribed copyright procedures to include eleven entities from three major groupings: key players, terror methods, and object location for potential terror. Table 2 illustrates the eleven terms I used for feature extraction in my proof-of-concept model of how ART-1 could be used to analyze OSINT data and in identifying and analyze potential terrorism warnings and indicators in the war against terrorism. The Data Source column indicates an email received with these key terms enumerated to binary "1" or "0" values from a parsed email, human intelligence (HUMINT) contact, document, multimedia file, web spider extraction, etc. In this simplistic example ten emails are inputted into the ART-1 code with a "1" implying "true" for that entity or a "0" for the entity if "false".

Table 2. Terrorist Threat Considerations

Data Source	Osama-bin-Laden	Dirty-Bombs	Boston	Khalid-Sheikh-Mohommed	Plane-Crash	Los-Angeles	Zein-al-Abidin-Mohammed-Hussein	Building-Explosion	Chicago	Abu-Bakar-Baasvir	Nuclear-Power-Plant-Explosion
0	0	0	0	0	0	1	0	0	1	0	0
1	0	1	0	0	0	0	0	1	0	0	1

2	0	0	0	1	0	0	1	0	0	1	
3	0	0	0	0	1	0	0	1	0	0	1
4	1	0	0	1	0	0	0	0	0	1	0
5	0	0	0	0	1	0	0	0	0	0	1
6	1	0	0	1	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	0	1	0	0
8	0	0	0	0	1	0	0	1	0	0	0
9	0	0	1	0	0	1	0	0	1	0	0

2.4 ART-1 Learning

The "prototype vector" is what each inputted "feature vector" is compared against. The "prototype vector" changes as "learning" takes place. At initialization, the ART-1 code has no rules and no learning – all conditions in the prototype vector are preset by the first received "prototype vector". As each email with the "feature vectors" it provides is parsed into the ART-1 code, this feature vector is tested against the "prototype vector"; this causes new clusters to be created or another input to be added to an existing cluster. As the ART-1 algorithm vacillates between adding an input to an existing cluster and creating a new cluster, the condition of "resonance" occurs during the learning process.

Since the feature vectors contain attribute information about an intelligence source, each of these attributes are compared against the prototype vectors; this allows for the creation of the applicable clusters. Clustering now allows for the analysis of commonality within each of these clusters to "learn" potential traits about them. Some of these traits may indicate similar interests within the clusters – in my model, they would display some of the threat consideration groupings like terrorists, terror methods, locations of terror focus, as well as other potentially emerging clusters. Figure 1 illustrates the four clusters developed under the current parameters:

- Cluster 0: Emails 0, 7, 9
- Cluster 1: Emails 3, 8
- Cluster 2: Emails 2,4, 6
- Cluster 4: Emails 1, 5

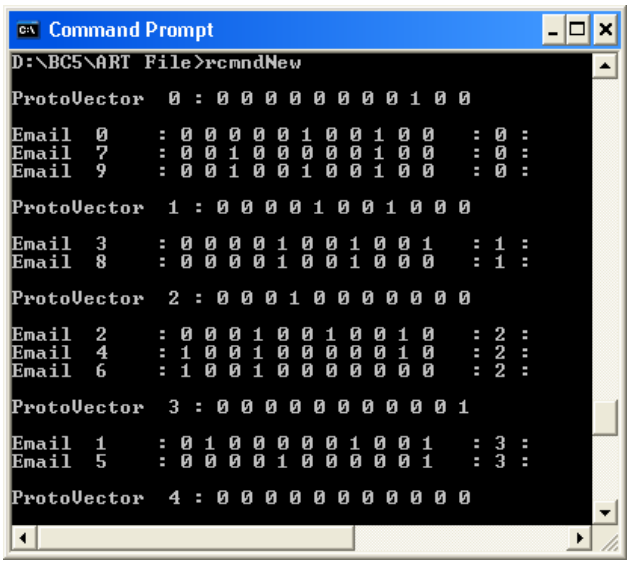

Figure 1. ART-1 Clustering

2.5 ART-1 Parameter Settings

The prototype vector is set to what is referred to as "the center of the cluster" during initialization of the ART-1 algorithm. The maximum number of prototype clusters is set with the value "N". The rho "ρ" is set for Vigilance parameter ($0 < p <= 1$). The beta "β" is set for some small positive value ($0< \beta < 1$). In this model, the variables are:

- $\beta = 1.0$
- $\rho = 0.9$
- $N = 4$
- Prototype Vector Initialization: Based on the first email Prototype Vector attribute received. (Shown in Table 2 as Email 0.)

