Proceedings of the 2016 Industrial and Systems Engineering Research Conference
H. Yang, Z. Kong, and MD Sarder, eds.

A Network-based Approach for Understanding Suicide Attack Behavior

Salih Tutun
Department of Systems Science and Industrial Engineering
Turkish Military Academy, Ankara, Turkey, and Binghamton University, Binghamton, NY

Sina Khanmohammadi and Chun-An Chou
Department of Systems Science and Industrial Engineering
Binghamton University, Binghamton, NY

Abstract

Terrorists are increasingly using suicide attacks to attack different targets. The government finds it challenging to track these attacks since the terrorists have learned from experience to avoid unsecured communications such as social media. Therefore, we propose a new approach that will predict the characteristics of future suicide attacks by analyzing the relationship between past attacks. The proposed approach first identifies relevant features using a graph-based feature selection (GBFS) method, then calculates the relationship between selected features via a new similarity measure capable of handling both categorical and numerical features. The proposed approach was tested using a second terrorism data set; we were able to successfully predict the characteristics of this new testing data set using patterns extracted from the original data set. The results could potentially enable law enforcement agencies to propose reactive strategies.

Keywords
Terrorism Networks, Suicide Attacks, Link Formation, Similarity Function, Feature Selection

1. Introduction

1.1 Background and Motivations

Crisis and chaos has been increasing around the world. Terrorist groups are using ever more complex tactics and strategies, which are not easily recognizable. In this regards, predicting suicide attacks, which encompass high uncertainty is almost impossible. The uncertain nature of terrorism is one of the main challenges in the design of counter-terrorism policies [1]. Nevertheless, recent studies have shown that there are some recognizable patterns in suicide attacks [2]. In this regards, counter-terrorism agencies have begun mining social media and telecommunication data to identify the intention behind suicide attacks, then using this data to predict future attacks. However, terrorist groups are becoming aware of such counter-terrorism tactics and rarely use social media and other unsecured communication techniques.

In this paper, we propose a network model to analyze the interaction between various suicide attacks. This research is an early attempt to identify meaningful patterns in suicide attacks using network models. Network models have been proven to be a valuable tool for understanding terrorism [3]. However, most researchers have focused on understanding the behavior of individual terrorists by modeling their relationship with each other within a particular terrorist group, hoping that such information can provide insights about the leader of that specific terrorist group. In this research, we focus on the spatial and tactical relationship of different attacks rather than the connection of individuals within a terrorist group. Furthermore, current network-based approaches in the literature concentrate more on prosecution than on prevention [4–6]. However, the ultimate goal of counter-terrorism agencies is preventing terrorist attacks.

Hence, in this study, we have incorporated data mining techniques with network analysis methods to predict future suicide attacks in Iraq. The results show that the proposed approach was able to successfully predict future attacks in

our testing data set. The results of this study could potentially help with detection and prediction of suicide attacks and enable law enforcement agencies to propose reactive strategies that reduce casualties, as well as financial and political losses. The rest of the paper is organized as follows. Section 2 provides the details of the proposed approach. In section 3, we test the proposed method on a case study of suicide attacks in Iraq. Finally, we provide a brief conclusion.

2. Materials and Methods

The proposed method consists of three main steps including feature selection, similarity calculation, and network analysis. In this section, we first describe the data sets and how they were preprocessed, followed by details of each section of the proposed approach.

2.1 Data Preprocessing

The National Consortium for the study of terrorism and responses to terrorism (START) is employed in this study [7]. The data set includes historical incidents of domestic as in the USA, and international terrorism. Data preprocessing is performed to remove the missing values and non-terrorist attacks. Considering the large size of the data set, we have selected as our case study only those suicide attacks occurring between 2003-2014 in Iraq. The data set from 2003-2013 is used for understanding the patterns of suicide attacks, and then the proposed approach is tested using data from 2014.

2.2 Graph-based Feature Selection (GBFS)

Feature selection is one of the most important steps in data mining. The goal of feature selection is to select a subset of relevant features that could provide useful information. Feature selection is especially important in terrorist activity analysis, where we want to focus only on features that influence the attacks. Feature selection is a combinatorial optimization problem that is computationally expensive. Hence, most of the feature selection methods analyze the features independently [8]. In this paper, we propose a graph-based feature selection (GBFS) that considers the effect of features on each other. The proposed feature selection approach consists of five main steps as follows:

Step 1: First, import the terrorist attack data including the number of the attacks (defining nodes in the network) and the number of features.
Step 2: Let $I = \{i_1, i_2, ..., i_n\}$ be a set of binary values of features, and $T = \{t_1, t_2, ..., t_m\}$ be a set of attacks. Associated rules of features are found by using the Apriori algorithm (classic associate rule algorithm). The similarity matrix is computed based on the support and confidence between feature vectors.
Step 3: Networks are formed based on 1000 rules as relations for features.
Step 4: From the constructed network, scores for each feature are calculated by using node centrality. Next, the rank (n_{top} features) is found according to the scores with high predictive ability.
Step 5: Lastly, we use selected features for classification to determine predictive ability in terms of precision and sensitivity. Therefore, reduced features are checked for validation. In this step, we use selected features as input for classification. In our experiment, we use a logistic regression classifier for classification. Figure 1 illustrates the five steps of GBFS feature selection. The selected features are used to calculate the similarity function of suicide attacks in the following steps of the approach.

