Analyzing Behavior of ISIS and Al-Qaeda using Association Rule Mining

Tamanna Goyal, Jaspal Kaur Saini and Divya Bansal

Abstract Social media have been exploited by terrorist groups to share their massacres plans, recruitment, buying weapons, or propagating their violent plans. Terrorist groups named ISIS and Al-Qaeda are the most active and well known for using social media to propagate their violent intents over online discussion forums. It becomes necessary to study the behavior of these terrorist groups over online social media. In this paper, we present association rule mining based approach to extract a feature set for terroristic groups named ISIS and Al-Qaeda. We used the Global Terrorism Dataset which contains systematic information on terrorist attacks worldwide since 1970. Entropy-based feature extraction technique is used to extract top features which are then further used to find association rules. Eclat (Equivalence Class Transformation) and Apriori algorithms are used to mine association rules from prepared data. Rules for ISIS and Al-Qaeda are computed separately, and are then further classified using machine learning classification algorithms. Our research contributes to the smart and novel application of data mining algorithms and computational intelligence to study the behavior of the most popular and active terrorist groups over social media.

1 Introduction

Recent years have witnessed exploitation of online social media by terrorist groups in many ways. Networking, sharing information, planning, recruitment, and mobilization are set of terroristic activities which have done over online social media. The Irish Republican Army, Naxalites, Boko Haram, Hezbollah, ISIS (Islamic States of

T. Goyal · J. K. Saini (✉) · D. Bansal
Punjab Engineering College, Chandigarh, India
e-mail: sainijassi87@gmail.com

T. Goyal
e-mail: goyaltamanna161@gmail.com

D. Bansal
e-mail: divya@pec.edu.in

© Springer Nature Singapore Pte Ltd. 2019

C. R. Krishna et al. (eds.), *Proceedings of 2nd International Conference on Communication, Computing and Networking*, Lecture Notes in Networks and Systems 46, https://doi.org/10.1007/978-981-13-1217-5_66

Iraq and Syria), Tehrik-i-Taliban Pakistan, Lashkar-e-Taeba, Taliban, and Al-Qaeda are terrorist groups which are getting active over online social networks such as Twitter or dedicated dark web forums [1, 2]. ISIS and Al-Qaeda are two most popularly known terrorist groups which are well known for using social media to propagate their agenda and violent activities. It becomes a challenging task for security agencies to explore and mine the behavioral features of these terrorist groups. Several research organizations have taken charge to examine trends and patterns of terrorists to study behavioral models of terrorists.

Figure 1a shows the word cloud of top 50 terrorist groups as reported by Global Terrorism Dataset (GTD). GTD is open source dataset which contains information about various terroristic incidents in a form of 134 attributes [3]. The font size of each terrorist group name is proportional to number of attacks done by terrorist group as reported by GTD. It is evident from the figure that Boko Haram, Al-Qaeda, ISIS, and Al-Shabaab are the most active terrorist group since 1970, which in terms is an open research challenge to researchers in the field to study their behavior and attack patterns. Figure 1b depicts the number of attacks done by ISIS and Al-Qaeda as per reported by GTD. It is seen that a number of attacks are increasing exponentially over years which opens an era for researchers to study the behavior and analyze the patterns of attacks by ISIS and Al-Qaeda in order to counter terrorism.

In this paper, we analyzed the behavior of two terrorist groups, i.e., ISIS and Al-Qaeda, by understanding the patterns of attacks done by both the groups in different countries. The data of these groups is compiled and preprocessed from GTD and analyzed further. We applied Apriori and Eclat algorithms to mine the association rules and classified these rules using naive Bayes, SVM, and decision tree algorithms to make predictions for the future. This paper is further organized as follows: Sect. 2 elaborates literature survey in the same domain. Section 3 explains the detailed proposed methodology. Experimental results are illustrated in Sect. 4, and finally, Sect. 5 concludes the paper and lists the future scope of the work done.

