Ninth LACCEI Latin American and Caribbean Conference (LACCEI'2011), Engineering for a Smart Planet, Innovation, Information Technology and Computational Tools for Sustainable Development, August 3-5, 2011, Medellín, Colombia.

Finding Patterns of Terrorist Groups in Iraq: A Knowledge Discovery Analysis

Steven Nieves
Polytechnic University of Puerto Rico, Hato Rey, Puerto Rico, USA, stevennieves@yahoo.com

Alfredo Cruz
Polytechnic University of Puerto Rico, Hato Rey, Puerto Rico, USA, alfredcross@gmail.com

ABSTRACT

In this paper we explore the application of data mining to model terrorism activity in Iraq. To this end, we experiment with the use of data mining algorithms to support in the process of the identification of terrorism patterns. We applied data mining techniques to real terrorism data from the Global Terrorism Database (GTD) of the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The data mining techniques used in this study discovers inherent information of different terrorist organizations according to the different types of terrorist acts they commit. The results in this paper should be practical not only for counter terrorism security analysts, but also to determine prioritization and geographical allocation of military and law enforcement resources.

Keywords: Data Mining, Clustering, Classification, Association

1. INTRODUCTION

After the tragic events of 9/11/2001, scholars were challenged to research and find feasible solutions that would help protect the public and the security of the nation. As result, domain experts realized that gaining knowledge on the structures of terrorist organizations and learning how they operate was the key to winning the so-called war on terror. Law enforcement and intelligence agencies are helpless without reliable data and data analyzing techniques. As an outcome of further analysis of the 9-11 events, came legislation and implementation of the USA Patriot Act. This Act resulted in government expansions such as new electronic communications surveillance systems, new law enforcement agency divisions for terrorism related crimes and major projects for developing systems for sharing of law enforcement data between national and international law enforcement agencies (Chen, 2007). These expansions represent new opportunities for data mining tools and techniques to demonstrate effectiveness on analyzing data and producing knowledge that assists in making critical decisions. Perhaps one of the most aggressive government implementations of data mining projects can be observed in systems such as the known Defense Advanced Research Program Agency (DARPA) Total Information Awareness (TIA). The TIA program focus was to integrate technologies to collective data, and apply sophisticated data mining techniques such as link analysis, to generate descriptive models and predictive hypothesis with the purpose to apply these resulting rules to new datasets and identify terrorists and terrorist groups. Also, the Multi-State Anti-Terrorism Information Exchange (MATRIX) project, managed by the Department of Homeland Security (DHS), was also used to analyze public records and help law enforcement produce results that focused on preventing terrorist attacks.

These projects intended to demonstrate the effective use of data mining for counter-terrorism, but eventually awakened the general public concerns of intrusions on privacy, their impact on liberties, and the possible violation of the Fourth Amendment. Consequently the United States Congress enforced measures to monitor, regulate and establish standards on the use of private information in government data mining projects (Thuraisingham, 2002). Congressional research agencies highlight how untested these programs are and possibly ineffective. In addition, these programs were found to search data without the approval of a warrant or other protective judicial measure.

Unlike the systems implemented and mentioned above, in this paper we will demonstrate the effectiveness of data mining techniques for the use of counter-terrorism without compromising privacy. By means of simple data mining algorithms, this paper will focus on the approach of discovering patterns in data of terrorist groups, based on recorded acts.

The main purpose of this paper is to propose a structure using simple data mining techniques that can be used to categorize terrorist groups and assist in prioritizing the geographical allocation of limited resources (Popp and Yen, 2006), and aid in generating valuable information relevant to security analysts. With the results presented in this paper, data mining opponents can have a new perspective on data mining as a tool that provides information for making intelligent and critical decisions.

2. TERRORISM

Terrorism has expanded to the point where terrorist movements have become a significant influencing factor in international politics (Lia, 2005). For this reason, numerous efforts to study terrorist activity have emerged. Most of these attempts have failed in evaluating the transforming conditions that inhibit changes of terrorism environments. Earlier forms of terrorism were mostly seen in assassinations that intended to influence political changes. Terrorism has evolved, and what it is known for today needs to be redefined. Perhaps the most used definition of terrorism has been adopted by the United States State Department, Department of Defense, and Central Intelligence Agency. In their terms, terrorism is the premeditated, politically motivated violence perpetrated against non-combat targets intended to influence an audience (Stout, 2002). But, modern terrorism acts include hijacking, taking of hostages, car bombs and human bombs. The targets not only include government and industry leaders, but also economic, military, political and religious as well. Terrorism can take many forms, but each requires to be addressed through different approaches. This is the purpose of modeling a process that can examine terrorist groups. An approach to address insurgency violence making use of heavy military artillery and airstrikes is not reasonable. As well as assigning police to undertake terrorist hideouts. In fact, different types of violence will require different solutions, depending on their specific character. By analyzing the data of terrorist acts, identifying and classifying terrorist groups, we can increase our perception on their mode of operations. Eventually, this will facilitate the discovery and development of significant targeted intervention strategies.

