



WEB DATA MINING FOR DETECTION OF TERRORISM ACTIVITY

Nilesh T. Gole¹, Tushar Sangole², Nilesh Bissa³, Pallavi Vyas⁴

¹ V.I.T, Nagpur, ² M.K.C.T, Akola, ³ V.I.T, Nagpur, ⁴ V.I.T, Nagpur,

¹tusharsangole@gmail.com, ²sheyanilu@gmail.com

ABSTRACT:

Technology is a foundation of modern society, it governs its dynamics, and therefore, it is normal that also the terrorism will benefit from it. We are living a very difficult period, the terrorism is threatening the society and cultures on a global scale, and differently from the past, it has reached a global dimension. We are facing a global alert. Intelligence and law enforcement agencies worldwide are warning about possible terrorist attacks in the major cities. Parigi, Istanbul, Jakarta, Kashmir, Iraq are last locations in order of time where the terrorist threat has demonstrated its strength, where terrorists have hit the defenseless population with a military strategy. The modern terrorism is instantaneous and unpredictable, a global threat that hit its targets, but that at same time hits a wide audience thanks to the technology, it is bringing the terror into our homes. The cyberspace is an environment without boundaries, a privileged place where terrorists find resources, make propaganda activities and from which it is possible to launch the attacks against enemies everywhere in the world. Social media is an essential element of modern terrorism; these powerful platforms allow terrorists to communicate, to make propaganda and recruit new sympathizers. Every terrorist attack is similar to a dramatic representation, in which the Internet is the stage.

INTRODUCTION

Terrorism has grown its roots quite deep in certain parts of the world. With increasing terrorist activities it has become important to curb terrorism and stop its spread before a certain time. So as identified internet is a major source of spreading terrorism through speeches and videos. Terrorist organizations use internet to brain wash individuals and also promote terrorist activities through provocative web pages that inspire helpless people to join terrorist organizations. So here we propose an efficient web data mining system to detect such web properties and flag them automatically for human review. Data mining is a technique used to mine out patterns of useful data from large data sets and make the most use of obtained results. Data mining as well as web mining are used together at times for efficient system development. Web mining also consists of text mining methodologies that allow us to scan and extract useful content from unstructured data. Text mining allows us to detect patterns, keywords and relevant information in unstructured texts. Both Web mining and data mining systems are widely used for mining from text. Data mining algorithms are efficient at manipulating organized data sets, while web mining algorithms are widely used to scan and mine from unorganized and unstructured web pages and text data available on the internet. Websites created in various platforms have different data structures and are difficult to read for a single algorithm. Since it is not feasible to build a different

algorithm to suit various web technology we need to use efficient web mining algorithms to mine this huge amount of web data. Web pages are made up of HTML (Hyper text markup language) In various arrangements and have images, videos etc intermixed on a single web page. So we here propose to use smartly designed web mining algorithms to mine textual information on web pages and detect their relevancy to terrorism. In this way we may judge web pages and check if they may be promoting terrorism. This system proves useful in anti terrorism sectors and even search engines to classify web pages into the category.

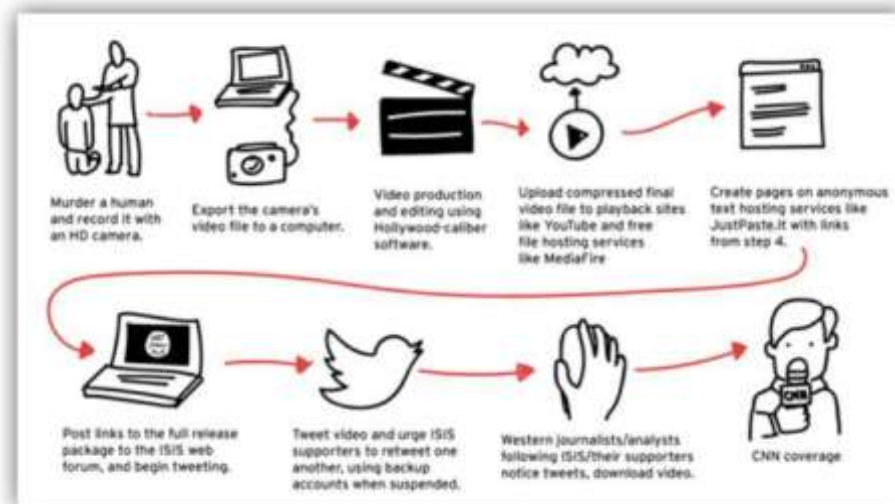


Figure 1 – Chain of Terrorist propaganda

1.WEB DATA MINING:

The term Web Data Mining is a technique used to crawl through various web resources to collect required information, which enables an individual or a company to promote business, understanding marketing dynamics, new promotions floating on the Internet, etc. There is a growing trend among companies, organizations and individuals alike to gather information through web data mining to utilize that information in their best interest. Data Mining is done through various types of data mining software. These can be simple data mining software or highly specific for detailed and extensive tasks that will be sifting through more information to pick out finer bits of information. For example, if a company is looking for information on doctors including their emails, fax, telephone, location, etc., this information can be mined through one of these data mining software programs. This information collection through data mining has allowed companies to make thousands and thousands of dollars in revenues by being able to better use the internet to gain business intelligence that helps companies make vital business decisions.

Challenges in Web Mining

The web poses great challenges for resource and knowledge discovery based on the following observations –

The web is too huge – The size of the web is very huge and rapidly increasing. This seems that the web is too huge for data warehousing and data mining.