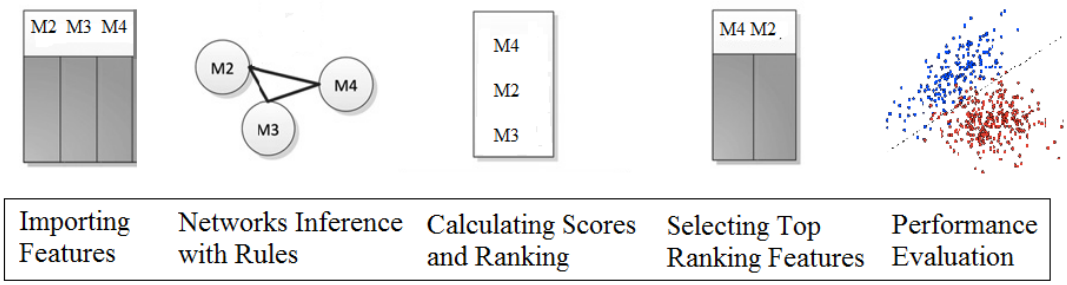

Figure 1: Graph-based feature selection

2.3 Similarity Function

After selecting the relevant features for our analysis, we need to construct a network using the selected features. For this purpose, first we need to calculate the similarity between nodes (suicide attacks). However, since our data set includes both categorical and numerical features, we cannot directly use the traditional approaches. In this regard, we propose a new heterogeneous similarity function to solve this problem. For categorical features, b_h is used by looking at binary similarity. Ratio value for numeric features is used to define similarity. Moreover, this data set has d categorical and continuous features where F_h denotes the h^{th} feature. Let the feature F_h take n_h values in the data set. The importance of features (weights) are found by using Equation (1). Features influence is checked on the success rate(as a class). After finding weights from logistic regression methods, the weights are normalized to use in Equation (2). Therefore, w_h is used (from Equation (1)) to give significance rate of the features. Finally, we combine similarities for categorical and continuous features to find final similarity value for links between suicide attacks, as seen in Equation (2).

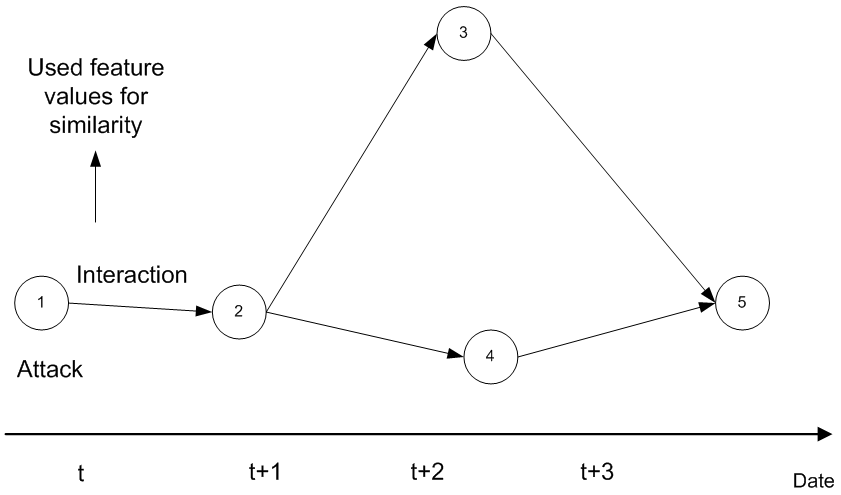

Figure 2: Link formation among nodes (terrorist attacks).

$$F_{Linear} = \frac{1}{1+e^{-(w_0+\sum_{h=1}^{d} w_h x_h)}} \tag{1}$$

The similarity value between X and Y belonging to the data set is as follows:

$$b_h = \left\{ \begin{array}{ll} 1 & if \;\; X_h = Y_h \; as \; categorical \; features \\ 0 & otherwise \end{array} \right\}$$

$$n_h = \left\{ \begin{array}{ll} (X_h/Y_h) & if \; X < Y \; as \; continues \; features \\ (Y_h/X_h) & otherwise \end{array} \right\}$$

$$S_h(X_h,Y_h) = \sum_{h=1}^{d} w_h \; (\sqrt{(b_h)^2} \; or \; (n_h)) \tag{2}$$

where $S_h(X_h,Y_h)$ is the similarity between two values for feature F_h and w_h is the weight of feature h.