This paper will focus on studying contemporary terrorism activity in Iraq. This country comes into view because of its strategic location in the center of the Arab community and it is also in immediate proximity to high risk targets such as the Persian Gulf oil fields. The terrorist groups in the Middle East region are constantly trying to find ways to gain access to operate and use this country as a launch pad to export terrorism and insurrection to adjacent countries. Even though an enduring United States military presence during the past years has weakened the incidence of terrorist groups in Iraq, the war on terror has prolonged almost a decade. Of course, the irregular character of terrorist activity makes triumph exceptionally complicated for United States military forces (Cordesman, 2003). Also, the proportions of military resources have reduced and currently are not enough for a successful mission. Consequently, the development of new intelligence tools is critical to establish priorities on the use of limited military resources at this phase of the war on terror

3. DATA MINING

Data mining is a multidisciplinary domain that includes efforts from areas of database technology, artificial intelligence, machine learning, neural networks, statistics, pattern recognition, knowledge base systems, high performance computing, and data visualization fields. As data mining has developed, it is commonly acknowledged to be a particular stage in a larger process known as Knowledge Discovery. The term Knowledge Discovery refers to the extensive process of finding knowledge in databases and it is focused on the actions leading to tangible data analysis including the assessment and presentation of results. The first phase of Knowledge Discovery starts with data selection. The goal of this phase is to extract only the data that is significant to the data mining analysis. This data extraction helps to efficiently accelerate the following sub processes. Next, the data preprocessing phase emphasizes with data cleansing and preparation tasks such as eliminating missing values in the data to guarantee that values have a consistent meaning.

The data transformation phase aims at converting the data into preferred practical formats. Subsequently, the data mining phase analyzes the data by means of proper set of algorithms in order to discover meaningful patterns and rules that are capable to produce models. Finally, interpretation and evaluation phase is aimed at selecting those models that are valid and useful for making decisions. Figure 1 illustrates the phases of the Knowledge Discovery Process. Every task in the Knowledge Discovery process produces some output which can be reliably used as input on the next step. Nevertheless, it is favorable to save the results which are output at each step. This will allow analysts to experiment with different algorithms in the data mining phase without having to repeat all the previous steps. Data mining itself has developed into a research area with escalating significance due to its capability of helping analysts extract useful information from large databases. A general misinterpretation about data mining is to expect that data mining can independently work all of the important knowledge that is set in a given large database, without human analysis involvement. Data mining is constituted by a set of tools and techniques that are used by analysts to extract patterns that subsequently expose knowledge. This can all be accomplished with various data mining techniques, such as clustering, association rules and classification.

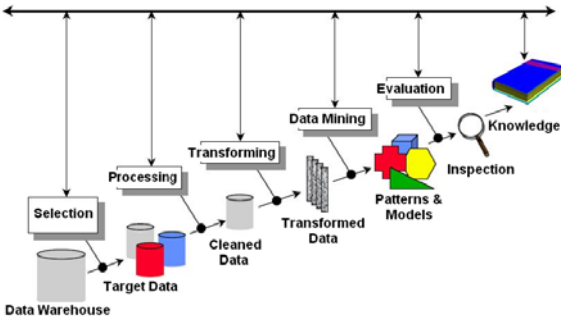

Figure 1. Phases of Knowledge Discovery Process

3.1 CLUSTERING

Clustering techniques group data according to logical relationships and organizes collections of truthfully characterized objects as clusters. The success of clustering is regularly considered intuitively in terms of how practical the outcomes appear to be to the analyst. Clustering analysis allocates each instance of the dataset to the cluster in which it naturally fits in (Fukunaga, 1990). For this study we experiment with K-means. K-means is a typical and commonly used algorithm for clustering operations due to its simple and efficient technique. In the K-means algorithm each cluster's center is represented by the mean value of the objects in the cluster. Subsequently, each point is assigned to its nearest cluster center, so the general outcome is to reduce the total squared distance from all points to their cluster centers. Selecting the mean as the cluster center minimizes the total squared distance from every cluster point to its center. With this technique it is achievable to characterize and isolate every object in the data set into different groups by reducing the mathematical distance between the consequential groups or clusters. Although there are many descriptions of mathematical distance, in this paper we will use the Euclidean distance due to its straightforwardness and ordinary use in clustering techniques. Equation (1) describes Euclidean Distance.

