Technical Evaluation Report

G.Wyman
UNITED KINGDOM

glyn.wyman@gmail.com

ABSTRACT

This paper presents the findings of the Technical Evaluator relating to the workshop on Big Data Challenges: Situation Awareness and Decision Support held in Budapest 15-16 October 2019. A summary of each invited presentation is included as an annex. Data capture remains a critical element for presenting succinct information, AI can assist in this aim but the users must be aware of the limitations. Confidence in the data presented and the conditions of use should be displayed. Visual Analytics need further research to supplement the existing tool set, evidence of progress in visualisation techniques was noted. It concludes that the vast amount of available data is already exploited for commercial and state purposes and that users must be educated to be aware of the impact that third parties can exert.

INTRODUCTION

The proliferation of data continues apace and the distribution is feasible over the internet for commercial and other enterprise. Data from a wide spectrum of sources is becoming readily available and requires analysis to present the necessary and sufficient information to the operators. Users need relevant information in a timely fashion but defining what is relevant can be difficult and will vary with the background, environmental context and psychological state of that person. AI development is now sufficiently mature to achieve some of the desires which coupled with advances in Visual Analytics gives enhanced understanding of the information displayed. A whole raft of AI tools have evolved and continue to be developed, critical to the models is the training data which needs to be extensive. In the military field some data can be sparse or highly sensitive, in this instance the training data could be supplemented with synthetic data derived from other models or through data farming. The ethical and legal implications need extensive thought but should not be insurmountable. The field of Visual Analytics and AI has attracted academic and commercial interest which encourages innovation in, for example, reinforced learning tools. The desire is to retain information superiority over an adversary.

The Workshop was called and invited scientists in the area to combine their knowledge with experts involved in current NATO projects. It is noted that the ‘Call for Papers’ restricted the attendance to 50 but it is unknown how the attendees were selected.

THEME

The theme of this meeting is how the availability of big data may assist the military in understanding the situations they will encounter, and how this may support them in making better decisions. The challenges for development and opportunities for the exploitation of Big Data are likely to vary from domain to domain the following were identified for consideration:

Applied vehicle technology	Human cognition and perception	Situation awareness and Decision Support
Artificial intelligence	Information warfare	System analysis and studies
Challenges in Big Data exploration / exploitation	Knowledge engineering	System concepts and integration
Cyber defence and security	Medicine	Visual Analytics and visualization
Evaluation frameworks, measures and metrics	Mission planning	War gaming
Genomics	Modelling and simulation	
Human Factors	Sensors and electronic technology	

PERCEIVED MILITARY ISSUES

Information superiority is paramount to achieve dominance, regrettably the military appear to lag the development and exploitation of big data observed in other fields. Information warfare has already been observed. Decisions are critical at all levels of the Military Command where succinct, timely information is essential presented with a high level of confidence. The significant increase in raw data obtained from,e.g. ad hoc sensors and social media will only increase. Further enlargement is envisioned as a result of the proliferation of small low earth orbiting satellites. Filtering of these massive amounts of data is essential to avoid overload of an individual viewing the information, demanding a level of data analysis with the ability to identify false sets of information whether explicit or subliminal. Trust in the systems is critical with any derived actions displayed to the users, ideally coupled to the reasoning for taking that approach. Visual Analytics is already employed with a significant array of commercial tools. The use of extant modules to ease the development of new systems can incur penalties introducing uncertainty and thus reducing trust; this can be mitigated through transparency since formal analysis is impractical.

The increased use of Artificial Intelligence (AI) to identify the salient aspects of a situation exposes issues of verification impacting on the confidence. AI tools are evolving with a further level of AI employed to establish the model structure. Research and Development in commercial applications have the potential to be exploited but care needs to be exercised with respect to the security implications. Third parties can readily influence the outcome both in terms of the raw data but also the data employed during the training process, detection of these activities is non trivial. Third parties can also have an impact on communities through social media which will have an impact on any interface with that group. The aspect of using social media in a pro active manner was raised but not pursued.

