COUNTERING TERRORISM ON SOCIAL MEDIA USING BIG DATA

Ali Alzahrani[1], Khalid Bashir Bajwa[2], Turki Alghamdi[3] and Hanaa Aldahawi[4]

[1,2,3]Department of Computer Science,
Islamic University of Madinah, Madinah, Saudi Arabia
[4]Department of Information Science,
King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Terrorism and violence are used by miscreant groups and individuals to disrupt the normal course of events. While not a new phenomenon, the information age offers new and innovative methods for spreading messages related to terrorism and expansion through recruitment on social media. These observations are alarming due to the broad reach and speed of propagation made available by social media. To ensure safety, harmony, and peace, it is important that the use of social media for terrorism is minimised. We discuss the various methods used by terrorists on social media to increase exposure and identify how the inherent structure of social media, the amount of data available, and language understanding pose challenges as well as opportunities for control efforts. We propose a strategy for restraining terrorist activities through data mining methods on big data created by social media in combination with natural language processing for language understanding and social network analysis for uncovering the underlying structure and association of terrorist groups and their activities

KEYWORDS

Big Data, Data Mining, Social Network Analysis, Natural Language Processing.

1. INTRODUCTION

There has been a war waging on between right and wrong since the dawn of time. The dictionary definition for terrorism is "the unlawful use of violence and intimidation, especially against civilians, in the pursuit of political aims," [1] which describes it as illegal and, hence, wrong by the prevailing norms. Since the commercial availability of the Internet in the late 1980s [2], users of computing devices are now connected affecting many aspects of society. While adding significant value, the Internet has opened loopholes and new forms of access for miscreants to exploit to perform their illegal efforts with greater ease. There is a dire need to curtail the use of emerging technologies for unlawful actions. New perspectives and methods have been developed with the study of counter-terrorism becoming a major research area. Scholars from social sciences, political sciences, and technological areas have joined efforts to better understand the phenomenon [3].

The Internet also gave rise to a new definition of social interaction through the extended use of social media. As social networks are less than two decades old, all earlier mediums of socialisation combined did not create such a wide audience with personalised experiences, largely due to Artificial Intelligence and Machine Learning methods. Services such as Facebook, Twitter,

David C. Wyld et al. (Eds) : AI, CSTY, NMOCT - 2018
pp. 35–42, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81404

YouTube, and Instagram have revolutionised the way people interact. These opportunities also opened doors for terrorist organisations to more efficiently recruit, grow, train, and communicate with followers, supporters, and donors. Social media is also an effective platform for spreading propaganda, ideological thoughts, and hate material [4]. Radicalisation that once required personal contact is now spread by social media through digital materials (audio, video, images, and messages). An alarming aspect is the access to local groups and societies from remote locations, which was not possible before. Social media has been driving and affecting events since its proliferation into the society, such as the international covert influencing of the 2016 U.S. presidential election [5].

Through this ability to influence, the power of social media has been used by terrorists to coordinate and execute attacks. The most effective use of social media for terrorism purposes was by Islamic State (ISIS) for spreading their ideology and propaganda materials. The distinguishing feature of their operation was the quality of the material and the ability to disseminate information through social channels [6]. Additionally, ISIS used high velocity content executed with great efficiency [7].

Terrorism-related information intended for the spread across social media targets a young audience who are within the prime recruiting age [8]. This information is also spread in multiple languages making it accessible to a broader audience [9]. Social media also enables terrorist organisations to engage directly with the public through the hijacking of the identities of credible figures. This method is performed, for example, by being the first to reply to a tweet from a notable person with a large number of followers. As the commenting thread expands, the illicit remarks gain more views. Confronting people and organisations who hold a credible reputation and enticing them into a discussion is another approach used by terrorist organisations to increase social mileage. A common strategy by insurgents is to spread their messages and drive communications through the use of disseminators located outside a war zone. In addition, algorithms developed by social services to connect people and provide suggestions based on personal preferences, history, and other information often favour terrorist organisations in spreading their messages [10]. These social algorithms differ in how they connect people. For instance, Facebook users see posts that are "personalised based on past clicks," including the 'Like' button, and on items' popularity among other users with similar preferences [11]. Google presents results based on the location and previous searches and clicks [11]. These algorithms become extremely accurate in projections over time, and miscreant groups can leverage these mechanisms for information propagation and dissemination.