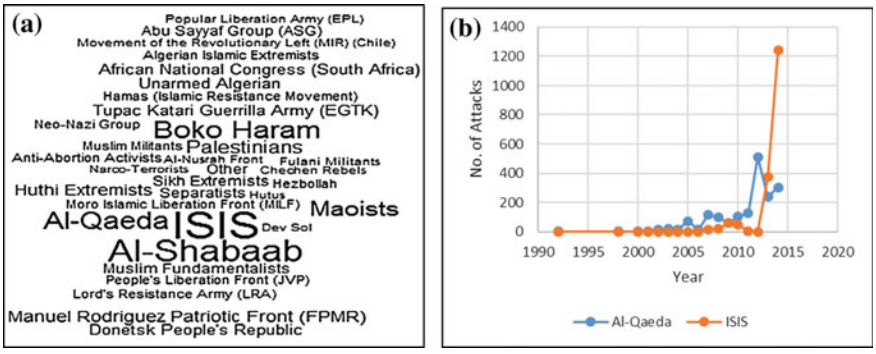

Fig. 1 **a** Top 50 Terrorist Groups Worldwide as reported by GTD **b** No. of Attacks since 1990 by ISIS and Al-Qaeda as reported by GTD

2 Related Work

Computational techniques from machine learning, data mining, and social network analysis are studied in literature to highlight which techniques are used by researchers to study behavior of terrorist groups [4–11]. Literature survey is divided into two parts. First, we studied what are techniques used in all research evidences as listed in Table 1. Second, we looked at which terroristic organizations have been targeted by research agencies. Table 2 describes work done on various terrorist groups.

Table 1 Different techniques used for the study of different terrorist groups

Technique used	Citation
Data preprocessing techniques, data mining, and classifiers	[12]
COSM (Crime ontology similarity measure)	[13, 14]
Natural Language Processing (NLP), event clustering, event trending, and narrative generation	[15]
Stochastic Opponent Modeling Agents (SOMA)	[16, 17]
Naive Bayes and Apriori algorithm	[18]
Class Association Rule mining (CPAR and CMAR)	[19]
Classification techniques (Naive Bayes, SVM, decision tree, and multilayer perceptron)	[20]
Data mining techniques, Classifiers—Naive Bayes, SVM, and AdaBoost	[21]
Clustering techniques	[22]
Logistic regression, Support Vector Machines (SVM), boosting, naive Bayes models, and classification trees	[23]
STONE	[24]
Rule mining techniques	[25]

Table 2 Research work done on different terrorist groups

Terrorist group	Work Done	Citation
Lashkar-e-Taiba (LeT)	LeTs activities, behavior, and environment are studied in Pakistan and J&K using SOMA	[4, 16, 17]
ISIS	Analyzed and classified the tweets of ISIS using machine learning	[5, 10, 21]
Al-Qaeda	The magazines produced by Al-Qaeda are examined to estimate the similarity and difference among ISIS, Al-Qaeda, and Taliban	[5]
Terrorist groups of Istanbul	On the basis of criminology perspective, COSM (ontology-based similarity measure) is used to classify them into three groups: extreme left, separatism, and extreme right groups	[13]
Terrorist groups of Turkey	COSM is used to classify the terrorist network on the basis of their similarity	[14]
Hezbollah	SOMA is used to predict the behavioral model of this terrorist behavior	[17]

Related work done in the field illustrates that there is no prior work done to extract features from GTD about popular terrorist groups ISIS and Al-Qaeda, and hence, demands to extract a feature set of these violent extremist groups which can assist any security agency worldwide to perform behavioral rule-based predictions about future attacks.

3 Proposed Methodology

For a given item set, association rule mining algorithms can be utilized to discover frequent patterns and draw inferences based on the discovered patterns. We used Eclat and Apriori algorithms which are described further. Eclat (Equivalence Class Transformation) works on the basis of a depth-first search, whereas Apriori works on the basis of a breadth-first search. Eclat uses simple intersection operations for equivalence class clustering along with bottom-up lattice traversal to mine frequent itemsets and works well for smaller datasets. Apriori counts the support of itemsets and uses a candidate generation function which exploits the downward closure property of support and works well for large datasets. Our proposed behavioral rule-based prediction algorithm can be described step by step:

1. Extract and prepare data about ISIS and Al-Qaeda from Global Terrorism Dataset.
2. Extract useful features from this data using entropy and information gain as mentioned below:

$$Entropy(set) = \sum_{i=1}^{k} -P(value_i) log_2 P(value_i) \tag{1}$$

   where P(value)$_i$ is the probability of getting the ith value when randomly selecting one from the set.
3. Association rules are calculated with extracted features using Apriori and Eclat algorithms by setting different support and confidence levels.
4. Classify the rules using different classification algorithms.
5. Compute different performance metrics accuracy, sensitivity, specificity, etc.
6. Verify and validate the results.

