UNDERSTANDING TERRORIST BEHAVIOR WHEN HOSTAGES ARE TAKEN

Jayden Carr

Phil Ramontal

Damian Echevest

St. Thomas University

CIS 543*-* Pragraming for Data Analytics

*Abstract*—The purpose of this paper is to help understand and predict terrorist behavior . The ability to cluster events, individuals and regions enables Data Scientist to try to enter the mind of a terrorist and determine reason for an attack and why and where future possible attacks may take place. Unfortunately, no single method or person can do this work, nor can a single algorithm figure it out. Multiple Algorithm’s including Linear, Naïve Bayes, classification and Decisions Tress’s must be implemented to provide some level of comparison and accuracy. The ability to deliver some level of accuracy provide decision makers the ability to make informed decisions. Although accuracy is paramount an acceptable level must be determined for each problem so decision makers can have a full picture to work with. Higher risk issues will require a higher level of accuracy and certainty and concrete explanations to back up the analysis being provided.

# Introduction

Terrorism can be defined as an attack on the rights of humans life, property, and freedom (Including mental freedom) The problem of terrorism goes back hundreds of years. The goal of terrorist is usually to kill, strike fear and control citizens of the world. Terrorist carry out these acts of violence for various reason. In recent years terrorist have attacked in innocent civilians because of religious beliefs, financial reasons, social reasons and political reasons. Understanding why and having the ability to understand the individuals who participate in these acts of violence is important

One of the most complex debates is how to stop the terrorist as well as how to provide opportunities for young men in areas that are known for breading impressionable men into vicious killers. Understanding why someone would give their life for a cause such as that is mind boggling. Understanding the religious, social, and economical reason why will give some sort of idea as to the reasoning behind such a decision.

The emergence of terrorism in the Western world has existed for some time. Globalization over the last century has put many countries and their citizens in a position of poverty and strife. They believe the cause of their issues are Western countries. Many terrorist act too place in that time frame. However, it culminated in 2001 when America was attack on Sept 11, 2001. American citizens did not expect nor were they ready for such a heinous attack.

Since 2001 many have undertaken the effort of understand why terrorist do what they do. Thousands of studies have taken place to get into the mind of those who commit these act. Researchers have engaged in working of figuring out why religious, social, and economical issues cause terrorist carry out attacks. Most international terrorist are from the middle east. The attackers on Sept 11 we all from the middle east. Those societies are male dominated with many of their youth lacking opportunities as well as being religious.

1. Problem

The problem being analyzed is the behavior of terrorist when they take hostages. In many terrorist attacks hostages are taken as ransom or just a pawns in the terrorist plans. Our analysis will also investigate the weapon of choice as well as the year the attack took place.

1. Motivation

The motivation for this analysis is to discover any trends in the way terrorist think when they decide to take hostages. It also seeks to understand the underlying issues of why terrorist do what they do.

1. Sub-Problems

Sub-problems include but are not limited to geographical location of the attacker, the reason or reason for the attack. These are important factors to understand when analyzing such a subject.

1. Approach

This problem cannot be completely solved however, solutions to limit the spread of terrorism around the world are invaluable. Providing economic assistance and opportunities to those most susceptible to the attraction of joining such groups. Secondly, educating individuals that religious beliefs are personal and violence to indoctrinate others does not solve any goals. If these goals can be achieved this will make for a much safer and prosperous world. To solve this problem a dataset consisting of all terrorist attacks from around the world for the last 50 years. 6 different algorithms will be used to analyze different metrics within the dataset.

The approach to solving this problem will take a similar approach to those before. Machine learning techniques will be implemented to provide the most accurate results possible. However, the difference between those researchers and this project will be the metrics used from out data to answer this pressing problem. After viewing the data a determination will be made as to which data points will be the most impactful for analysis. Naïve Bayes, Tree, Linear models are just a few of the algorithm which will be used in this analysis.