Terrorist groups using social media's capabilities to disseminate information and recruit also makes them vulnerable. Social media operate through global platforms with open access, which leave digital signatures for law enforcement agencies to tracking them down. Successful operations to curb malicious activities and track those responsible have been performed with the help of open source information [10]. Analysing available information using intelligent tools and methods enables localisation and riddance of terrorists. By acting swiftly on this available information, it also becomes possible to stop incidents before they occur.

Social networks provide a graph of linkages of individual social accounts and the information contained in them. Graph analysis techniques such as cluster graph structure and graph vertex analysis [21] are used to associate people to groups and identify relationships. Terrorist organisations operate in such a way that that social networks can provide the first point of contact with potential candidates for recruitment. Once initial contact is established further communication is shifted to an encrypted channel where tracking becomes nearly impossible, which is further complicated by privacy and protection laws prevalent in various countries around the world. Thus, capitalising on this initial window of opportunity through the information available from open source social networks is of utmost importance.

In this section, we identified the role of social media in human society. We identified the vulnerabilities of social media and how it can be used by miscreant groups to gather traction, recruit followers, and spread their messages. We looked at how terrorist organisations prepare professional quality materials with high velocity and leverage learning algorithms. We identified that although social platforms offer advantages for terrorists in spreading their message, they also include vulnerabilities with traceable signatures. The mitigation goal is to identify and stop the spread of malicious content and contain it within the realm of social platforms.

The remainder of the paper is organized as follows. The next section discusses the inherent problems faced in addressing these issues. We then propose strategies that can be used to counter terrorist activities to achieve harmony and peace. This is followed with a conclusion and references.

2. THE INHERENT PROBLEM

In the information age, enormous amounts of data are created with exponential increases in recent years. With social media, huge volumes of data of different varieties are generated rapidly. Dealing with all this data is challenging as traditional methods of analysis such as guessing, constructing hypothesis and testing with data based experiments do not perform well [22] because of the sheer volume and variety of data, hence new methods uncovering the insights in data must be devised. Research in Big Data is considering new techniques [23-27] with the core challenge of how to process data to extract useful information. Computational resources seem to be second runners-up in this race with Big Data leading the way. Classic learning and intelligence methods also fail to perform well on large data set, and new "Deep Learning" techniques [23, 24] are being devised to take advantage of the available information.

Interpreting natural language comes easily to humans. However, the same is not true with computing machines, and processing natural language data is critical in uncovering terrorism-related information. This poses another challenge for the processing of Big Data produced by social media sources.

Social interactions are complex as are the relationships between entities in a social network. Complex graphs must be evaluated to extract useful information. Uncovering these relationships and making sense of them is critical to identifying terrorist groups, their recruitment strategies, and their information dissemination methods. Complex graph analysis techniques such as spectral clustering, information maps [28] must be employed for this.

Extracting meaningful information and curbing terrorist-related activities require dealing with Big Data, natural language processing and network analysis. In the next section, we propose strategies for dealing with these problems.

3. PROPOSED STRATEGY

The following strategies address the three problem domains identified with the aim of combating terrorism using the latest technology.

3.1. Data Mining on Reduced Data

Extracting information on terrorism and its allies from Big Data requires processing large amounts of data, which is sometimes not feasible due to limitations on computational resources and timing constraints. Innovative methods are required to reduce the data while preserving information content, such as the following data reduction methods.

- Dimension reduction techniques based on clustering, map-reduce implementations of existing dimension reduction methods, feature selection techniques, and fuzzy logic implementations. PCA, SVD, eigenvalue/eigenvector decompositions.

- Reduce the velocity of data streams before entering into storage (pre-processing). In a specific use case from [13], the proposed algorithms show that data reduction performs effectively, and the memory requirement is reduced from 3 TB to 300 GB of RAM.

- Data sampling [14] is useful when data sizes become too large to practically deal with the entire dataset simultaneously and has been used extensively in data mining applications. Sampling techniques include simple random sampling, stratified sampling, systematic random sampling, and cluster sampling.