Information specific to the operator taking cognisance of environment, background, ability and psychological disposition requires development. Collection of data using autonomous systems to supplement identified deficiencies need to be integrated. A pragmatic position on risk assessment must be understood by the decision makers made easier if recommended action made by intelligent agents are displayed.

EVALUATION

Structure

A Call for Papers was issued to solicit contributions and participation from scientists who have been engaged in this field. From the submissions, fifteen were selected with an additional two keynote papers designed to give a background with a general assessment of the research in hand. The technical papers were divided into four sessions: 1) AI and Machine Learning, 2) Human Factors, 3) Information Warfare and Social Media, and 4) Visual Analytics and Visualisation. All papers were relevant and benefited from the selection process, the quality was generally good. The papers and presentations were, in some instances, endorsed by demonstrations of the tools during the breaks. The inclined reader can access the full papers on the CSO website with a synopsis of each provided as an annex to this report. Leverage from other working groups was evident both from within the IST and other STO groups. It is strongly recommended that groups or individuals submit technical activity proposals (TAP) to establish research areas which can be supported by the IST or other STO Panels.

Overview

To successfully engage in any conflict information superiority is critical. Data relating to situational awareness proliferates and is compounded by the improvements in resolution and quantity of sensors. The internet of things (IoT) allows searches and is exploited in the commercial field. Social media also plays a significant role by targeting perceived weaknesses in a population or dividing opinions within a coalition. The so called Digital Influence Machine (DIM) allows the use of information as an effective weapon. DIM can search data associated with individuals and target their activities and derived views. This then allows propaganda to be focused, subtly modifying their views and changing their actions. Evidence was shown in which third parties are exploiting social media to this effect. Not all these aspects need be negative and the military could mobilise support; an example was cited in which the Boston Police force defused a volatile situation during the Boston Marathon by using social media. The use of Deep Fake Videos was raised as a means to bias opinion, it is reported that this media can have a strong influence. Lax regulations over the use of the internet allow fake news to proliferate which is no longer countered by reputable journalists; the issue to detect and remove false information is non trivial.

AI and Machine Learning

Trust was an issue raised extensively during the workshop and a strong recommendation to display a measure of confidence was offered. The size of the supporting tools and the nature of neural networks makes formal analysis impractical. The inherent nature of neural networks makes ascribing a confidence interval very difficult with a quantitative assessment at best. The degree of trust can be improved by knowing the provenance and having active trusted data organisations. If in specific instances the error bounds and distribution of the parameters is derived a figure can be computed but in most instances it is the result of a coupling of entities with unknown characteristics. AI and Deep Learning are trained in specific aspects and show promise but should not be overestimated, if applied within the design constraints they can assist operators and filter raw data rapidly. In principle training based machine intelligence can use generic architectures (several exemplar templates are available). Its functioning critically relies on the availability of data that is reliable and selected to be carefully in tune with the targeted objective. A strong recommendation was made to retain the human as part of the process.

Natural language processing has been advanced with appropriate dictionaries available, discussion relating to establishing an ontology for this aspect did not reach a conclusion but received support. Sentiment analysis where groups are identified with similar views is a potential enabler to direct propaganda but could be refined.

Data is the corner stone of this topic and the dissemination is crucial, the nature of military operations demands that some data will be classified and have a restricted distribution, however the wider the dissemination the better the models are trained. A limited set of training data can be mitigated by introducing simulated data but care must be exercised to avoid any bias which will skew the outcome. It is also feasible that third parties can poison the data which could modify the weightings within the model. The ability of the tools to identify trends was observed and the ability to extrapolate to make predictions. This is enhanced with visual analytics to display the results in a coherent fashion.

Evidence was provided to isolate subsets and outliers to enable the analysts to focus on specific effects, the concept of coupling notionally independent parameters to identify causality remains heuristic. If causality could be identified a solution space rapidly opens, false correlations remain an issue where similar characteristics are incorrectly correlated.

