
HOMELAND SECURITY BY USING DATA MINING PATTERN DETECTION FOR COUNTER TERRORISM AND CRIMES

D.V. Patange

Arts & Science College, Chikhaldara, Dist. Amravati (M.S.) India
deepti_pethkar@rediffmail.com

ABSTRACT

Data mining can be used to model predictive analytical pattern detection for crime and counter terrorism problems. Terrorism and crimes are social tragic events which our society has to pay cost in drastic effects on many aspects. In the field of computational and information science data mining is an analytical process that primarily involves searching through vast amounts of data to spot useful, but initially undiscovered patterns in solving crimes faster way. About 100% of crimes only 10% of the criminals commit about 50% of the crimes detected. Here we use clustering algorithms from data mining to detect the crimes patterns and speed up the process of solving crimes. We will look at k-means clustering with some enhance to apply these techniques to real crime data from sheriff's office and validated our result. We also use semi-supervised learning techniques here for knowledge discovery from the crime records and to help increase the predictive accuracy. We also develop a weighting scheme for attributes here to deal with limitations of various out of the box clustering tools and techniques. This easy to implement data mining frame work with the geospatial plot of crime and helps to improve the productivity of the detectives/agencies. It can also be applied for counter terrorism for homeland security.

Key Words: Homeland security, Data mining, Crime-patterns, clustering, algorithm.

Introduction

The tragic event of September 11th have caused drastic effects on many aspects of society for security of their life and homeland. Terrorists and criminals are typically indistinguishable from the local civilians. They aren't part of an organized, conventional military force-rather, they form highly adaptive organizational webs based on tribal or religious affinities. Historically solving crimes has been the prerogative of the criminal justice and law enforcement specialists with the increasing use of computerized systems to track crimes, computer data analysts have started to help law enforcement officers and detectives to speed up the process of solving crimes. As well as they conduct quasi-military operations using instruments of legitimate activity found in any modern

society, making extensive use of internet, cell phones, schools and temples (worship house), prisons, hospitals, hotels, commercial vehicles and financial systems. Terrorist deliberately attack civilians by the aim to kill as many people as possible and try to create threat and destruction in society for homeland security. Here we will take an inter disciplinary approach between computer science and criminal justice to develop a data mining paradigm that can help to solve crimes faster. More specifically, we will use clustering based models to help in identification of crime patterns (Chen H. et.al., 2003).

Of the numerous challenges to countering terrorism, none are more significant than being able to detect, identified and similarly, suspect refers to the person that is believed to have committed the crime is not a convict until

proved guilty. Hence the suspect may be identified or unidentified but the victim is a person who is target of the crime. Most of the time the victim is identifiable and in most cases is the person reporting the crime. Additionally, the crime may have some witnesses. In our judgment, if preemption is a goal, the key to detecting terrorists is to look for patterns of activity indicative of terrorist plots based on observations of current plots and past terrorist attacks, including estimates about how terrorist will adapt to avoid detection.

Cluster (of crimes) has asocial meaning and refers to a geographical group of crimes in region. Such clusters can be visually represented using geo-spatial plot of crime overlaid on the map of the police jurisdiction. The densely populated group of crime is used to visually locate the 'hot-spots' of crime. However, when we talk of clustering from a data mining standpoint, we refer to similar kinds of crime in the given geography of interest. Such clusters are useful in identifying a crime pattern. Our hypothesis about these well known crimes like a serial-rapist, DC sniper or a serial killer and terrorist plan to launch an attack, the plot must involve people (the terrorist, their finances and so forth). The transactions all these people conduct will manifest in database owned by public, commercial and government sectors and will leave a signature-detectable clues in the information space. Because terrorists operate worldwide data associated with their activities will be mixed with data about people who aren't terrorist. If the government wants to access to such activity data, then it must have some to protect the privacy of those who aren't involved in terrorism.

It Enabled Counter Terrorism Infrastructure and Crime Reporting Systems in Developed Countries