# Background

All of the possible supervised learning algorithms have limitation. They are all dependent on the data available.  “The performance of any supervised learning model is entirely dependent upon the representation of the data it receives”[1]. Machine learning is being used by most researchers to figure out what to do and how to profile a possible terrorist. This quantitative approach is relatively new considering that terrorism is been around for hundreds of years. The development of social media, televisions and globalization has thrust the threat of terrorism into every home in American an countries worldwide.

# Literature Review

Using Big Data has become a very important role in figuring out what terrorist may have planned. Almost all of the current analysis use large amounts of data to analyze the question being asked. With the advent of social media and the use of cell phones over the last 20 years mush more data is available for use. These pieces of information, each be unique and distinct are useful information for analysis. This information put the Data Scientist in the mind of a terrorist to determine their way of thinking and can help determine what they will do next.

With the boom of social media over the last 15 years studies of terrorist interaction on sites like Facebook and twitter have been closely monitored. “A complete social network contains a quantity of members with incoming and outgoing ties. If our information is limited to someone’s connections to other network members and does not include, in turn, their interconnections, then this would represent a so-called ego-network An ego-centered network is still valuable for some research analyses, such as the investigation of an individual’s general embeddedness, but it is less comprehensive and thereby lacks many analytic possibilities with regard to the investigation of structural patterns” [2].

‘PNA (Paragon Network Analysis) algorithms directly incorporate time-dependent data about communication events to characterize the dynamical evolution of a network. The initial tests indicated that PNA algorithms can track the changes in a terrorist cell as it transitions from a covert “sleeper” state to an active state”.[3] This method was more reliable than previous methods. It was able to capture a time-dependent data of communication between a network of terrorist giving the ability to see how changes occur in a network.

Another method being implemented is the behavioral surveillance approach. Monitoring how a subjects behaviors changes can shed light on their mindset and possible give the authorities an idea as to by who and when an attack will take place. “Behavioral surveillance seeks to detect physiological behaviors, conditions, or responses and the attendant biological activity that indicate that an individual is about to commit an act of terrorism. Specifically, behavioral surveillance seeks to detect patterns of behavior thought to be precursors or correlates of wrongdoing”. [4]

Classification techniques have been used to build probabilistic models to determine the probability of something occurring. “It is used for several research applications including fraud detection, terrorism prediction, target marketing, performance prediction, medical diagnosis, finance, weather prediction, business intelligence, homeland security”[5].

“Based on this analysis the belief more data lead to better predictive models is not always true, because noise in the data can overwhelm predictive models” [6]. This analysis suggest that bots being used on social media can skew results. Determining accurate data is important in providing accurate results.

Geographic profiling of possible terrorist can give the authorities the ability of find terrorist based on location. It can also provide advice insight to prepare for possible attacks and help with the rescue of victims. “Most terrorism researchers have traditionally tended to utilize historical case studies, descriptive statistics, or time-series analysis. The first to systematically examine geographical aspects of terrorism were scholars from the discipline of geography” [7]. Geography is an important aspect of being able to identify and catch terrorist.

“Artificial Intelligence is another method being implemented to fight terrorism. Automated data analytics are used to support the activities of the intelligence and security services, particularly through data visualization. Algorithms prioritize terrorist suspects and routinely assess the risk of air-travel passengers. Information can be collected and stored by default, to be analyzed at a later time with a view to revealing patterns and links that expose terrorist networks or suspicious activities”[8]. Amazingly, over the last 18 years since 9/11 a vast list of individuals has been added to no fly list. The issue is some of these individuals are not a threat. This is one of the problems profiling faces.

Big data poses notable challenges to the intelligence process. When massive amounts of data are gathered, stored, processed and used simultaneously, the intelligence process is compressed and there is a risk that inherently human activities, such as analysis, are undermined by an over-reliance on automation[9]. In the age of technology gathering data has become a top priority for intel agencies. However, gathering this data and analyzing takes out the human factor. This writer of this paper inquires as to how to fill the human gap.