$$dist = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2} \quad (1)$$

Beforehand, the first step is to indicate k as the number of clusters that are required. Then k points are chosen at random as cluster centers. All the data set instances are allocated to their neighboring cluster center according to the regular Euclidean distance metric. After that, the mean of the instances in each cluster is calculated. These means become new center values for their particular clusters. Lastly, the whole process is repeated with the new cluster centers and the process continues to loop uninterrupted until all instances are assigned to a cluster. As a result the cluster centers become stable and remain the same everlastingly. Figure 2 describes the pseudo code for K-means. Perhaps the main disadvantage of k-means is that the number of clusters must be specified as an input to the algorithm prior to initiation of the procedure. As intended, the algorithm is not able to determine a proper

number of clusters and relies upon the analyst to identify this prior to start. In this study, the k input is assigned as the number of values in the dataset for a specified attribute. This approach can assign the instances to the respective clusters based on the dataset attribute we will focus in this study.

3.2 CLASSIFICATION RULES

Classification is a type of data analysis which can anticipate future data trends, categorical labels or discrete values. Various classification techniques have been proposed by researchers in machine learning, expert systems and statistics. Most algorithms are memory inhabitant; normally presuming small data sets. Modern data mining research has constructed upon these efforts, extending scalable classification techniques proficient in handling large quantities of data. OneR is a simple algorithm proposed by Holte (Holte, 1993). The algorithm builds one rule for each attribute in the training data and then selects the rule with the smallest error rate. To create a rule for an attribute, the most frequent class for each attribute value must be determined. The most frequent class is simply the class that appears most often for that attribute value. A rule is simply a set of attribute values bound to their majority class. OneR assumes that all attributes are independent and one attribute is more powerful than the rest. In the event that two or more rules have the same error rate, the rule is chosen at random. The pseudo code for OneR is described in Figure 3.

```
Inputs:
   I = {i1,…,ik} (Instances to be clustered)
   n (Number of clusters)
Outputs:
   C = {c1,…,cn} (cluster centroids)
   m : I → C (cluster membership)

procedure KMeans
   Set C to initial value (e.g. random selection of I)
      For each i, ∈ I
            m(ii) = argmin distance(ii,ck)
                    k∈ {1..n}
      End
      While m has changed
            For each j ∈ {1..n}
               Recompute i, as centroid of {i|m(i) = j}
            End
            For each ij ∈ P
               m(ij) = argmin distance (ij, ck)
                       k∈ {1..n}
            End
      End
return C
End
```

Figure 2. K-Means Pseudo Code

```
For each attribute,
 For each value of the attribute, make a rule as follows:
   count how often each class appears
   find the most frequent class
   make the rule assign that class to this attribute-value
 Calculate the error rate of the rules
 Choose the rules with the smallest error rate
```

Figure 3. OneR Pseudo Code

OneR generates a set of rules that test one particular attribute. OneR is a simple method that often finds excellent rules to differentiate the data structure. It turns out that simple rules frequently achieve surprisingly high accuracy (Holte, 1993). For this reason, one attribute is adequate to resolve the class of an instance with precision. The goal is to build rules that experiment with a single attribute. Clearly the best classification is to use the class that occurs most often in the data. The error rate of the rules can effortlessly be determined by counting the errors that occur on the data. This is the number of instances that do not have the majority class. Each attribute generates a different

set of rules, one rule for every value of the attribute. The error rate for the rule set of each attribute is evaluated and the rule set with the minimum error rate is selected.