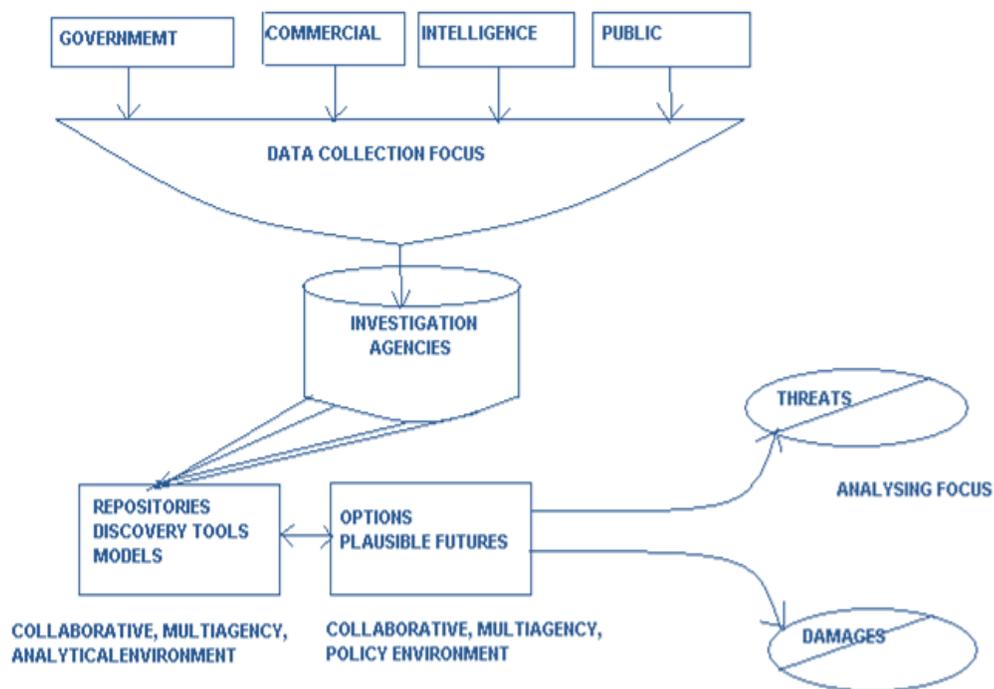
Law enforcement agencies in developed countries have become more vigilant about criminal and terrorist activities. But the data for crime often presents an interesting dilemma. While some data is kept confidential, some becomes public information. Data about prisoners can often be viewed in country or sheriff's site. Most sheriff's offices and police department use electronic system for crime reporting that have replaced the traditional paper-based crime reports. These crime reports have the following information categories namely-type of crime, date/time, location etc. then there is information about suspect (identified or unidentified), victim and the witness. Additionally, there is narrative or description of the crime and Modus Operandi (MO) that is usually in text form. The police officers and detectives use free text to record most of their observations that cannot be included in checkbox kind of pre-determined questions. While the first two criteria of information are stored in computer by databases as numeric, character or date fields of table, the last one is often stored as free text. This free text field in data mining is a main challenging task, because free text field can give the newspaper columnist, a great story line, converting them into data mining is not always an easy job.

Countering modern terrorism is a multi and interdisciplinary activity. That is why researcher in the natural, computational

and social science as well as engineering, medicine and many other fields have directed their research in science and technology to help enhance the capabilities in fighting the new counterterrorism war. Amongst all the technologies, Information Technology has been cited as the most important tool in making a country/homeland safer from terrorism. IT

can support intelligence and knowledge discovery by collecting, processing, analyzing and developing applications for terrorism and crime related data. Federal, state and local authorities can use the results to make timely decisions, select effective strategies and tactics and allocate appropriate resources to detect, prevent and respond to future attacks.

Figure 1: Counter Terrorism Framework



In the report of National Strategy for Homeland Security, the US Department of Homeland Security (DHS) has identified six critical mission areas where IT can contribute to accomplishing strategic national security objectives.

1. Primitiveness through intelligence warning.

2. Smart Borders.
3. Domestic Counterterrorism.
4. Protecting critical infrastructure and key assets.
5. Defending against catastrophic terrorism.
6. Emergency preparedness and responses.

Data Mining and Crime Patterns with Clustering Techniques Used

We will look at how to convert crime information into a data-mining problem (Chen H. et.al., 2004), such that it can help the detectives in solving crimes faster. We have seen that in crime terminology a cluster is a group of crimes in a geographical region or a hot spot of crime. Whereas, in data mining terminology a cluster is group of similar data points – a possible crime patterns. Thus appropriate clusters or a subset of the clusters will have one-to-one correspondence to crime patterns and counter terrorism.

Thus clustering algorithms in data mining are equivalent to the task of identifying groups of records that are similar between themselves but different from the rest of the data. In our case some of these clusters will useful for identifying a crime spree committed by one or same group of suspects. Given this information, the next

challenge is to find the variables providing the best clustering. These clusters will then be presented to the detectives to drill down using their domain expertise for automated detection of crime patterns, allows the detectives to focus on crime sprees first and solving one of these crimes results in solving the whole “spree” or in some cases if the groups of incidents are suspected to be one spree, the complete evidence can be build form the different bits of information from each of the crime incidents. For instance, let us take an oversimplified case of crime record that suspect has black hair, the next witness reveals that suspect is middle aged and third one reveals there is tattoo on left arm, all together it will give a much more complete picture from bits of information from different crime incidents. A crime data analyst or detectives will use a report based on this data stored in different orders, usually the first sort will be on the most important characteristic based on detective`s experience.