The beta, rho, and "N" should be tested with varying conditions based on data population to ensure the best results are achieved [11].

2.6 ART-1 Operation

The ART-1 operation is a sequence of steps. As each step is completed, the next one is accomplished. Upon completion, a final check of clustering is rechecked.

- Once the first feature vector is parsed into the ART-1 algorithm, the ART-1 prototype vector is initialized.
- Each feature vector continues to be parsed into the ART-1 algorithm for comparison of how close that feature vector is compared to the prototype vector.
- The beta is used as a "tie-breaker" that favors "1s or trues" over "0s or falses".
- If the closeness or proximity is met, the ART-1 algorithm then incorporates the vigilance test

with rho. In most cases, the vigilance parameter will be set to a number closer to "1" to ensure fewer groups with more group members. If the vigilance is set low (closer to 0), extreme fragmentation of clusters may occur.

- If the vigilance test passes, the current feature vector and the current prototype vector go through a bitwise AND operation. If the proximity and vigilance test does not pass, the next prototype vector is checked, if this one does not pass all other prototype vectors are tested until they have been exhausted. If after testing all prototype vectors, both tests do not pass – a new cluster is created.
- Once all of the feature vectors have been assigned to clusters, the ART-1 algorithm rechecks all clusters to make sure that they were not inadvertently placed in an incorrect cluster during the process. To prevent excessive oscillations, the number of iterations is controlled within the ART-1 algorithm; this number is set high enough to prevent the ART-1 from achieving a premature solution.

2.7 ART-1 Personalization/Recommender

Just as business intelligence uses personalization with ART to personalize and recommend a book to the consumer on Amazon using techniques called "Collaborative Filtering" which provides critical tuning and filtering; OSINT analysts trying to determine the next steps that terrorists may also employ similar approaches. ART can be performed online or offline – this just-in-time decision making could provide one more of the tools to assist analysts with time critical decision making. A simplistic approach to this filtering is by analyzing the differences seen in a cluster and based on the difference; this may indicate a "future threat consideration". Figure 2 illustrates some of the recommendations for each of the ten emails. For each of the emails, a "best threat consideration" based on terrorists, terror methods, or terror locations are asserted. Other entities within the same cluster illustrate their possible threats associated with the recommended "terrorist threat".