2.4 Network Analysis

After calculating the similarity measure between various features of the data, the network $G = (V,E)$ is constructed, where V is the vertices of the network showing different incidents, and E is the calculated similarity measure between incidents. Two nodes are connected to each other if the similarity measure between them exceeds 80%. We use a directed graph because our network also has a temporal dimension showing the time of the attack. If we have a directed

graph, the $in-degree$ of node v is the number of nodes with v (terminal node). At the same time, the $out-degree$ of node v is the number of nodes with v (internal node) [9].

The in degree and out degree represent total experience and individual experience for attacks, respectively. This information can be used to understand how tactics change for suicide attacks. By looking in-degree and out-degree for every two years, we can capture patterns that show future tactics.

3. Results and Discussion

In this section, we present the experimental results using the START data set [7]. Next we will discuss how the results can be used to find patterns for predicting future terrorist attacks [10]. First, to validate the proposed GBFS feature selection method, we applied it to the START data set and used logistic regression to classify whether the attack was successful or not. By calculating attractions for each attack, we calculate the weights for each feature. In order to define relations between terrorist attacks, the approach needs to use the relevant features. As can be seen in Table 1, the proposed method was able to detect the relevant features, and the classification results were almost identical to using all the features in the data set.

Table 1: Comparing the results from the feature selection method and all features with Logistic Regression (LR). It shows that some of the features are redundant for the similarity function.

Methods with Logistic Regression (LR)	**Number of Features**	**Accuracy**	**Precision**	**Recall**	**F-Measure**
All features (without feature selection)	23	95.38%	0.94	0.95	0.95
GBFS	6	95.16%	0.94	0.95	0.94

Table 2: Suicide attacks being the highest in-degree and out-degree dynamically. Note: The numbers of the nodes show the attacks, and values in nodes are increasing from right to left in the table.

Years	**In-degree**	**Out-degree**
2003-2005	52-46-43-44-95-96	6-3-17-21-20-24-16-11-8-13
2004-2006	46-52-43-44-55-80-39	17-21-22-18-20-24-11-16-13-8
2005-2007	43-44-80-84-90-108-62-71-114	17-21-22-18-20-24
2006-2008	43-44-80-84-62-95-71-96	31-36-33-37-42-39-43-44
2007-2009	80-84-90-108	31-32-33-36-42-43-44-38-39
2008-2010	81-77	43-44
2009-2011	77-81-108-91	46-49-52-55
2010-2012	96-114-113-95-81-77	80-62-71
2011-2013	114-96-113	80-62-63-71
2013-2015	**80-84-90-108**	

Next we applied the proposed similarity measure to the selected features to construct the network of suicide attacks. The results of analyzing the in-degree and out-degree properties of the constructed network (shown in Table 2), expose a repeating pattern where the tactics of attacks with a high out-degree are repeated after some period. For example, the attacks with a high out-degree at during 2010-2013 are repeated in the 2007-2009 period as in-degree. This means that they learn individual experience for future attacks. After finding the patterns for 2013-2015, we compared these patterns and real data for sensitive analysis to show that the proposed approach works. Figure 3 shows the identified patterns for the 2013-2015 period. The finding patterns will be used as suicide attacks in the future.

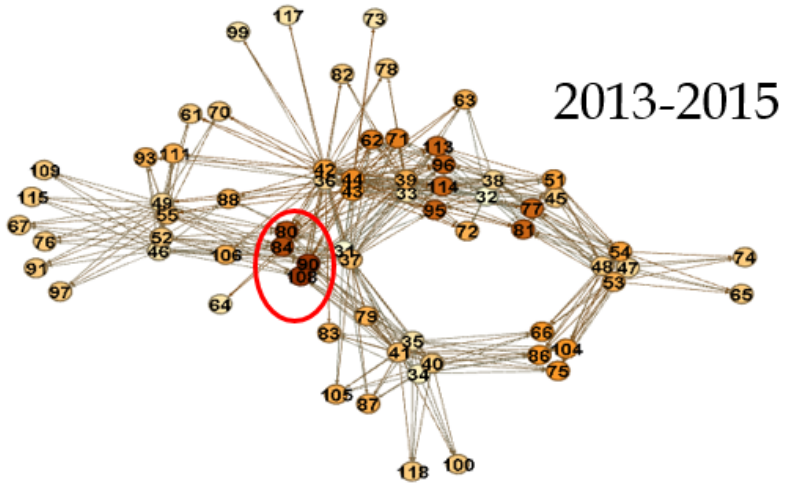

Figure 3: The finding patterns with networks analysis.