Crime Type	Suspect Race	Suspect Sex	Suspect Age	Victim Age	Crime Time /Date	Weapon
Robbery	B	M	Middle	Elderly	10 pm	Knife
Robbery	W	M	Young	Middle	12pm	Bat
Robbery	B	M	?	Elderly	11 pm	Knife
Robbery	B	F	Middle	Young	-	Piston

Table A: Simple Crime example

We look at table-A with a simple example of crime list, where 1 and 3 row show a simple crime pattern where the suspect description matches and victim profile is also similar.

The main intension of data mining to detect much more complex patterns. Because in real life situation there are many attributes or factors for crime and terrorism which often gives us partial information available about them. In such situation for computer

analyst it is not so easy to identify these patters by simple querying. Thus clustering technique using data mining comes in handy to deal with enormous amount of data and dealing with missing data about

Results of Crime Patterns Analysis

The proposed system is used along with the geo spatial plot. The crime analyst may choose a time range and one or more type of crime from certain geography and display the results graphically.

Subsequently, we cluster the crimes based on our weighing techniques, to come up with crime groups which contain the possible crime patterns of crime sprees. The geo-spatial plot of these crime patterns along with the significant attributes to quantify these groups is presented to the detectives who has now vary easy task to identify crime sprees than from the hundred of listed crimes. On this same way for our homeland security our countries should support the range of technologies and systems to identify and find individual terrorist. Among the various technologies can help to find terrorist before they strike must include: Improved data sharing; “Smart ID cards” with biometric information; “Smart visas” and improved border security; Digital surveillance data and Face recognition technology.

Counter Terrorism Domain Challenges

1. Distribution of activities
2. Heterogeneous sources and formats
3. Crime and intelligence analysis

These problems can be addressed by

the crime evidences.

We used k-means clustering technique here, as it is one of the most widely used data mining clustering technique.

addressed by implementing the following

- Integrate information shearing across the federal government.
- Integrate information shearing across state and local governments, private industry and citizens
- Adopt common “meta-data” standards for electronic information relevant to homeland security
- Improve public safety emergency communications and
- Ensure reliable public health information.

Conclusion and Future Direction

We looked at the use of data mining for identifying crime patterns crime pattern using the clustering techniques. Our contribution here was to formulate crime pattern detection as machine learning task and to thereby use data mining to support police detectives in solving crimes. We identified the significant attributes; using expert based semi-supervised learning method and developed the scheme for weighting the significant attributes. Our modeling technique was able to identify the crime patterns from a large number of crimes making the job for crime detectives easier.

Some of the limitations of our study include that crime pattern analysis can only help the detectives, not replace them. Also data mining is sensitive to quality of input data that may be incorrect, have missing information, be data entry error prone etc.

Also mapping real data to data mining attributes is not always an easy task and often required skilled data miner and crime data analyst with good domain knowledge. They need to work closely with a detective in initial phases.

As a future extension of this study we will create models for predicting the crime hot-spots that will help in the deployment of police at most likely places of crime for any given window of time, to allow most

effective utilization of police resources. We also plan to look into developing social link networks to link criminals, suspects, gangs and study their interrelationship. Additionally the ability to search suspect description in regional, FBI databases, to traffic violation databases from different states etc. to aid the crime pattern detection or more specifically counter terrorism measures will also add value to this crime detection paradigm for homeland security.

References

Chen, H., Chung, W., Qin, Y., Chau, M., JieXu, J., Wang, G., RongZheng, HomaAtabakhsh (2003). Crime Data Mining: An Overview and Case Studies, AI Lab, University of Arizona, proceedings National Conference on Digital Government Research.

Chen, H., Chung, W., Qin, Y., Chau M., JieXu, J., Wang, G., RongZheng, HomaAtabakhsh (2004). Crime Data Mining: A General Framework and Some Examples, IEEE Computer Society.

Chen, H. (2008). Homeland Security Data Mining using Social Network Analysis, IEEE ISI Keynote Talk (III).

Vani, S. (2010). Counter Terrorism, IEEE International Conference on Computational Intelligence and Computing Research.

Poppetal, R. (2004). Countering Terrorism through Information Technology, Comm. ACM, vol. 47, no. 3, pp. 36-43.

Seifert, J.W. (2004). CRS Report for Congress "Congressional Research Service ~ The Library of congress, Updated December 16.