Human Factors

The aim is simply expressed to present the available raw data in a form readily assimilated, underlying this requirement is a whole raft of processes. AI is one element where assistance is available, other areas include conversational analysis which contains its own analysis history relevant to the topic. The display selected will be: a) user specific; adjusted to reflect ability, experience, and psychological state b) context specific; adjusted to matched the environment and c) transparent. Software can assist in this by applying smart filters, applying fast querying, and identifying gaps. The importance of trust is again significant where not only a confidence indication is present but the reasoning behind the display. It is important to understand the demands of the user and reflect their requirements; some users will have different priorities.

Diversity should be considered to enhance confidence not only from the perspective of different views but ideally to employ different methods to derive the display. It is widely acknowledged that AI has considerable potential for Defence. However, its current state of progress, with its operational requirements, anticipated benefits and risks, requires that its users exercise a high level of critical judgement and awareness if it to be used as a component in critical decision making. Decision makers can only assimilate a small quantity of data so that any displayed must be relevant and succinct.

Information Warfare and Social Media

Social media is experiencing a rapid expansion and is accessible in a large number of forms, a representative list is available in the papers but the list is not static and subject to evolution. The use of social media is complex and ranges from an individual informing friends to states attempting to influence targeted sections of the public. Evidence of use for nefarious purposes is readily available and should be regarded as a potential battle ground. NATO will need to question the actions it needs to take with a recommendation to consider a proactive approach. This could entail funding of bots to promoting trusted users. It was suggested that prominent scientists should be active to improve the public perception of the organisation. It should also be identifying clusters with particular views and accumulating information on detected bots to hold a library to identify future instances. Evidence was demonstrated of a current capability. The internet as a conduit is not only capable of distributing information but is a rich source to accumulate situational awareness where interested observers can identify an adversary preparing an event.

Visual Analytics and Visualisation

A wide variety of visual analytic tools were discussed and a selected set demonstrated. Decision makers have an inherent limitation on the information which can be absorbed. An attempt was made to quantify the figure during the breakout discussions with a figure of tens of kilohertz indicated. This emphasises the need to focus on the salient points. Further research is required to establish the degree of visual occlusion required to extract information created by the visualisation methods.

Adaptive screens were advocated allowing the users to drill down for specific detail or use their initiative to interrogate aspects observed in the initial view. Enhanced relational data structures were advocated and demonstrated to reduce the reaction interval to display the association. Humans are adapt at extracting patterns from displays particularly if they have practice in that scenario.

Facilities

The room was well appointed with the delegates sat in rows with a limited number of tables giving all delegates full visibility of the screen. WiFi access was provided giving access to the cso website and other internet facilities, however the nature of the WiFi provided by the Hotel required security considerations. Coffee was served during the breaks with ample opportunity to engage with the presenters to elaborate on their research and to seek points of clarification. I observed very productive discussions during the breaks.

CONCLUSIONS

The Workshop achieved its objectives in informing the invited subject experts of the issues affecting NATO and provided an interface with scientists working for The Organisation. The knowledge associated with the discipline was advanced with a number of aspects identified which require further research. Raw data is expanding at a rapid rate dictating that AI assistance is required for decision makers to comprehend the information. Humans in the loop remains essential but smart filters and visual analytics play a significant role in assisting the users. A raft of tools have been presented and demonstrated with the recommendation that more research is undertaken to allow tools to evolve.

Trust of the information displayed is paramount but recognised as difficult to achieve. Defining the source and provenance would assist as would only accepting data from a known repository, but this may be too restrictive. Identifying corrupted data particularly for training neural nets remains critical.

Social media coupled with the internet of things permits rapid access and dissemination of information to a wide audience. It also permits adversaries to influence the opinion of communities by giving fake news to a wide audience with bots relaying the data as if it were a true actor. Another tactic is to drive a wedge through the community to create division. NATO needs to consider a proactive stance.

I commend the co-chairs and their technical committee for a successful outcome.

ANNEX

Papers/Presentations

The inclined reader is directed to the CSO website where a file of the papers and power point slides are available, the following are extracts of the salient points, with comments.