“Effective investigation, pursuit and arrest after the event is little deterrent, particularly when dealing with would-be suicide attackers. In counter terrorism, prevention is most certainly better than cure”[10]. Essentially, catching the terrorist before an attack is the only way to deter.

1. Approach

The approach taken to solve this problem come in various forms. Times series is used to predict the amount of attacks for the next 5 years based on the previous 48 years of data collected. Clustering is used to group variables together which will in turn enable us to predict using other predictive models without having to many variables in a model. “Clustering algorithms partition data into a certain number of clusters (groups, subsets, or categories). There is no universally agreed upon definition”[11]. A Linear models is used to predict the number of hostages released based on the number of hostages taken. Naïve Bayes will predict unknown number of hostages taken.

Subproblem

Subproblem 1 relates to the geographical location of the attacks. Most attacks take place in the Middle East. However, terrorist have become much more savvy and over the last 20 years have begun to attack lager metropolitan areas to exact mass casualties.

Subproblem 2 relates to the motivation of terrorist and why it is they take hostages. Most of the attacks are to strike fear however, there are many instances that hostages are held against their will. These problems will be solved using machine learning techniques such as Time Series, Naïve Bayes and Linear Regression models

1. Data

The data represent a wide range of events dating back to 1970. The data represents every region of the world. It is stored in a CSV file which will be uploaded to a database. The data is structured. R will connect with this database with the dbconnect function to pull the data for analysis. The data set has over 150 variables consisting of both categorical and numeric values. The file has over 11k rows of data along with the 150 variables.

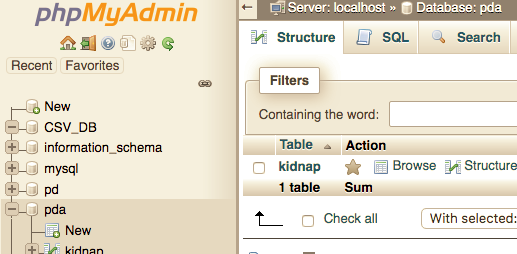
1. Data Source

Data was source from <https://www.kaggle.com/cemilisilak/global-terrorism/>

This choice was a last resort due to our inability to use AWS. Instead we extracted the data directly from github.com. After which a subset was done to prepare for analysis.

Intentions were to put on AWS however, due to unforeseen circumstances we were unable to.