- Network theory approaches are used to extract topological structures of unstructured data [15].

Using these methods, we propose to reduce the data while preserving the information content. Only the techniques that preserve the underlying semantics of the data will be helpful as we will be applying data mining techniques on the reduced dataset.

Data mining is a powerful approach for discovering valuable information by analysing data from different dimensions, categorising it, and summarising the data relationships identified in the database. Subsequently, decisions can be made or improved based on this information. In data mining solutions, algorithms can be used independently, or more than one can be applied to achieve the desired results. It can be employed using some algorithms to explore data, while other algorithms are used to extract specific data to find a specific outcome. For example, clustering algorithms, which recognise patterns in data, can be used to group data into different n-groups. Data contained in each group are considered reasonably consistent so that a decision model can be created based on the results. Multiple algorithms can be applied within one solution to perform separate tasks.

Stored data is divided into predetermined groups. The classification algorithm uses a training data set, where each record is predefined in a different class for building a learning model, and a testing dataset, which classifies and labels every record with an unknown class. Classification is sometimes called supervised learning because class labels are known in the training dataset.
Clustering, or unsupervised classification, is a method that separates data into groups of members that belong together based on some characteristic. Class labels in clustering are initially unknown, so a clustering algorithm discovers acceptable classes and assign each item to the corresponding group. Typically, clustering provides a broad view to the user of what is happening in the database. Clustering does not require prior knowledge of the groups and their members, which is useful for separating terrorist-related data from other data.

Association rules are used to discover elements that frequently co-occur in a dataset that contains multiple independent selections. The association rules approach includes two phases. The first is support, where frequent item sets are identified. The second is confidence, where conditional probabilities are identified in transactions in which items continually appear together. In the context of this article, association rules are particularly useful in extracting relationships between terrorist groups and their recruits.

Sequential patterns are anticipated from the data by mining it for patterns that appear frequently. Known patterns learned from a terrorist dataset as the one maintained by the Global Terrorism Database (GTD) [29], extracted using natural language processing methods can be useful to identify similar patterns from the data to uncover additional terrorist-related content.

Regression algorithms are useful in predicting future values of data by analysing the behaviour of data over a period. These techniques can be useful for predicting upcoming terrorist-related activities that may occur.

Once data has been reduced to a manageable size, the data mining techniques of classification, clustering, association, sequential pattern discovery, and regression can be used more effectively for identifying terrorism and extracting recruitment-related information and strategies of different groups of terrorists. With a reduced data set it is relatively easy to identify terrorism and the hidden relationships among terrorists that may result in terrorist acts. Data mining on the reduced set with clustering to gather specific data into groups can be helpful in segregating terrorist groups. Grouping terrorist data into homogeneous classes or clusters can provide a comprehensive understanding of terrorist behaviours, while predicting the likelihood of terrorist activities in a reasonable amount of time.

The process of using data mining involves the following steps:

- Establish domain (terrorism and allied activities) understanding with relevant prior knowledge and identification of end-user goals.
- Build a dataset using natural language methods, as described in the next section, by using known keywords and entities related to terrorism and terrorists.
- Pre-process data for reduction and cleanliness.
- Identify useful features in the data.
- Appropriately apply the data mining strategies of classification, clustering, association, sequential pattern discovery, and regression to understand and predict terrorist activities.
- Use data mining algorithms to search for patterns in the dataset to identify additional malicious activities.
- Extract interesting patterns that distinguish terrorist-related activities from the rest of the data.
- Document and report the observations.

3.2. Natural Language Processing

Making sense of natural language is fundamental in identifying malicious information. Natural language processing is the automatic analysis and representation of human language, which enables computers to perform a wide range of natural language-related tasks [16]. With the availability of large amounts of data, deep learning methods can be used that employ multiple processing layers to learn hierarchical representations of data. A variety of model designs and methods have blossomed in the context of natural language processing [17].

Leveraging the available language data, we can employ deep methods to extract terrorism-related information. This approach involves making semantic sense of the text and identifying underlying patterns to characterise terrorist-related information. Crawling social media sites and using natural language processing methods to analyse the content against keywords using state-of-the-art matching methods, such as distance-based matching. Identifying entities with connections to both organisations and names that have been banned and declared terrorist can be accomplished

using language processing techniques. Natural language processing tools can be used to scan the web for unwanted material and report for further analysis and processing. These methods must perform above a certain threshold for the system to work appropriately and result in as few false alarms as possible.