4 Experimental Results

By using information gain and entropy, 37 attributes are selected out of 134 attributes from Global Terrorism Dataset. The top seven out of 37 variables that contributed the most are as follows:

Table 3 Sample extracted rules using Apriori algorithm

Rule No	Rule (lhs=> rhs)
1	attacktype1_txt=Bombing/Explosion=> country_txt=Iraq
2	weaptype1_txt=Explosives/Bombs/Dynamite, weapsubtype1_txt=Vehicle=>attacktype1_txt= Bombing/Explosion
3	attacktype1_txt=Bombing/Explosion=> weaptype1_txt=Explosives/Bombs /Dynamite
4	targtype1_txt=Military,natlty1_txt=Yemen =>country_txt=Yemen
5	attacktype1_txt=Bombing/Explosion, natlty1_txt=Iraq=>weaptype1_txt =Explosives /Bombs/Dynamite

1. Target1: Target of incident like military, public place, government, transportation, etc.
2. City: It defines name of the city, village, or town in which the incident occurred.
3. Location: Information about location of incident.
4. Weapsubtype1_txt: Type of weapon used for the attack like bomb, firearms, biological, etc.
5. Nalty1_txt: Nationality of target of attack.
6. Country_txt: Name of country attacked.
7. Attacktype1_txt: Type of attack like assassination, hijacking, kidnapping, bombing/ explosion, etc.

Feature names used above are same as used in Global Terrorism Dataset Codebook [3]. The rules are extracted by Apriori algorithm by considering different supports and confidences. Sample extracted rules are shown in Table 3.

Rule (1): When the type of attack is bombing or explosion, then the country is Iraq.

Rule (2): When the main weapons used are explosives, bombs, or dynamite and a vehicle is used as another weapon, then attack tends to be bombing or explosion.

Rule (3): When a bombing or explosion is used for attack, then the main weapon used for attack would be explosives, bombs, or dynamite.

Rule (4): When the target is military and nationality of attack is Yemen, then country would be Yemen.

Rule (5): When the type of attack is bombing or explosion and nationality of attack is Iraq, then weapon used for the attack would be explosion, bombs, or dynamite.

After associative rule mining, these rules are classified using different classifiers including Naive Bayes, SVM, and decision tree. In the classification method, to construct a model, a training set is required which comprises a set of attributes with one attribute being the attribute of the class. Thus, a random sample of one-third of the dataset of rules is considered for training and the trained model is used for classification on the basis of the terrorist group name and by using this trained model the instance is classified. The remainder of the dataset is used for testing the model.

Table 4 Performance measures

S. No.	Performance measure	Naive Bayes	Decision tree	SVM
1	Specificity	0.98	0.95	0.99
2	Sensitivity	0.96	0.80	0.99
3	Positive predictive value	0.95	0.63	0.99
4	Negative predictive value	0.99	0.98	0.99
5	Kappa measure	0.91	0.69	0.98
6	Accuracy	0.98	0.93	0.99

We used R programming to perform all experiment. The output of these classifiers is compared on the basis of different parameters, as shown in Table 4.

It can be inferred from Table 4 that SVM emerged as the best classifier for the classification of our association rules of ISIS and Al-Qaeda.

5 Conclusion

Social media emerges to give all of us a platform to share and discuss any information or opinions. But violent extremists have exploited these online social networks for their malicious intent. We proposed a novel approach to find associations among two popularly known terrorist groups: ISIS and Al-Qaeda. We further automated the process to identify whether attacks were done by ISIS or Al-Qaeda using machine learning classification algorithms. SVM is the best classifier for our case as compared to naive Bayes and decision tree. We discovered seven critical feature sets named as target, city, location, weapon type, nationality, country, and attack type description, to find association rules. It is seen that when type of attack is bombing or explosion and nationality of attack is Iraq, then the attacking group is ISIS, and when the main weapons used are explosives, bombs, or dynamite and vehicle are used as another weapon, then the attacking group is Al-Qaeda.

In future, this work can be extended to study more terrorist groups. Different features can be accommodated from social media also to study behavior of terrorist groups.