3.3 ASSOCIATION RULES

The research on the association rule mining originated from market basket analysis and has been widely used in different applications such as selective marketing, decision analysis, business management, medical application and fault detection. There are two metrics that are used to describe an association rule; support and confidence. These are based on the structure of discovered patterns and the statistics underlying them. An objective measure for the association rules of the form X ==>Y is rule support, representing the percentage of data samples that the given rule satisfies. Another objective measure for association rules is confidence, which assesses the degree of certainty of the identified association. It is defined as the conditional probability that a pattern Y is true given that X is true. More formally, support and confidence are defined as follows:

$$\text{Support: } (X \Longrightarrow Y) = \text{Prob } \{X \cup Y\}$$
$$\text{Confidence } (X \Longrightarrow Y) = \text{Prob } \{X \mid Y\}$$

The confidence factor indicates the strength of the implication rules and the support factor indicates the frequencies of the occurring patterns in the rule. Figure 4 describes the pseudo code for the Apriori algorithm, which is commonly used for generating association rules (Agrawal and Srikant, 1994). Association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Such thresholds can be set by analysts or domain experts. The goal of mining association rules is to find strong rules or in other words, rules with high confidence and strong support. Moreover, many strong rules that appear to be interesting may represent common knowledge, and consequently, are actually uninteresting (Witten and Frank, 2005).

```
Ck: a set of candidate item sets of size k
Lk : the set of frequent item sets of size k
L1 = {frequent items};
for (k = 1; Lk !=∅; k++) do begin
  Ck+1 = candidates generated from Lk;
  for each transaction t in database do
  increment the count of all candidates in Ck+1
  that are contained in t
  Lk+1 = candidates in Ck+1 with min_support
  end
return ∪k Lk-1;
```

Figure 4. Apriori Pseudo Code

Some attention-grabbing observations are based on the analyst's point of view for the data. These patterns are to be interesting if they are unexpected and disagree with the analyst's certainty or present practical information on which the analyst can take action. In this case, such patterns are referred to as actionable. Patterns that are anticipated can be interesting if they prove a hypothesis that the analyst desired to confirm, or are close to the analyst's intuition.

4. DATA SET

The data set for this study has been extracted as a subset portion from the Global Terrorism Database (GTD). GTD was obtained from the National Consortium for the Study of Terrorism and Responses to Terrorism (START), a U.S. Department of Homeland Security Center of Excellence. The START program is also tasked by the Department of Homeland Security's Science and Technology Directorate based at the University of Maryland. GTD is an open-source database including information on terrorist events around the world from 1970 through 2007. GTD includes data of domestic as well as foreign international terrorist incidents that have occurred during this time period and also includes more than 80,000 cases. The subset used for this paper contains terrorism records ranging from 1991 to 2007. For each GTD incident, several attributes are presented such as the date and

city location of the incident, the type of weapons used to commit the terrorist act, the number of casualties, the amount of wounded victims, the type of attack and the identified terrorist group responsible.

5. DATA MINING TOOLS

It is essential to have data mining system tools that can mine multiple kinds of patterns to accommodate to different analyst purposes or applications. Furthermore, data mining systems should be able to discover patterns at various levels of abstraction. To support interactive and investigative mining, data mining systems should be able to allow users to straightforwardly become involved in the output representation. The *Weka* tool, created by the University of Waikato, was selected as the primary data mining tool for the experiments presented in this paper. It is a collection of machine learning algorithms for data mining tasks. *Weka* contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. We selected *RapidMiner* for this study for its user friendly graphical user interface (GUI). In addition the tool has data loaders for relational databases, spreadsheets, and comma delimited files. *RapidMiner* is available as an open source GPL licensed tool developed in Java that presents an integrated development environment (IDE) for data mining, machine learning and business intelligence. It also contains the learning schemes and attributes evaluators of the *Weka* environment. We also explored the use of an emerging technology called *Centrifuge*. The software uses an interesting approach on interactive data visualization, and unified data views. This approach is known as Interactive Analytics and has been applied in problem domains such as homeland defense and cyber crimes. This technology helps analysts understand and reveal key non-obvious relationships in data. By including geographical coordinates of locations this tool is able to plot and export the results to geospatial applications.

6. KNOWLEDGE DISCOVERY ANALYSIS

We present a complete Knowledge Discovery Analysis. Every task in each sub process produces information which can be dependably used as input on the next step.

6.1 DATA SELECTION, PREPROCESSING AND TRANSFORMATION

The GTD database was provided by the START program from the University of Maryland. We extracted a subset of the data that contains terrorism records ranging from 1991 to 2007, for the purpose of using contemporary information. From the extracted dataset we also selected those acts of terrorism documented for the country of Iraq, eliminating unrelated instances belonging to other geographical locations and giving importance to only the geographic area under study for this paper.