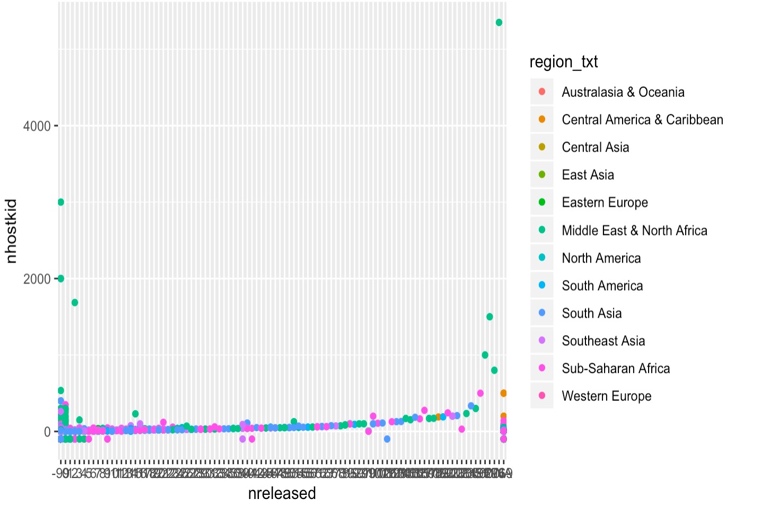
Data Schema



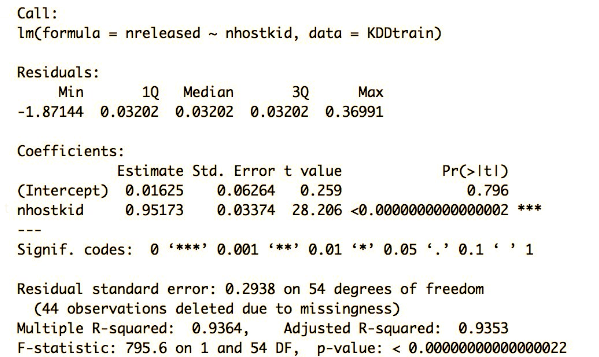
|  |  |
| --- | --- |
|  |  |

1. Results

Linear Model

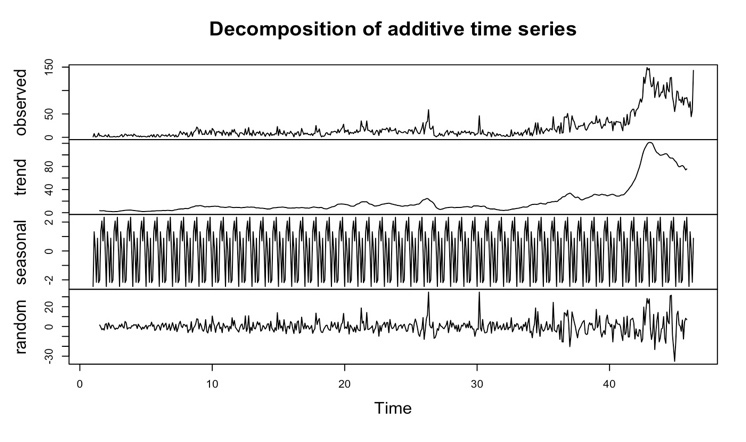


Above is a ggplot from a Linear regression model



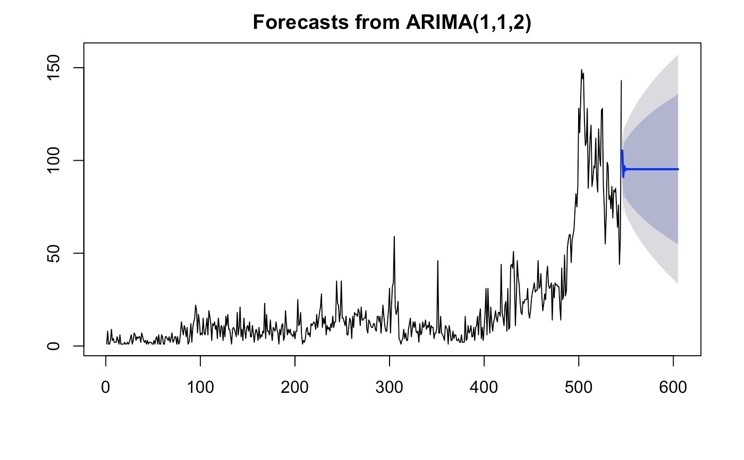
Based on the Linear Model the R-Squared is predicting with .93% accuracy the amount of hostages released from the total taken. This model has provided a high level of accuracy.

Time Series



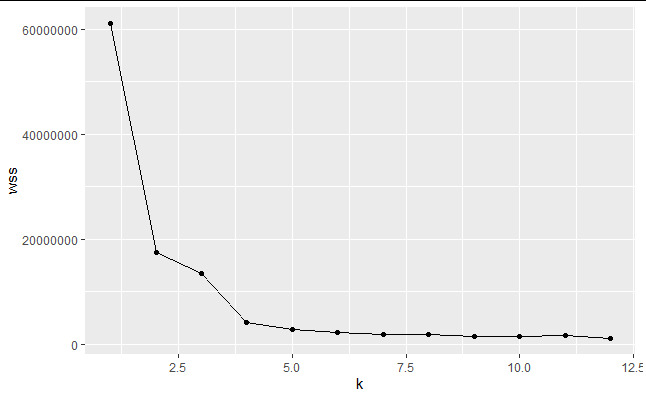
Decomposition Plot

There is no seasonality or trend but it is evident that there is a boom occurring in the last few years. Further analysis is required.