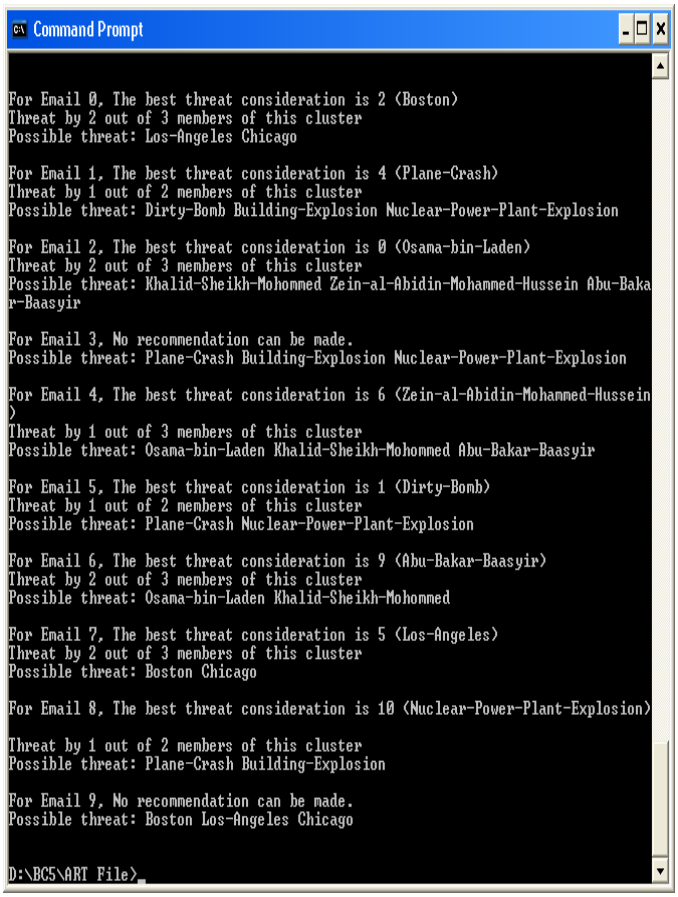

Figure 2. ART-1 Personalization/Recommender

3. Conclusion

In conclusion, I will address the advantages, limitations, possible applications, and future work planned for this project.

The advantages of using ART-1 are that it is fast and adaptable to a wide range of data types, once properly pre-conditioned for the data and with some degree of tuning to get the ART-1 algorithm configured for the data set under test.

Some of the limitations include a model that works entirely on binary data. This requires pre-formatting of data prior to processing; in many case this data is collected in a database for later processing and retrieval. I have three options planned for the future as options:

- Use the ART-2 algorithm which supports binary or analog data inputs. ART-2 expands the F1 Layer into several different additional layers [11].
- Create a search/parser module to pre-parse the data. This will offload the functionality to another module that is not directly ART-related. An added benefit is that this capability could be added to a parallel processor set of systems to more rapidly process large data sets.

- Use the capability of a database to perform the pre-processing prior to inputting the data to the ART-1 capability.

As I have outlined, ART as well as other ANNs offer a significant capability for use with handling OSINT data from a variety of sources. As these models for handling this type of data are improved, I expect that improvements in feature extraction, clustering, and tuning of the algorithms will also improve.

My plans for future work on this model will be to improve on the handling of binary data, improve on the ability of the ART program to except a file which can be directly parsed into the program. This expansion will include an ability to dynamically create the database array. Another smaller feature would also include a timestamp so that future recreation of events can be generated from this clustering.

4. Acknowledgements

Many thanks to my professors at Nova Southeastern University for providing challenging work, diverse research opportunities, and commitment to excellence in teaching. Additionally, I would like to thank my wife, Sharon, and my two kids for being patient and affording me the opportunity to pursue this endeavor.

References:

[1] Thayne Coffman, Seth Greenblatt, Sherry Marcus, Graph-based technologies for intelligence analysis, *Communications of the ACM, 47*(3), 2004, 45-47.

[2] Robert Popp, Thomas Armour, Ted Senator, and Kristen Numrych, Emerging technologies for homeland security, *Communications of the ACM, 47*(3), 2004, 36-43.

[3] John Yen, Emerging technologies for homeland security, *Communications of the ACM, 47*(3), 2004. 32-35.

[4] Jesus Mena, *Investigative Data Mining for Security and Criminal Detection* (Oxford, UK: Butterworth-Heinemann, 2003.)

[5] C. Lee and D. A. Landgrebe, Feature extraction based on decision boundaries, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(15), 1993. 338-400.

[6] M. Tim Jones, *AI Application Programming* (Hingham, MA: Charles River Media, 2003.)

[7] Bhavani Thuraisingham, *Web Data Mining and Applications in Business Intelligence and Counter-Terrorism* (Boca Raton, FL Auerback/CRC Publications, 2003.)

[8] Mehmed Kantardzic. *Data Mining: Concepts, Models, Methods, and Algorithms*. (Hoboken, NJ: John Wiley & Son, 2003.)

[9] G. Carpenter, S. Grossberg, A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition, Machine. *Computer Vision, Graphics and Image Processing. 37*(1), 1987. 54-115.

[10] Marc Sageman, *Understanding Terror Networks* (Philadelphia, PA: University of Pennsylvania Press, 2004.)

[11] Beale Hagan Demuth, Mark Beale, Martin T. Hagan, Neural Network Design (Boston, MA PWS Publishing Company, 1996.)