Using the resulting dataset, we eliminated irrelative attributes that would have not added significance in the analysis processes. The attributes maintained for the dataset were the date and city location of the incident, the type of weapons used to commit the terrorist act, the number of casualties, the amount of wounded victims, the type of attack and the identified terrorist group responsible. The dataset instances with missing information were also eliminated from the database. This step resulted in a reduced number of 189 clean instances in the dataset.

The GTD was imported into a Microsoft Access database for the ease of generating queries. After completing the data preprocessing phase, the resulting instances were exported from our local database and converted to comma delimited format. The comma delimited format allows our data mining tools to read directly into the resulting files without additional formatting.

6.2 DATA MINING

In the data mining phase we analyze the data using a set of algorithms. In this approach we applied the K-means algorithm for our clustering technique. For classification learning, the OneRule algorithm will be used. And finally, we put together the Apriori algorithm for generating association rules.

6.3 K-MEANS

The dataset was fed into the RapidMiner tool for performing clustering, the first data mining technique applied. Next the k-means operator is added to the project and the k parameter is assigned. For this case, the number of clusters that will be generated is 6, conforming to the number of values for the dataset attribute 'AttackType1_txt' which represents the type of attack for each instance in the dataset. The result of the K-Means clustering operation in Figure 5 shows the dispersed terrorism acts in relation to the attack types.

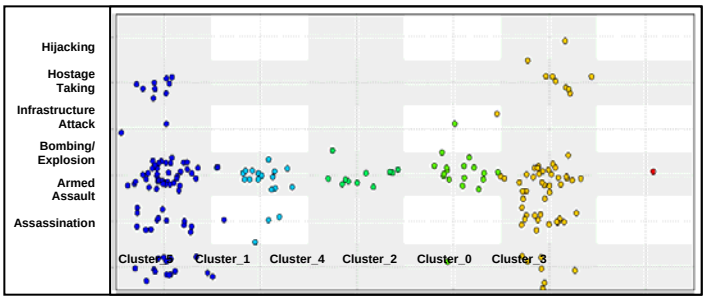

Figure 5. K-Means Clustering using *RapidMiner*

6.4 ONERULE

This next step will call the clusters individually and the OneR classifier is applied. The algorithm generates one rule for every value of the attribute and selects the best rules. For this step we used the WEKA data mining tool using as input the results of the previous clustering operation and applied individually to each of the cluster's subset. The goal of this step is to find rules that identify terrorist groups and target type within the clusters. Figure 6 shows the generated rules using OneR in cluster_0. We then selected only the terrorist groups that contain the military target type, as shown in Table 2.

Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.rules.OneR -B 6
Relation:     cluster0
Instances:    60
Attributes:   4
              Attack Type
              Target Type
              Terrorist Group
              Weapon Type
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Terrorist Group:
Mujahedeen Brigades     -> Private Citizens & Property
Tawhid and Jihad        -> Maritime
Jaish al-Ta     -> Business
Al-Qa'ida       -> Private Citizens & Property
Ansar al-Sunna  -> Government (General)
Takfir wal-Hijra (Excommunication and Exodus)   -> Military
Brigades of Iman Hassan-al-Basri        -> Journalists & Media
Factions of the Mujahedeen Army -> Other
Ansar al-Tahwid wal Sunna       -> Business
Mujahedeen Shura Council        -> Police
Islamic Army in Iraq (al-Jaish al-Islami fi al-Iraq) -> Military
Anbar Salvation Council -> Terrorists
Ansar al-Islam  -> Police
Islamic State of Iraq (ISI)     -> Police
Diyala Salvation Council        -> Terrorists
Kurdistan Free Life Party       -> Military
Kurdish Democratic Party-Iraq (KDP)     -> Military
```

Figure 6. OneR using *Weka*

6.5 APRIORI

In this step of our analysis we will initiate a through review of each cluster by applying Apriori to generate association rules. By default, APRIORI tries to generate ten rules. It begins with a minimum support of 100% of

the data items and decreases this in steps of 5% until there are at least ten rules with the required minimum confidence, or until the support has reached a lower bound of 10%, whichever occurs first. Also, the minimum confidence is set to 0.9. The goal of this step is use different attributes to complement information gained in the previous step. We applied this step only to clusters where military target types were found. As an example, Table 1 presents association rules results for cluster_0 using Apriori with the *Weka* tool.

6.6 EVALUATION AND PRESENTATION

Now that our data mining techniques have been applied, it is possible to analyze and explore the data representations to assess practical and relevant information for understanding the terrorist groups. Results are presented below.

Table 1. Association Rules

1	Terrorist Group=Al-Qa`ida Weapon Type=Explosives/Bombs/Dynamite 17 ==> Attack Type=Bombing/Explosion 17 conf:(1)
2	Target Type=Private Citizens & Property Weapon Type=Explosives/Bombs/Dynamite 7 ==> Attack Type=Bombing/Explosion 7 conf:(1)
3	Attack Type=Bombing/Explosion Target Type=Private Citizens & Property 7 ==> Weapon Type=Explosives/Bombs/Dynamite 7 conf:(1)
4	Target Type=Police Weapon Type=Explosives/Bombs/Dynamite 7 ==> Attack Type=Bombing/Explosion 7 conf:(1)
5	Attack Type=Bombing/Explosion Target Type=Police 7 ==> Weapon Type=Explosives/Bombs/Dynamite 7 conf:(1)
6	Target Type=Military Weapon Type=Explosives/Bombs/Dynamite 7 ==> Attack Type=Bombing/Explosion 7 conf:(1)
7	Target Type=Military Terrorist Group=Al-Qa`ida Weapon Type=Explosives/Bombs/Dynamite 6 ==> Attack Type=Bombing/Explosion 6 conf:(1)
8	Weapon Type=Explosives/Bombs/Dynamite 27 ==> Attack Type=Bombing/Explosion 26 conf:(0.96)
9	Attack Type=Bombing/Explosion 27 ==> Weapon Type=Explosives/Bombs/Dynamite 26 conf:(0.96)
10	Attack Type=Bombing/Explosion Terrorist Group=Al-Qa`ida 18 ==> Weapon Type=Explosives/Bombs/Dynamite 17 conf:(0.94)

7. RESULTS

The proposed analysis structure is used along with various data mining techniques and tools. The K-means clustering technique grouped the dataset with possible attack patterns and served as a foundation for organizing the data. Figure 7 shows the distributed data assigned to each cluster.

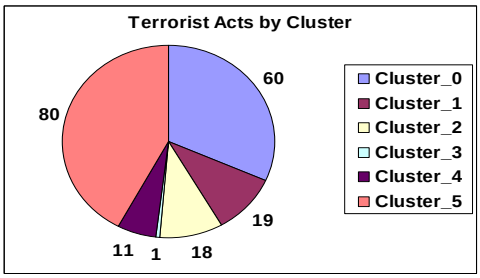

Figure 7. Terrorist Acts by Clusters Chart

For each cluster, we implemented the OneR classification technique which in return generated one attack target for each terrorist group within the clusters. The results in Table 2 present terrorist groups with military targets capability and the membership cluster.

Using the Apriori algorithm we generated association rules. The association rules can serve as complement for determining attack type and weapon type the identified terrorist groups are most likely to use in an attack, based on the information of historic data. The rules that are redundant and irrelevant have been discarded in order to maintain the most coherent. Therefore an analyst intervention is required to inspect the association rules that will give additional information necessary for describing the selected terrorist groups. This can be achieved by

combining the analyst intuition and the evaluation of the rule's support and confidence metrics. Table 3 describes the generated association rules selected for the identified terrorist groups.

Table 2. Terrorist Groups Targeting Military

Terrorist Group	Target	Cluster
Takfir wal-Hijra (Excommunication and Exodus)	Military	cluster_0
Islamic Army in Iraq (al-Jaish al-Islami fi al-Iraq)	Military	cluster_0
Kurdistan Free Life Party	Military	cluster_0
Kurdish Democratic Party-Iraq (KDP)	Military	cluster_0
Ansar al-Sunna	Military	cluster_2
Al-Qa`ida	Military	cluster_5

