Research Article

Model for Predicting Terrorist Group by CLOPE Algorithm

Pawan H. Pilley and S. S. Sikchi

Information Technology, Sant Gadge Baba Amravati University, India

Accepted 10 May 2014, Available online 01 June 2014, Vol.4, No.3 (June 2014)

Abstract

In the present scenario terrorist attacks are biggest problem for the mankind and whole world is under constant threat from these well-planned, sophisticated and coordinated terrorist operations. Now every country is focusing on counter-terrorism. Counter-terrorism is the practices, tactics, strategies and techniques that governments, militaries, police and security agency uses to prevent or in response to terrorist threats. Main objective of this paper is to focus on terrorist group prediction model technique for counter-terrorism. Also, analysis of predicting the responsible group using CLOPE algorithm.

Keywords: Group Detection; Clustering; Terrorist Group; predicting terrorist attack; CLOPE algorithm; counter – terrorism.

1. Introduction

Prediction of terrorist group using historical data of attacks has been less explored due to the lack of detailed terrorist data which contain terrorist group’s attacks and activities. The reasons may be its confidentiality & sensitivity. Intelligence agencies are having large amount of data. They are continuously monitoring terrorist activities. But they are not having enough trained officers to process bulk data in very less time period for the purpose of decision-making about terrorist attacks. In counter terrorism first step after any incident/attack is to find the group names that were involved and to make strategy to catch them. This paper discusses a model for the identification of responsible groups for an attack based on the available information.

2. Data Mining for Counter-Terrorism

Security is an important aspect that has been given top priority by all political and government worldwide and are aiming to reduce crime incidence (David G., 2006, A. Malathi and Dr. S. Santhosh Baboo, 2011) Intelligence analysis might be applied to any of several recognized intelligence sources like Signals Intelligence (SIGINT), Imagery Intelligence (IMINT), and Open Source Intelligence (OSINT) (H. Chen, D. Denning et al., 2011)

Current terrorism informatics, which aims to help security officials using data mining techniques, is mainly focused on using social network analysis (SNA) for structural and positional analysis of terrorist networks where required information is provided from non-crime data (Coffman, T.R., Marcus, S.E 2004). Prediction of terrorist group using historical data of an attack has very less work; this is because of lack of detailed terrorist data which contain terrorist group’s attacks and activities (Faith Ozgul, Zeki Erdem and Chris Bowerman, 2009). The use of data mining technologies in counter terrorism has been flourishing since the U.S. Government encouraged the use of information technologies (Taipale KA, 2003). This paper, focus on developing a prediction model using historical data to predict the terrorist group involved in a given attack. Our database includes terrorist attacks in different countries from year 1970 to 2011.

3. Group Prediction Technique

Prediction of terrorist group after an attack is one of the most important steps for counter terrorism. As soon as we are able to find the involved group name, we will be able to make strategies to catch the culprits. Generally, the terrorist group responsible for an attack is detected by using email, telephone signal information, terrorist web sites, social network analysis etc. Terrorist activities occurred in past are available in criminal data/historical database. This database can be used to detect terrorist group responsible for an attack. Frequency of terrorist attacks can also be analyzed.

4. Terrorist Group Prediction Model

Terrorist group prediction model (TGPM) which learns the pattern of terrorist attacks from the available historical data and make an association between terrorist group and previous attacks. Every terrorist group can be differentiated based on the style of attack, targets like police, private organizations; public property etc. so by analyzing these patterns TGPM will predict the group that may be involved in a given incident.
TGPM is developed to detect the responsible terrorist group by using historical data. TGPM uses the concept of Crime Prediction Model, Group Detection Model (GDM) and Offender Group Detection Model (OGDM) (Ozgul, F., Bondy, J., Aksoy, H Erdem, Z., 2008). TGPM uses various parameters like attack type, location, target type, weapon type, hostage/kidnapping and suicide attack etc. TGPM uses terrorist corpus, parameter’s value and parameters weight as input.

A. Group Detection Model (GDM)

GDM is a general detection model based on co-offending clustering. It works by linking co-offenders with inner join query using unique crime reference id numbers. These links collected together to create criminal networks as graphs, then Strongly Connected Components (SCC) algorithm (T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, 2001) is applied so that each connected component can be treated as an individual criminal network. SCC algorithm is a way of dividing big graphs into smaller chunks of sub graphs or components. A directed graph is called strongly connected if for every pair of vertices there are paths towards each of them. SCC components of a directed graph are its maximal strongly connected sub graphs. In GDM, each component represents a unique criminal network and one criminal can only belong to a single criminal network. GDM simply finds and links all criminals each other who commit the same crime with the same friends.

B. Offender Group Detection Model (OGDM)

Just like GDM uses co-offending feature, OGDM uses similarity of three crime features which are criminals’ choice of crime location, date, and modus operandi by assuming that similarly behaving criminals would inevitably meet or know each other since they are likely minded. OGDM steps include creation of Spatial, Temporal, and Modus Operandi links, followed by clustering for detecting criminal networks. OGDM is similar to GDM for creating links although GDM counts links between criminals as link weights, whereas OGDM calculates link weights counted and normalized against all links.