Keynote 1: Big Data Challenges-Like War: M. Wunder The keynote speaker provided a sound background to the issues concluding that information has already been observed as a weapon. Civilian companies have unintentionally provided the means to implement the war and NATO should be vigilant and be aware of the consequences. The STO has over 100 sponsored activities associated with this discipline with additional aspects still requiring research, in particular the ability to transfer data whilst maintaining trust. AI needs new training of personnel and availability of relevant data to establish the models

Paper 1: Extracting Value from NATO Data Sets through Machine Learning and Advanced Dat Analytics: I. Mestric A paper describing the activities within the NATO Agencies to enhance the visualisation of events during exercises and to enhance the products through lessons learned. The specific exercise is ‘Trident

Junction 2018' It concludes that Military specific ontology would be a considerable advantage but that NATO specific research is necessary.

Presentation 2: Mission-Oriented Research for AI & BD for Military Decision Making (AIBD4MDM) Theme: F. Desharnais The presentation was given by the chair of IST 173, a full report on the activities of the group will be published in March 2020. It addresses the need to ensure AI and Big Data can be both operational and cost effective. It advocates a mission orientated approach from a multidisciplinary perspective coupling stakeholders and scientists. The need for cross panel communication was emphasised.

Paper 3 Exploitation of Sensor Data Using Artificial Intelligence for Battlefield Sensemaking: V. Lavigne The paper describes the work undertaken in Canada to augment situational awareness. The emphasis was on labelling objects to use as training data. They identified challenges specific to military applications and encouraged common practice between nations to share best practice.

Paper 4: Ukraine Conflict in Media: Two Approaches to Narrative Analysis: T. Krilavicius The use of natural language extraction and the inference drawn from data obtained during the Ukrainian Crisis. The primary aim was to track and understand public opinion. The need to automate the process became evident. Visualisation of the occurrence of words were displayed from the various sources which conform with the anticipated outcome through reputation and political alliance of the sources.

Paper 5: Developing Transparent Conversational Agent Systems for Intelligence Analysis: S. Hepenstal The presenter has produced a prototype tool to assist analysts to retrieve information using natural interactions. The work modelled the actions of analysts derived by undertaking a cognitive task analysis of four experienced analysts. A practical demonstration of the tool should be available in the near future.

Paper 6: Determining the Effectiveness of Representing Intrusion Detection System Log Files with Visualisation Techniques: G. Thomas The paper is an attempt to optimise the display of data to identify a cyber attack. It is recognised that identifying the structure of an attack in near real time is hard. A comparison was made between a) scatter plot and b) parallel coordinates. The results show a slight preference for the use of scatter plots.

Paper 7: An Exploration of Maintaining Human Control in AI Enabled Systems and the Challenges of Achieving It: M. Boardman A paper presenting the findings of HFM-ET-178 looking at the increased use of AI in robotics and the legal, moral and ethical questions which arise. The recommendation that a level of human control is maintained was presented because of the complexity and the adaptability to the situation. The human will however need support tools to ensure meaningful control.

Keynote 2: What's next for Visual Analytics? K. Cook The second keynote speaker presented the flavour of the current position with a view to extrapolation. No major step functions were identified but evolutionary progress predicted. New approaches are envisioned by combining computational techniques with interactive visualisation whilst acknowledging the limitations of human resolution. Aspects of trust to enable a measure of calibration were addressed. No solution was offered but mitigated by the ability to drill into the provenance. Assistance to visualise development and exploration was advocated which would be an extension of the common aids associated with, for example, smart speakers. Visualisation coupled with new relational data libraries may be worth pursuing. The new generation of machine learning will allow humans to create smarter systems removing the mundane elements and aiming for a shared understanding between humans and machines.

Paper 8: Security Perspectives on Social Media Exploitation: B. Forrester The work reports on the activities within IST-177. The number of platforms in which individuals can access social media is extensive and continues to grow with very little regulation this paper provides a snapshot taken a short time ago. Three significant aspects were considered a) the content b) the meta-data and c) the links and networks formed.

Content analysis and natural language processing are both applied in this context to create situational awareness and undertake predictive analysis. The presentation gives evidence of significant advances in natural language processing and identifying sentiments expressed in the media.