1. Complexity of Web pages – The web pages do not have unifying structure. They are very complex as compared to traditional text document. There are huge amount of documents in digital library of web. These libraries are not arranged according to any particular sorted order.
2. Web is dynamic information source – The information on the web is rapidly updated. The data such as news, stock markets, weather, sports, shopping, etc., are regularly updated.
3. Diversity of user communities – The user community on the web is rapidly expanding. These users have different backgrounds, interests, and usage purposes. There are more than 100 million workstations that are connected to the Internet and still rapidly increasing.
4. Relevancy of Information – It is considered that a particular person is generally interested in only small portion of the web, while the rest of the portion of the web contains the information that is not relevant to the user and may swamp desired results

Advantages

1. This system will help to reduce terrorism spread around the world.
2. This system proves helpful for anti terrorism departments/agencies.
3. This system helps agencies to detect suspicious web pages and track them from their sources.
4. This system alerts authorities to block those web pages on time.



Disadvantages

1. If terrorists communicate in codewords not fed in system, the system will not be so efficient.

2. Intrusion Detection System:

With the increasing popularity of the wireless network, the security issue for mobile users could be even more serious than we expect. In this paper, we discuss several aspects of network security. Then we focus on the intrusion-detection and cloning mobile phones problems, which are currently attracting considerable interest from both research communities and commercial companies. An (IDS) constantly monitors actions in a certain environment and decides whether they are part of a possible hostile attack or a legitimate use of the environment (Debar et al.1999). The environment may be a computer, several computers connected in a network or the network itself. The IDS analyzes various kinds of information about actions emanating from the environment and evaluates the probability that they are symptoms of intrusions. Such information includes, for example, configuration information about the current state of the system, audit information describing the events that occur in the system (e.g., event log in windows XP), or network traffic.

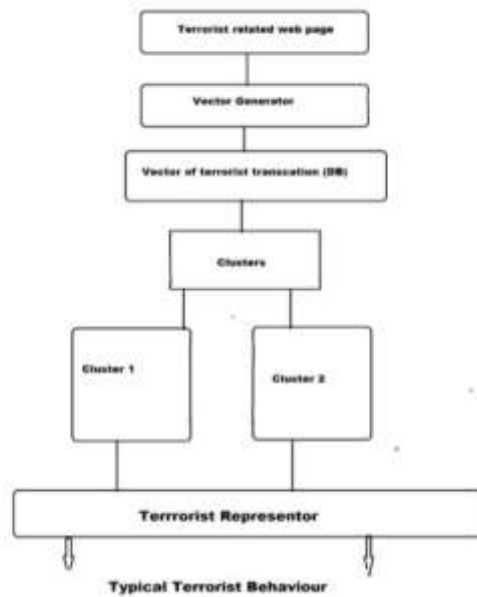


Fig: Detection Terrorist Activity

3. Clustering;

Clustering is the process of making a group of abstract objects into classes of similar objects. A cluster of data objects can be treated as one group. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups. The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups. Clustering also helps in identification of areas of similar land use in an earth observation database. It also helps in the identification of groups of houses in a city according to house type, value, and geographic location. Clustering also helps in classifying documents on the web for information discovery. Clustering is also used in outlier detection applications such as detection of credit card fraud. As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

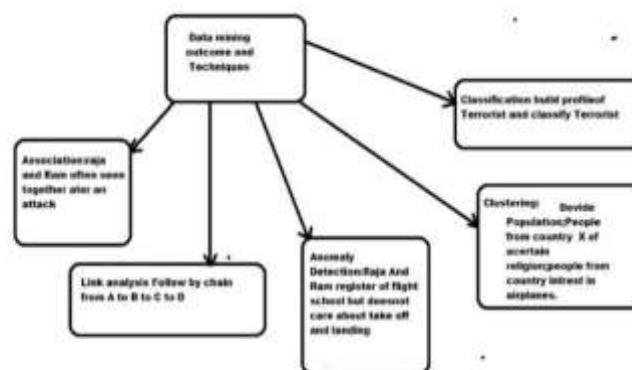


Fig: Data mining Techniques for counter Terrorism

4. CONTENT- BASED DETECTION OF TERROR-RELATED ACTIVITY

This study suggests a new type of knowledge-based detection methodology that uses the content of Web pages browsed by terrorists and their supporters as an input to a detection usually viewed by terrorists and their supporters can be used as training data for a learning surveillance is identified as a 'user's computer' having a unique IP address rather than by his or her name. In the case of a real-time alarm, he detected IP can be used to locate the computer and hopefully the suspected terrorist who may still be logged on to the same computer.

Conclusion:

In this paper innovative knowledge based methodology for terrorist activity detection on the web is presented. The above paper suggest that the method or a technique useful for detecting terrorist and their supporters using a internet access to view terror-related content a a series of evasive website

Reference:

- Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: A Overview and Case Studies", AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003, available at: <http://ai.bpa.arizona.edu/>
- Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: A General Framework and Some Examples", IEEE Computer Society April 2004.
- C McCue, "Using Data Mining to Predict and Prevent Violent Crimes", available at: <http://www.spss.com/dir/video/richmond.htm?source=dmpage&zone=rtstdebar>
- Whitepaper, "Oracle's Integration Hub For Justice And Public Safety", Oracle Corp. 2004, available at: http://www.oracle.com/industries/government/IntegrationHub_Justice.pdf