5. Dataset and Collection Methodology

The Global Terrorism Database (GTD) is an open-source database including information on terrorist events around the world from 1970 through 2011 (with additional annual updates planned for the future). Unlike many other event databases, the GTD includes systematic data on domestic as well as transnational and international terrorist incidents that have occurred during this time period and now includes more than 104,000 cases. For each GTD incident, information is available on the date and location of the incident, the weapons used and nature of the target, the number of casualties, and---when identifiable---the group or individual responsible. Statistical information contained in the Global Terrorism Database is based on reports from a variety of open media sources.

6. Implementation

The Global Terrorism Database (GTD) was developed to be a comprehensive, methodologically robust set of longitudinal data on incidents of domestic and international terrorism. Its primary purpose is to enable researchers and analysts to increase understanding of the phenomenon of terrorism. The GTD is specifically designed to be amenable to the latest quantitative analytic techniques used in the social and computational sciences.

A. Characteristics of GTD

- Contains information on over 104,000 terrorist attacks.
- Currently the most comprehensive unclassified data base on terrorist events in the world.
- Includes information on more than 47,000 bombings, 14,000 assassinations, and 5,300 kidnappings since 1970.
- Includes information on at least 45 variables for each case, with more recent incidents including information on more than 120 variables.
- Over 3,500,000 news articles and 25,000 news sources were reviewed to collect incident data from 1998 to 2011 alone.

Data preprocessing is an important step in which missing values are filled up, redundancies are removed and filtering is performed so that the database will be ready for use. Missing values can be filled by using various terrorist databases available over internet for terrorism research (http://www.start.umd.edu/gtd/, 2012).

Figure 1: Terrorist Group Prediction Model (TGPM)

After pre-processing of database, percentage of attacks of each group is calculated based on input parameters. Each parameter is assigned a weight based on its impact over the incident. The group weight is calculated by using the percentage of attacks of each group and the parameters weight. Different clusters are created. Association between these clusters is being performed and highest value from these associations is obtained. Group name corresponding to the highest value may be the most probable responsible terrorist group.
Evaluation of TGPM is performed on standard Global Terrorism Database (GTD). Data of the GTD needs to be processed because some of the values of different fields of the database were missing. Database size is reduced according to the requirement. We have used terrorist incident data from 1970 to 2011. The different parameters used in our model are attack type, location, target type, weapon type, hostage/kidnapping and suicide attack. This paper is based on TGPM model and its architecture has been used as a reference.

Figure 2: Flow chart diagram of Predicting Terrorist Group

7. Selection of the Clustering Algorithm

Most clustering algorithms define some criterion functions and optimize them, maximizing the intra-cluster similarity and the inter-cluster dissimilarity. The criterion function can be defined locally or globally. CLOPE is specific to categorical attributes, which forms a large part of our database. At each level of the hierarchy; it groups instances within clusters, again with respect to a distance function. Finally, CLOPE calculates the extent to which categorical attributes match with each other, for two given instances. If the level of this matching is greater than a certain threshold, the instances are clustered together. Considering limitations, CLOPE was the only algorithm which clustered data with missing values. Also, it doesn’t require much tuning to efficiently obtain good clusters. Finally, it’s designed specifically for categorical attributes, and its performance is comparable to other algorithms for categorical clustering. Therefore, we selected CLOPE for our experiments.

8. Implementation of CLOPE Algorithm

Like most partition-based clustering approaches, we approximate the best solution by iterative scanning of the database. Our implementation requires a first scan of the database to build the initial clustering. After that, a few more scans are required to refine the clustering and optimize the criterion function. If no changes to the clustering are made in a previous scan, the algorithm will stop, with the final clustering as the output. The output is simply an integer label for every transaction, indicating the cluster id that the transaction belongs to. For categorical attributes, it can be expected that those instances can be clustered which contain matching values for one or more attributes. CLOPE attempts to increase the intra-cluster overlapping of categorical values by increasing the height-to-width ratio of the cluster histogram. It also uses a parameter called repulsion, to control the tightness of the cluster. Different number of clusters can be obtained by varying this parameter.

9. Results

The following Screen shot 1 shows the Attack Prediction page of the system. Here we implemented CLOPE clustering algorithm using the ASP.NET tools. So for that we have created one form which is shown in the following screen shot.

User can predict the names of responsible terrorist groups involved in attack based on different parameters such as

- Nation and Attack type
- Nation and Target type
- Region and Attack type
- Region and Target type

Screenshot 1: Search on the basis of Nationality & Attack Type

Here weight represents particular values assigned to parameters as given in GTD codebook and height represents total number of incidents fetched from
database. After that, it calculates total summation based on height and weight.

**Screenshot 2: Terrorist Group Names Involved**

Screenshot 2 displays all the terrorist group names involved in an attack and based on the higher number of occurrences it finally displays probable terrorist group names.

**Conclusion**

In this study, it can be concluded that by using historical terrorist data it is possible to predict the group involved in the given attack. However, further study can be carried out to detect a terrorist group using historical data. For further research point of view it is suggested to use different artificial intelligence techniques of group detection and include more parameters like phone calls, email data etc. so that more accurate results can be obtained. Method of criminals and terrorists profiling can also be useful.

**References**


Global Terrorism Database (05/02/2012), http://www.start.umd.edu/gtd/. Retrieved on.