References

1. Accessed on 15 Aug 2017, http://listovative.com/top-10-most-dangerous-terrorist-organizations-in-the-world/
2. V.S. Subrahmanian ed., Handbook of Computational Approaches to Counterterrorism (Springer Science & Business Media, 2012)
3. Global Terrorism Database. Accessed 19 May 2016, https://www.start.umd.edu/gtd/
4. A. Mannes et al., A computationally-enabled analysis of Lashkar-e-Taiba attacks in Jammu and Kashmir. 2011 European Intelligence and Security Informatics Conference (EISIC) (IEEE, 2011)

5. T. Mahmood, K. Rohail, Analyzing terrorist events in Pakistan to support counter-terrorism-Events, methods and targets, in *2012 International Conference on Robotics and Artificial Intelligence (ICRAI)* (IEEE, 2012)
6. Jacob R. Scanlon, Matthew S. Gerber, Automatic detection of cyber-recruitment by violent extremists. Sec. Inf. **3**(1), 1–10 (2014)
7. F. Spezzano, V.S. Subrahmanian, A. Mannes, Reshaping terrorist networks. Commun. ACM **57**(8), 60–69 (2014)
8. P. Su et al., Mining actionable behavioral rules. Decis. Support Syst. **54.1**, 142–152 (2012)
9. U.K. Wiil, J. Gniadek, N. Memon, Measuring link importance in terrorist networks, in *2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, 2010)
10. D.B. Skillicorn, Empirical assessment of al qaeda, ISIS, and taliban propaganda, in *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)* (IEEE, 2015)
11. D.B. Skillicorn et al., Understanding South Asian Violent Extremist Group-group interactions, in *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, 2014)
12. WITS - Home Page. Accessed 19 May 2016, http://wits.worldbank.org/WITS/WITS/Default-A.aspx?Page=Default
13. International Terrorism: Attributes of Terrorist Events (ITERATE) | Duke University Libraries. Accessed 19 May 2016, http://library.duke.edu/data/collections/iterate
14. S. Ginsberg, 2012 Database Spotlight: Minorities at Risk Organizational Behavior (MAROB) | START.umd.edu. Accessed 19 May 2016, https://www.start.umd.edu/news/database-spotlight-minorities-risk-organizational-behavior-marob
15. E. Spertus, M. Sahami, O. Buyukkokten, Evaluating similarity measures: a large-scale study in the orkut social network, in *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD '05)* (ACM, New York, NY, USA, 2005), pp. 678–684
16. J Pagn, Improving the classification of terrorist attacks a study on data pre-processing for mining the Global Terrorism Database, in *2010 2nd International Conference on Software Technology and Engineering (ICSTE)*, vol. 1 (IEEE, 2010)
17. E. Serra, V.S. Subrahmanian, A survey of quantitative models of terror group behavior and an analysis of strategic disclosure of behavioral models. IEEE Trans. Comput. Soc. Syst. **1.1**, 66–88 (2014)
18. F. Ozgul, C. Atzenbeck, Z. Erdem, How much similar are terrorists networks of istanbul?, *2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (IEEE, 2011)
19. F. Ozgul et al., *Intelligence and Security Informatics*, Specific similarity measure for terrorist networks: how much similar are terrorist networks of Turkey? (Springer, Berlin, 2011), pp. 15–26
20. L. Guohui, et al. Study on correlation factors that influence terrorist attack fatalities using Global Terrorism Database. Procedia Eng. **84**, 698–707 (2014)
21. J. Lauten chlager, et al., Group Profiling Automation for Crime and Terrorism (GPACT). Procedia Manuf. **3**, 3933–3940 (2015)
22. A.R. Kulkarni, V. Tokekar, P. Kulkarni, Identifying context of text documents using Nave Bayes classification and Apriori association rule mining, in *2012 CSI Sixth International Conference on Software Engineering (CONSEG)* (IEEE, 2012)
23. Bangaru Veera Balaji, Vedula Venkateswara Rao, Improved classification based association rule mining. Int. J. Adv. Res. Comput. Commun. Eng. **2**(5), 2211–2221 (2013)
24. A. Al Deen, M. Nofal, S. Bani-Ahmad, Classification based on association-rule mining techniques: a general survey and empirical comparative evaluation. Ubiquitous Comput. Commun. J. **5.3** (2010)
25. M. Ashcroft et al., Detecting jihadist messages on twitter, in *EISIC 2015*, Manchester, UK. IEEE Computer Society (2015)
26. RAND Corporation, RAND HRS Data File (v.N) | RAND (1994). Accessed 19 May 2016, http://www.rand.org/labor/aging/dataprod/hrs-data.html