These results generate information that describes the terrorist groups that are capable of attacking military installations and interfere with the armed forces operations. We validated our results by confirming the terrorist groups listed with the START Terrorist Organization Profile (TOPs) database. The TOPs database has descriptive and verifiable information of each terrorist group and their founding philosophies. The identified terrorist groups are all described as highly hostile. Also, these terrorist groups have been confirmed to have attacked or that are capable of attacking military targets, even though they are classified as religious and ideological terrorist groups. By adding the geographical coordinates of the city locations where each terrorist incident occurred we can plot and export the information to geospatial applications. This will enable a visual representation that allows analyst to better understand the data, track mobilizations of terrorist groups and perform further investigations. Figure 8 shows an example of data exports to geospatial applications.

Table 3. Association Rules

Rule
Target Type=Military Weapon Type=Explosives/Bombs/Dynamite 7 ==> Attack Type=Bombing/Explosion 7 conf:(1)
Target Type=Military 2 ==> Attack Type=Bombing/Explosion 2 conf:(1)
Attack Type=Assassination 15 ==> Weapon Type=Firearms 15 conf:(1)
Weapon Type=Explosives/Bombs/Dynamite 40 ==> Attack Type=Bombing/Explosion 38 conf:(0.95)

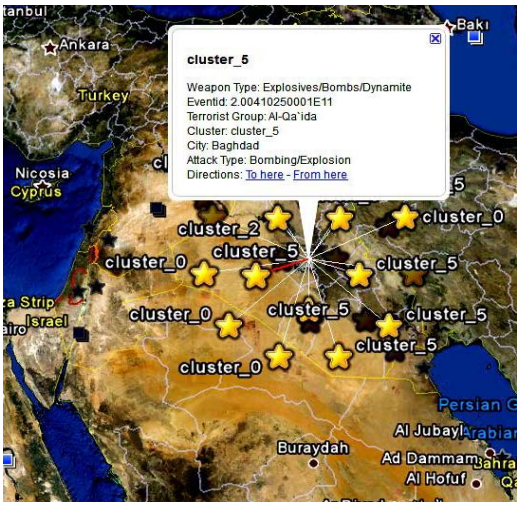

Figure 8. Geospatial Visualization

8. CONCLUSIONS

We looked at the use of data mining for identifying and classifying terrorist groups using various data mining tools and techniques. The scope of this project is to aid in the planning of military allocation to counter-strike terrorists, our goal is to gain all possible information from the terrorist groups that are capable of attacking military installations and interfere with the armed forces operations. Armed with the results of each of our data mining techniques, we can raise our insights on terrorists operations and provide security analysts with the information needed to put into action intervention strategies.

From a military perspective, neutralizing the terrorist groups with military attack capabilities should be a main priority. Military capable terrorist groups are known to often affiliate with other groups and are capable of providing training and undergo large scale operations. But, as it is known, this can only be achieved by gaining knowledge and in depth understanding on terrorist group mode of operation. It must be emphasized that this paper uses the simplest data mining analysis procedures since; the main objective of this paper is to demonstrate the advantages of using straightforward data mining techniques for military strategy and resource allocation.

Future collections of terrorism data should be centralized and made available for researchers and the private sector willing to invest and present emerging technologies. Perhaps, the most interesting observation is that terrorist groups are classified by what they are said to be in terms of the reasons for their actions, ethnic beliefs and political views, but not for their actions themselves. As a recommendation, counter-terrorist and law enforcement agencies should use all available means to establish a method that enables reclassification of terrorist groups based on collected data and ongoing analysis. This approach can greatly contribute in the effort of prioritizing counter-strikes without underestimating terrorist group capabilities.

REFERENCES

Agrawal, R. and Srikant, R. (1994). Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., 487-499.

Chen, H. (2007). Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security, New York, Springer Verlag.

Cordesman, A.H. (2003). The Iraq War, Greenwood Publishing Group.

Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition, Boston: Academic Press.

Global Terrorism Database, START: CD-ROM.

Holte, R.C. (1993). "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets", Machine Learning, vol. 11, 63-90.

Lia, B. (2005), Globalization and the Future of Terrorism, New York NY: Routledge.

National Consortium for the Study of Terrorism and Responses to Terrorism (START), "Terrorist Organization Profiles," http://www.start.umd.edu/start.

Popp, R.L. and Yen, J. (2006). Emergent Information Technologies and Enabling Policies for Counter Terrorism, John Wiley and Sons.

Stout, C.E. (2002). The Psychology of Terrorism: Theoretical Understandings and Perspectives, Greenwood Publishing Group.

Thuraisingham, B. (2002). "Data Mining, National Security, Privacy and Civil Liberties", ACM SIGKDD Explorations Newsletter, vol. 4, 1-5.

Witten, I.H. and Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann.

Authorization and Disclaimer

Authors authorize LACCEI to publish the paper in the conference proceedings. Neither LACCEI nor the editors are responsible either for the content or for the implications of what is expressed in the paper.