Paper 9: Propagation Filters:Tracking Malign Foreign Interventions on Social Media in Real Time: B. Forrester Suspicion had been aroused as to whether foreign agents were influencing public opinion in Canada prior to an election. Models were applied to identify the agents and to isolate the bots to populate the library of known networks. Narratives propagated by the identified agents were similar to those detected in the US elections of 2016. The technique is to drive a wedge between groups to diversify opinion. NATO should ensure that we remain steadfast and promote unity.

Paper 10: People, Bots or Cyborgs? Analysis of a Hashtag War: R. Goolsby Information threats can arise from a combination of artificial and enhanced actors who manipulate community structures, content and code. The aim of the project presented is to create a common framework to analyse the data and assess the potential impact and prepare countermeasures. Attacks have been identified which manipulate perception, emotions, stance and opinion. The counter is to build social trust by engaging with topic groups. The framework 'BEND' has been postulated to build community resilience. Concern was expressed about a amygdala hijack forcing individuals from rational thought. Understanding the mechanism and making early recognition of an attack is crucial.

Paper 11: Towards Big Data in the Tactical Domain: F. Johnsen The paper addresses the benefits and challenges of the Internet of Things (IoT) which can contain physical and virtual objects. It is assumed that in general use the communications will be available on demand, for use in a military environment intermittent service should be anticipated. The rapid expansion of available entities is driven by miniaturisation of both sensors and processing equipment coupled with the power of cloud storage. The military should be able to take advantage of the leverage from commercial products but will require interoperability.

Paper 12: Assessment of IoT Data Ingest Reliability for Urban Environments: J. Michaelis This paper expands on paper11 to encompass the Internet of Battlefield Things (IoBT). Again using the leverage from the development of smart cities in the civilian field. A protocol was selected for use in an urban environment where communications could not be guaranteed with test results presented. Additional research is recommended to establish the necessary network bandwidth, the computing power demanded and the cognitive load on the soldiers.

Paper 13: Storytelling Exploratory Visual Analytics for Counter-Improvised Explosive Device Incidents: V. Lavigne The paper and presentation give the results of using a tool to display IED events in Ukraine. The tool is web based and is readily available. It provides four viewing perspectives: geospatial, data set, incident type and text analysis. The raw data is from a NATO source through work undertaken by SAS-117, a complementary research group to IST within the Science and Technology Organisation (STO).

Paper 14: A Preliminary Experimentation for Meteorological Data Visual Analytics: F. Pisano The paper focuses on meteorological data but could readily be applied to other fields. The chaotic nature of weather makes forecasting difficult but benefits from the extensive observation sites and the accumulation of historic data. The aim is to assist meteorologists predict extreme events to allow adequate preparation to protect life. The tool was demonstrated during the breaks and forms part of the evolutionary progress by refining a relational database. Different views can be displayed allowing the user to interact and focus on specific aspects.

Paper 15: Microbial Genomic Data Analysis for Infectious Diseases: M. Varga Advances in visual analytics has had a positive impact in a wide set of disciplines, this paper identifies the significant advances in diagnosing infectious diseases with the opportunity to assess pathogenesis. The impact on the military is

significant because of the heightened exposure both deployed and in barracks. A library of tools is available to the developers enabling the right information to be available at the right time in the right format.

Workshop 1 Data Challenges.

- Data key factor
- Need to establish infrastructure
- Machine Learning Eco system
- Build common data sets
- Establish repository for data.

Workshop 2 Human Factors.

- AI to assist Decision Making with Big Data
- Training and Education (Gaming)
- Transparent
- Recommend interdisciplinary workshop
- Trust calibration

Workshop 3 Social Media.

- Ethics of Social Media (Define unethical !!)
- Can NATO be proactive? (Define the constraints)
- Information wargame
- Educate regarding Social Media
- Detect inflation of ethnic tension

Workshop 4 Visual Analytics.

- Chaotic data, fractal characteristics (identify transform to display strange attractor)
- Display Risk/ Uncertainty
- Listen to users' requirements
- Display source