Natural language processing methods based on keywords and known entities are also useful for preparing and refining a continuously evolving dataset of terrorist groups, organizations, and linked people. This dataset is a prime resource for further learning and processing using Big Data analysis methods reviewed above as well social network analysis methods discussed in the following.

3.3. Social Network Analysis

With a dataset in place and a continuous stream of available data from social media, data mining methods can extract useful information. However, this processing can be further improved using social network analysis techniques to unravel any underlying structures, relationships, and associations. Social network analysis is a collection of techniques that support statistical investigations on the patterns of communication between groups. Social scientists use these to analyse connected groups, and they form a basis of techniques for situational awareness and decision making in law enforcement applications [18].

A linkage map of terrorist organisations can be created using social network analysis [12] from which a frequency of co-occurrence of names of organisations can be used as a basis for inferring the intensity of the links. Concepts from graph theory play a pivotal role in network analysis, such as how an adjacency matrix will reflect the closeness of organisations. In addition, graph-theoretic concepts of centrality and between-ness provide further insight into the operation and structure of terrorist organisations. Using nodal analysis in social networks, a terrorist group can be rendered impotent by identifying and targeting their nodal or key points.

Social network analysis assesses the examined social aspects based on structures, which incorporate group members represented by nodes and their interconnections. As opposed to other quantitative strategies that centre around the portrayal and total investigation of the qualities of the actors who make up the exploration populace, social network analysis expects that to understand the social phenomenon, it is helpful to guide out and break down the arrangement of ties among the actors and the manners by which these social patterns shape actions. Thus, this approach can provide information on the decision making, group dynamics, and the outcomes of collective actions. The methodology for studying violent groups is broken down in the following, as suggested by [3]:

- Mapping the group with characteristics of parallel ties, symmetric/asymmetric, negative/positive, and the quality of ties, including measurement of time spent together, recurrence of communication, and size of associations.

- Division of power within a group, including progressive gatherings, actors having greater number of ties, thick structures or a heap of associated subgroups situated in vital areas. Proportion of status or centrality measures, idea of impact as an element of actor's significance.

- Structure and subgroups, including levels of cohesion and degree hierarchies, balance between efficiency with either a low number of repetitive connections or a high level of group centrality.

- Robustness and survivability with high density and a large number of redundant ties.

This framework leads to a deep analysis of terrorist groups and their modus operandi. Implementation of this is made possible using tools such as NetworkX [19] and SNAP [20].

4. CONCLUSION

In this paper, we presented the methods used by terrorists to spread their messages using social media. It is understood that containing terrorism-related material on social media is critical. We analysed the associated problems and proposed strategies towards a solution for containment of terrorism-related activities. The proposed strategy includes the use of natural language processing to build and expand a dataset by looking for terrorist related data on social networks, the reduction of the data using data sampling techniques, the use of data mining methods on reduced data to identify the patterns and extract useful information and the use of social network analysis to uncover the associations and relationship between individuals and terrorist groups, their structure and their modes of operations.

REFERENCES

[1] Terrorism definition, Available at https://en.oxforddictionaries.com/definition/terrorism

[2] History of Internet, Available at https://en.wikipedia.org/wiki/History_of_the_Internet

[3] P. Arie and P. Ami, "Social Network Analysis in the Study of Terrorism and Political Violence" Southern Illinois University Carbondale, OpenSIUC Working Papers 2010

[4] Md. S. Hossain, "Social Media and Terrorism: Threats and Challenges to the Modern Era" South Asian Survey, Vol 22 (2), pp. 136-155, 2018

[5] B. Bender, "SOCIAL MEDIA AND POLITICS: THE 2016 US PRESIDENTIAL ELECTION", BrandBa, 2017

[6] Alexander Meleagrou-Hitchens, et al., The Travelers: American Jihadists in Syria and Iraq, George Washington University Program on Extremism, February 2018, 33. Available at: https://extremism.gwu.edu/events/travelers-american-jihadists-syria-and-iraq