Table 3: Comparing finding patterns and occurred attacks for the year of 2014

Features	**Predicted Feature Value**	**Accuracy**
Weapontype	Explosive bombs and dynamite	100%
Geocoding Specificity	Attacks will occur in city/village/town	96%
Multiple	Attacks are multiple	88%
Group Name	ISIS and Al-Qaede	100%
Extented Incident	The duration of attacks is less than 24 hours	100%
Vicinity	Attacks will occur in city centers	100%
Goals of attacks	Attacks has political, economic, religious with larger audiences	100%
Target Type	Government (Politician, political party movement, head of state, government personel, government building)	96%
International-domestic	Attacks are domestic and unknown	100%
Hostages-kidnaping	No kidnapping or hostages	92%
Total number of fatalities	number of killing is less than 10 killing and 20 wounded	96%
Extent of property damage	Minor (likely 1 million dolars)	52%

To validate the predictive capabilities of our approach, we used the identified pattern in the suicide attack network to predict the features of attacks in 2014. As it can be seen in Table 3, comparing the results with existing data shows that our proposed method was able to successfully predict most of the characteristics of suicide attacks with more than 90% accuracy. The only exception was the extent of the property damage, which was accurate in 52% of the attacks. These results support the previous findings that terrorists tend to emulate the behavior of other terrorist groups and learn from their mistakes and successes [11]. As a results, terrorists will learn tactics from the years 2007-2009 to attack Iraq in the years 2013-2015. As seen in Table 3, in the these years, number of killing people will be less than 10 people, and number of wounded people will be less than 20 people for suicide terrorism. They will attack less than 24 hours. Multiple attacks for each time will be used for suicide terrorism. Attacks will occur in city centers. Politician, political party movement, head of state, government personnel and government building will be mostly attacked for this terrorism. Governments needs to be careful for ISIS and Al-Qaede.

4. Conclusion

Preventing terrorist threats before they happen is the ultimate goal of every counter-terrorism agency. However, to predict future attacks we need to understand how terrorist groups behave and what is the relationship between different

terror attacks. In this regards, we have proposed a approach to analyze the suicide attacks in Iraq to understand how different attacks are related to each other. We showed that using the extracted patterns; we can predict various attributes of future attacks. These results can be used by policy makers to develop precise global and/or local counter-terrorism policies that target the predicted attributes of attacks to reduce casualties, as well as financial and political losses. Furthermore, this information can be very useful for law enforcement agencies to propose reactive strategies.

5. Acknowledgments

The research is supported by the Turkish Military Academy (TMA). The authors wish to thank the Global Terrorism Database and the Turkish Military Academy for their help in providing data and supporting research.

References

1. Jackson, B. A., and Frelinger, D. R., 2009, "Understanding Why Terrorist Operations Succeed or Fail," Technical Report, DTIC Document.
2. Khan, A. N., 2010, "Strategic Response to Suicide Terrorism in Pakistan," retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.357.4694
3. Top, N. M., 2009, "Counterterrorism's New Tool:'Metanetwork' Analysis," retrieved from http://www.johnbohannon.org/NewFiles/mosaic.pdf.
4. Carley, K. M., 2005, "Dynamic Network Analysis for Counter-Terrorism," retrieved from http://www.nap.edu/read/12083/chapter/6
5. Krebs, V. E., 2002, "Mapping networks of terrorist cells," Connections, 24(3), 43–52.
6. Xu, J. J., and Chen, H., 2005, "CrimeNet Explorer: A Framework for Criminal Network Knowledge Discovery," ACM Transactions on Information Systems (TOIS), 23(2), 201–226.
7. National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2016, "Global Terrorism Database," retrieved from http:/www.start.umd.edu/gtd
8. Zhang, Z., and Hancock, E. R., 2011, "A Graph-based Approach to Feature Selection," Graph-Based Representations in Pattern Recognition, Springer, 205–214.
9. Barabasi, A. L., and Frangos, J., 2014, "Linked: the New Science of Networks Science of Networks,". Basic Books, Perseus Publishing, Philadelphia, Pennsylvania.
10. Shang, R., Zhang, Z., Jiao, L., Liu, C., and Li, Y., 2016, "Self-representation Based Dual-graph Regularized Feature Selection Clustering," Neurocomputing, 171, 1242–1253.
11. Chenoweth, E., and Lowham, E., 2007, "On Classifying Terrorism: A Potential Contribution of Cluster Analysis for Academics and Policy-makers," Defence & Security Analysis, 23(4), 345–357.