



BIG DATA AND THE FIGHT AGAINST EXTREMISM

Date: November 22, 2019

Disclaimer: This briefing note contains the encapsulation of views presented by the speaker and does not exclusively represent the views of the Canadian Association for Security and Intelligence Studies.

KEY EVENTS

On November 22, 2019, Dr. Fred Popowich presented “Big Data and the Fight Against Extremism” at the second annual West Coast Security Conference. The presentation was followed by a group panel for questions and answers. During such period, main discussion topics in the panel included slaughter-bots, the dark web and collaboration against increasing cyber threats.

NATURE OF DISCUSSION

Presentation

Dr. Popowich began by defining big data as the ‘Vs’: velocity, volume, variety, veracity, visualization, and value—big data can be considered as data too vast for human analysis alone, thus requiring the assistance of Artificial Intelligence (AI). By combining big data and AI with Natural Language Processing (NLP), analysts can use these tools to solve problems and identify new ones that occur quicker than ever before.

Question Period

The questions directed to Dr. Popowich during this section focused primarily on the biases that occur both in AI and in the data being analyzed.

BACKGROUND

Presentation

Dr. Popowich studies NLP, a branch of artificial intelligence which assists computers in understanding, interpreting, and manipulating human language;

NLP makes it possible for computers to read text, hear speech, measure sentiment, and determine which parts are important. In general terms, NLP breaks down language into smaller elemental pieces, tries to understand the relationship between the pieces, and explores how the pieces connect together to create meaning. Basic NLP tasks include tokenization and parsing, lemmatization/stemming, part-of-speech tagging, language detection, and identification of semantic relationships.

To further this research with NLP and how it can be merged with AI, Simon Fraser University's (SFU) Key's Big Data Initiative is a project directed by Dr. Popowich. Its purpose is three-fold: to fill a critical big data talent gap in the public and private sector, to make data-intensive approaches more accessible through customizable training and data literacy programs, and to collaborate with multidisciplinary experts to accelerate research breakthroughs and thus, drive innovation forward. This project works with a variety of academic and practitioner experts in SFU to understand the impact of big data and how it can be used both positively and negatively in communities around the world. One area of this project aims to research how big data combat fake news utilizing NLP to understand how language can be created to wrap false information around facts.

Dr. Popowich illustrated the significance of how language, which shapes the reality of humans, is ambiguous. Context is applied with human-to-human communication and context does not translate in human-to-technology communication, thus demonstrating the ambiguity.

To provide an example of how to best translate human-to-technology communication, the work of Bdour Alzeer was sampled. There was emphasis placed upon her research project, which utilizes NLP analysis to understand hate speech and kinetic action online. Dr. Popowich stressed how in most cases, analyzing hate speech data could have negative impacts on the analyst. Utilizing NLP analysis with technology allows for the analyst to be a step removed from the impact of hate speech and provides the analyst with extracted topics, features, entities, and sentiments of the data. Involving the analyst in the end stages allows for the ambiguity of context to still be considered and analyzed.

Question Period

In the question period, questions arose regarding how bias in AI programming can influence the results when analyzing big data. It was noted in response that

there is the possibility of minute quantities bias always existing in programming, as a result of the programmer owning their own biases. As a conclusion, it was derived that bias may always exist in the analysis of big data even with the assistance of AI and an analyst.

KEY POINTS OF DISCUSSION

Presentation

- Combining big data with AI allows for analysts to identify and solve problems of larger scales more efficiently.
- Context in human communication adds ambiguity that cannot be easily translated to AI for processing.
- NLP can assist in lessening the ambiguity when programmed into the AI.
- Using AI to process abusive and disturbing hate speech data lessens the impact it could have on an analyst.
- Application of NLP and machine learning techniques to provide insight into such complex data could possibly help to identify the targets of hate speech and to extract the topics, features, sentiment, and entities.

Question Period

- Bias will exist when examining big data with technology and human analysts.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

© (CASIS-VANCOUVER, 2020)

Published by the Journal of Intelligence, Conflict, and Warfare and Simon Fraser University

Available from: <https://jicw.org/>