Chapter 118
Detecting Linguistic Markers of Violent Extremism in Online Environments

Fredrik Johansson
Swedish Defence Research Agency (FOI), Sweden

Lisa Kaati
Uppsala University, Sweden

Magnus Sahlgren
Gavagai, Sweden

ABSTRACT

The ability to disseminate information instantaneously over vast geographical regions makes the Internet a key facilitator in the radicalisation process and preparations for terrorist attacks. This can be both an asset and a challenge for security agencies. One of the main challenges for security agencies is the sheer amount of information available on the Internet. It is impossible for human analysts to read through everything that is written online. In this chapter we will discuss the possibility of detecting violent extremism by identifying signs of warning behaviours in written text – what we call linguistic markers – using computers, or more specifically, natural language processing.

INTRODUCTION

In recent years, there have been many examples of various types of terrorist attacks taking place all over the world. There have also been several severe school shootings that resulted in many victims. When studying the reason behind why these attacks took place, Internet often has an important role to play. For example, the use of the Internet for terrorist recruitment and operations has increased significantly in recent years (Torok, 2013), not least due to an emergence of social media services such as Facebook, Twitter, Instagram, and YouTube.

There are several examples of terrorists and terrorist organisations that use or have used the Internet and social media in different ways. One example of a terrorist who used the Internet extensively is Jose Pimentel, who was arrested for planning attacks with home-made pipe bombs against police vehicles and postal facilities in the United States. Pimentel was very active on the Internet, where he maintained a website on Blogger and a YouTube channel containing radical works which connected him with like-minded individuals (Weimann, 2012). Another example of a terrorist that used the Internet is the Norwegian terrorist Anders Behring Breivik. He used the Internet to obtain the necessary knowledge on how to construct a large fertiliser bomb, and to express and discuss his critical view on Islam and socialism (Ravndal, 2013). The suicide bomber Taimour Abdulwahab al-Abdably that killed himself in the middle of Stockholm in 2010 is another example of a person that used the Internet for various reasons. Abdably was active on various forms of social media such as YouTube and Facebook and he also searched for a second wife on Islamic web pages.

Due to the nature of the Internet and social media, it is possible to communicate and express radical views and intentions as well as to connect to other persons with similar interests. This is also noted in Europol's annual terrorism situation and trend report of 2012, which states the following: "Online social media sites attract high numbers of users. Internet forums are an effective means to address targeted audiences, including supporters who have no off-line links to terrorist organisations" (Europol, 2012, p. 10).

The ability to disseminate information instantaneously over vast geographical regions makes the Internet a key facilitator in the radicalisation process and preparations for terror attacks. This can be both an asset and a challenge for intelligence and security agencies. While it is troublesome that radical and violent extremism content can be spread globally with very low costs, this fact also provides an opportunity for intelligence analysts and police to act preventively. By collecting, fusing and analysing 'weak signals' or 'digital traces' that are present on the Internet, there is a possibility to detect attackers before they strike.

One of the main challenges in doing such analysis is the sheer amount of information available on the Internet. For this reason, analysts need support from computerised tools to be able to perform analysis on a large scale. Although web data have become an important source of information for law enforcement agencies working to prevent terrorist attacks, it is impossible for human analysts to read everything that is written on the Internet.

As an example, a human analyst capable of speed-reading may be able to read up to 1,000 (or even more) words per minute (cf. the average adult reading speed is around 300 words per minute). If the analyst is able to read at the same speed consecutively for eight hours (which is perhaps not very likely), he/she will have read about 480,000 words, which given an average word length of six characters, equals 2.8 megabyte of data (cf. the average human would have read less than one megabyte of data in the same time period). It has been estimated that the Internet carries around 1.8 exabyte of data per day. Furthermore, approximately 80 percent of all available data is in unstructured form, and a large portion of this unstructured data is textual data. This means that it would require several hundred million incredibly focused speed-reading human analysts (and consequently more than a billion humans with normal reading capacity) to be able to read all the text data generated on the Internet in only one day. This example illustrates the need for computer support to be able to detect and analyse content that is of interest for law enforcement agencies if we do not know where to find the information of interest.

Related Content

Minimax Probability Machine: A New Tool for Modeling Seismic Liquefaction Data
Pijush Samui, Yldrm Huseyin Dalkiliç, Hariharan Rajaduraiand J. Jagan (2015). *Handbook of Research on Swarm Intelligence in Engineering (pp. 182-210).*
www.irma-international.org/chapter/minimax-probability-machine/131250

PSK Method for Solving Type-1 and Type-3 Fuzzy Transportation Problems
P. Senthil Kumar (2017). *Fuzzy Systems: Concepts, Methodologies, Tools, and Applications (pp. 367-392).*
www.irma-international.org/chapter/psk-method-for-solving-type-1-and-type-3-fuzzy-transportation-problems/178403

Intelligent Water Quality Monitoring System Based on Multi-Sensor Data Fusion Technology
Qiuxia Liu (2021). *International Journal of Ambient Computing and Intelligence (pp. 43-63).*
www.irma-international.org/article/intelligent-water-quality-monitoring-system-based-on-multi-sensor-data-fusion-technology/289625

A Hybrid Model for Rice Disease Diagnosis Using Entropy Based Neuro Genetic Algorithm
K. Lavanya, M.A. Saleem Duraiand N.Ch.S.N. Iyengar (2017). *Fuzzy Systems: Concepts, Methodologies, Tools, and Applications (pp. 273-291).*
www.irma-international.org/chapter/a-hybrid-model-for-rice-disease-diagnosis-using-entropy-based-neuro-genetic-algorithm/178398

Medical Diagnosis Based on Distance Measures Between Picture Fuzzy Sets
Palash Dutta (2018). *International Journal of Fuzzy System Applications (pp. 15-36).*
www.irma-international.org/article/medical-diagnosis-based-on-distance-measures-between-picture-fuzzy-sets/211984