Arima Model predicting the number of attacks over the next 5 years. There is not trend predicted.

K-means



This graph shows the number of K’s the model will use for clustering.

Discussion

During the first section of the project, during the exploratory analysis, hypothesis test was performed as there were number of variables which information was recorded only after year 1997. This result led to the rejection of the null hypothesis that populations were not significantly different, and the subset of the data was now an obligation.

After the data manipulation process, continuing the exploratory analysis, we started with the process of clustering those numerical variables that had too many different values as to use for other machine learning algorithms. This process took place by developing two different functions, one to perform the elbow plot which will help determine K and then another one to predict the clustered data.

Following, the data was ready to start predicting the “Unknown” values of the wanted variables. A function that would take a given model and a wanted vector was developed in order to predict those needed unknown values. Naïve Bayes and Decision trees were implemented and accuracy values were compared for the “nhostkid” clustered variable, which explains the number of hostages taken into custody. The accuracy of the first method was 0.93, compared to 0.54 from the decision tree.

Finally, following the same chain of command, correlation analysis helped us to determine that the number of hostages released was strongly correlated (0.94) with the number of hostages taken into custody. Linear regression model was implemented by using the number of hostages taken into custody variable to predict the number of hostages released. The accuracy of the model was 0.93.

Independently of the other analysis, Time Series method helped us to understand the change in number of attacks along the years. The results explain that there was neither trend nor seasonality. As a result, the predicted future 5 years had a straight line with high variance.

References

V. Kumar, M. Mazzara, A. Messina, and J. Lee, “A Conjoint Application of Data Mining Techniques for Analysis of Global Terrorist Attacks,” *Advances in Intelligent Systems and Computing Proceedings of 6th*

*International Conference in Software Engineering for Defence Applications*, pp. 146–158, 2019.

V. Kumar, M. Mazzara, A. Messina, and J. Lee, “A Conjoint Application of Data Mining Techniques for Analysis of Global Terrorist Attacks,” *Advances in Intelligent Systems and Computing Proceedings of 6th International Conference in Software Engineering for Defence Applications*, pp. 146–158, 2019.

COMMITTEE ON TECHNICAL AND PRIVACY DIMENSIONS OF INFORMATION FOR TERRORISM PREVENTION AND OTHER, *Protecting individual privacy in the struggle against terrorists*. Place of publication not identified: NATIONAL ACADEMIES Press, 2008.

“Big Data: An Approach for Detecting Terrorist Activities with People’s Profiling ,” *Big Data: An Approach for Detecting Terrorist Activities with People’s Profiling*, 14-Mar-2018. [Online]. Available: http://www.iaeng.org/publication/IMECS2018/IMECS2018\_pp196-201.pdf. [Accessed: 08-Dec-2019].

“Social Network Analysis in the Science of Groups: Cross-Sectional and Longitudinal Applications for Studying Intra- and Intergroup Behavior,” *Social Network Analysis in the Science of Groups: Cross-Sectional and Longitudinal Applications for Studying Intra- and Intergroup Behavior*, 215AD. [Online]. Available: https://www.apa.org/pubs/journals/features/gdn-0000021.pdf. [Accessed: 08-Dec-2019].

“Terrorism Research Initiative,” *Terrorism Research Initiative*, Feb-2013. [Online]. Available: https://www.jstor.org/stable/pdf/26296908.pdf?refreqid=excelsior:ca97a5e8409218eabfb63ee4def42397. [Accessed: 08-Dec-2019].

“An Integrated Machine Learning Approach To Studying Terrorism,” *An Integrated Machine Learning Approach To Studying Terrorism*, 20-Apr-2018. [Online]. Available: https://cogsci.yale.edu/sites/default/files/files/Thesis2018Peng.pdf. [Accessed: 08-Dec-2019].