[7] Charlie Winter, "Fishing and ultraviolence: So-called Islamic State is known for its brutality. But it's also hooking people in far subtler ways," BBC, August 1, 2015. Available at: http://www.bbc.co.uk/news/resources/idt-88492697-b674-4c69-8426-3edd17b7daed

[8] Aris Roussinos, "Jihad Selfies: These British Extremists in Syria Love Social Media," Vice, December 5, 2013. Available at: https://www.vice.com/en_us/article/gq8g5b/syrian-jihadist-selfies-tell-us-a-lot-about-their-war.

[9] A. Alexander, Digital Decay: Tracing Change Over Time Among English-Language Islamic State Sympathizers on Twitter, George Washington University Program on Extremism, October 2017

[10] G. Ratnam, B. Misztal, S. Hughes, J. Geltzer, R. L. Strayer, "Digital Counterterrorism: Fighting Jihadists Online", Report Bipartisan Policy Center, 2018

[11] M. El-Bermawy, "Your Filter Bubble is Destroying Democracy," Wired, November 18, 2016. Available at: https://www.wired.com/2016/11/filter-bubble-destroying-democracy/.

[12] B. Aparna. "Social network analysis of terrorist organizations in India", Institute for Defence Studies and Analysis 2005.

[13] H. Bronnimann, B. Chen, M. Dash, P.J. Haas, and P. Scheuermann, "Efficient Data Reduction with EASE," Proc. ACM SIGKDD, 2003.

[14] Weinstein M et al (2013) Analyzing big data with dynamic quantum clustering. arXiv preprint arXiv:1310.2700.

[15] M. Trovati "Reduced topologically real-world networks: a big-data approach" Int J Distrib Syst Technol (IJDST) 6(2):13–27

[16] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," IEEE Computational Intelligence Magazine, vol. 9, no. 2, pp. 48–57, 2014.

[17] T. Young, D. Hazarika, S. Poria,& E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]". IEEE Computational Intelligence Magazine, 13, 55-75, 2018

[18] P. Svenson, P. Svensson, H. Tullberg, F. Ledningssystem, F. Ledningssystem, F. Ledningssystem, "Social Network Analysis And Information Fusion For AntiTerrorism", In Proc. CIMI, 2006

[19] H. Aric, S. Pieter, & S Chult, Daniel. "Exploring network structure, dynamics, and function using networkx". United States, 2008.

[20] J. Leskovec and R. Sosic, "SNAP: A General Purpose Network Analysis and Graph Mining Library", CoRR, 2016

[21] Q. D. Truong, T. Dkaki, Q. B. Truong, "Graph Methods for Social Network Analysis", 2nd EAI International Conference on Nature of Computation and Communication – ICTCC, pp-276-286, 2016

[22] R. Kitchin, "Big Data, new epistemologies and paradigm shifts", Big Data & Society, Sage Publications, April, 2014.

[23] J. Cała, P. Missier, "Selective and Recurring Re-computation of Big Data Analytics Tasks: Insights from a Genomics Case Study", Big Data Research, Vol 13, pp 76-94, 2018

[24] H. Estiri, B. A. Omran, S. N. Murphy, "kluster: An Efficient Scalable Procedure for Approximating the Number of Clusters in Unsupervised Learning", Big Data Research, Vol 13, pp 38-51, 2018

[25] D. Lakshmipadmaja, B. Vishnuvardhan, "Classification Performance Improvement Using Random Subset Feature Selection Algorithm for Data Mining", Big Data Research, Vol 12, pp 1-12, 2018

[26] K., Savita. "Impact of big data and social media on society". Global Journal for reseach Analysis. Vol 5, pp 437-438, 2016

[27] C. Debas, "Big data analytics for exploratory social network analysis", International Journal of Information Technology and Management, Vol 16(4), 2017

[28] M. William, C. Campbell, K. Dagli, and C. J. Weinstein, "Social Network Analysis with Content and Graphs", Lincoln Laboratory Journal, Vol 20(1), 2013

[29] National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2018). Global Terrorism Database [Data file]. Retrieved from https://www.start.umd.edu/gtd