Science Can Solve Terrorism

Abstract

For Task One, we extract six features from internet usage data of monitored people. Additionally, we get the data vectorized and normalized to lay a good foundation for later quantitative analysis of models. Then, considering the fact that we may not obtain terrorists' internet usage data, we thus build a model with prior knowledge and a model without prior knowledge respectively based on the six features. And we evaluate the risk index of each monitored person based on Euclidean distance between features. Finally, we take the processing of missing data into consideration.

For Task Two, with an effective, fast and automatic clustering algorithm, we use unsupervised classification to categorize monitored people based on internet usage data. Then we evaluate risk index of each colony based on models built in task one.

For Task Three, we propose comprehensive solutions in four aspects. Firstly, we can identify people who have potential terrorism tendency based on risk index in previous calculation. We should monitor them and help them out of abnormal psychology to eliminate terrorism tendency from origin. Secondly, we can find out critical persons in terroristic organization based on the centrality in social network analysis. By striking these persons, we can reach the purpose of collapsing terroristic organization effectively. Thirdly, we can find out the 'off-ramps' by triad analysis and help them out of terrorism. Last but not least, we mine the rules of terrorism attacks by association rules analysis to take effective anti-terrorism measures. In addition, we validate our association model with practical data. In conclusion, we summarize the four measures above to give President Obama a thorough anti-terrorism proposal.

After finishing the three tasks, we discuss the robustness of our models for practical data. At last, we draw a conclusion that science can solve terrorism.

Key words: Risk index; Clustering; Social network analysis; Centrality; Traid; Association rules analysis

Contents

1 Introduction

1.1 What is terrorism

Terrorism is not new, it has been used since the early times of recorded history, which is hard to define correctly. In its broadest sense, terrorism is any act designed to cause terror [1]. In a narrower sense, terrorism can be understood to feature a political objective. The word terrorism is politically loaded and emotionally charged [2]. Since the 9/11 and the later 7/7 terrorist strikes in New York and London, respectively, the terrorism has caused the attention of people all over the world. Recently, the attack in Paris and the Islamic State group's killing of captives make the terrorism a hit topic again. The governments and security officials are making great efforts to protect the public against attacks, and hence the analysis of terrorism and terrorists has become a prior issue.

1.2 Why people turn to terrorism

As Amy Zalman Ph.D. says, all terrorist acts are motivated by two things [3]. One is the social and political injustice, which corresponds to the bigger push factor. People choose terrorism when they are trying to right what they perceive to be a social or political or historical wrong. The push factors include alienation, shared anger or outrage, frustration, disillusionment, a sense of victimization by the actions, or inactions of others. The other is the belief that violence or its threat will be effective, and usher in change. These are 'lures', which include the perceived benefits of turning, like adventure, excitement, camaraderie, a sense of belonging, being part of something far bigger etc.

1.3 How people turn to terrorism

How people turn to terrorism is a more meaningful question since we cannot say that the presence of one factor above provokes terrorism in the same way that we can say with certainty that certain toxins cause diseases. Recruiters will use whatever tools in order to pull someone in. The strategies include convincing them that it's their duty to go fight in defense of others, or involvement offers them a way out of the humiliation and victimization, which otherwise they will be destined to face at home.

Another consideration is how radicalization relates to recruitment. Some traits increase in severity, they become more advantageous for attracting more mates and even producing more offspring. This would characterize these traits as risky shortcuts to fitness, owing less to failures than to the twists and turns make by genes in order to perpetuate themselves [4].

2 Our Work

Some terrorists report a sense of suffocation-being unable to leave for fear of retaliation and being equally afraid of their disillusionment being detected by those close to them in the movement. Consequently, we need to do a better job of providing

rescue measures not only for people who are on the road to terrorism, but also to those who have gotten themselves in a jam and want to get out before it's too late. Psychological characteristics are key to this problem, which can be reflected through internet using habits. Without giving the content of internet usage, the mechanics of out usage-how often we email others, chat, online, stream media, or multi-task (switch from one application or website to another), can also predict psychological characteristics. According to these internet habits, we can predict users' psychological characteristics and do something about anti-terrorism.

2.1 Task One

Based on the internet usage, we build a model with prior knowledge and a model without prior knowledge respectively to evaluate the risk index of each monitored person using it. With prior knowledge of terrorists, we cluster terrorists' internet using habits by six indexes, average packets per flow, peer-to-peer usage, chatting, email, flow duration entropy and ftp and remote file usage. Then we can evaluate the risk index of the monitored person according to his similarity in the six indexes with the terrorists' standard.

Without the terrorists' internet usage data, we evaluate the risk index directly by the above six indexes. After ranking, we can predict who are likely to be terrorists, or who are likely to be on the road to terrorism.

2.2 Task Two

Experts use the expression big data to indicate huge amounts of information, such as those shared by billions of people on computers, smartphones and other electronic devices at any time. We'll get lots of monitoring data, but in order to use these huge amounts of data, we have to understand them. Thus we use a rapid clustering algorithm to categorize these big data in an effective, fast and automatic manner. After that, we can calculate each risk index of the centers of clustering using model one or model two. If having prior data of terrorists' internet usage, we can calculate the risk index of the clustering centers, and then categorize them by matching with classification standards in model one to predict whether they are terrorists or not.

2.3 Task Three

If President Obama asked for advice on fighting terrorism, what we will tell him is we need long-term investment in more scientific research. Our specific advice goes as follows:

- Based on Model One, Model Two and Model Three, we can make some predictions through psychological analysis and take some preventive measures to those predicted terrorists.
- Based on social network analysis, we can predict the possible leader in a terrorists' organization by calculating node centricity, whom we should spend more attention to monitor and strike.
- Based on triad model to find out those terrorists who want to get out, we should rescue them in time.

- Based on prior data, we can predict where attacks terrorists may take in the future for a period of time using association rules analysis. The government and security officials can take relevant actions on the basis of these predictions.

3 Definitions

Definitions of symbols employed in this paper are listed in **Table 1**.

Table 1: variables and definitions

variables	definitions
F	The features of monitored people
$P(x)$	The probability that X is in state x
$H(X)$	The result of Shannon Entropy
P_i	The node of the colony
C_i	Colony
d_c	The minimum distance between two colonies
d_{ij}	The distance between different colonies
$a(P_i)$	The average distance between P_i and other points in P_i's colony
$b(P_i,k)$	The average distance between P_i and C_k (where $\mathrm{P_i} \notin \mathrm{C_k}$)
d_{xo_j}	The distance between monitored person x and the cluster centroid of the j^{th} group
x_{f_i}	The i^{th} feature of individual x
$o_{j_{f_i}}$	The i^{th} feature of the cluster centroid of the j^{th} group
$s(P_i)$	The Silhouette's value
s_{f_i}	The standard value of the i^{th} feature (in our model, this value is 1)
n	The number of features
dis_{ij}	The distance between two points
dis_c	The interceptive distance
ρ_i	Local density of the particle
δ_i	The minimum value of dis_{ij}
$D_C(P_J)$	The degree centrality
$C_B(k)$	The betweenness centrality
$C_C(j)$	The closeness centrality
outer	The number of positive triad of monitored terrorist
inner	The number of negative triad of monitored terrorist

4 Models for Tasks

4.1 Task One

In consideration of the availability of prior data, our solutions for task one are divided into two parts (See **Figure 1**). We build two models to evaluate risk index depending on whether we have prior data of terrorist's internet usage features.

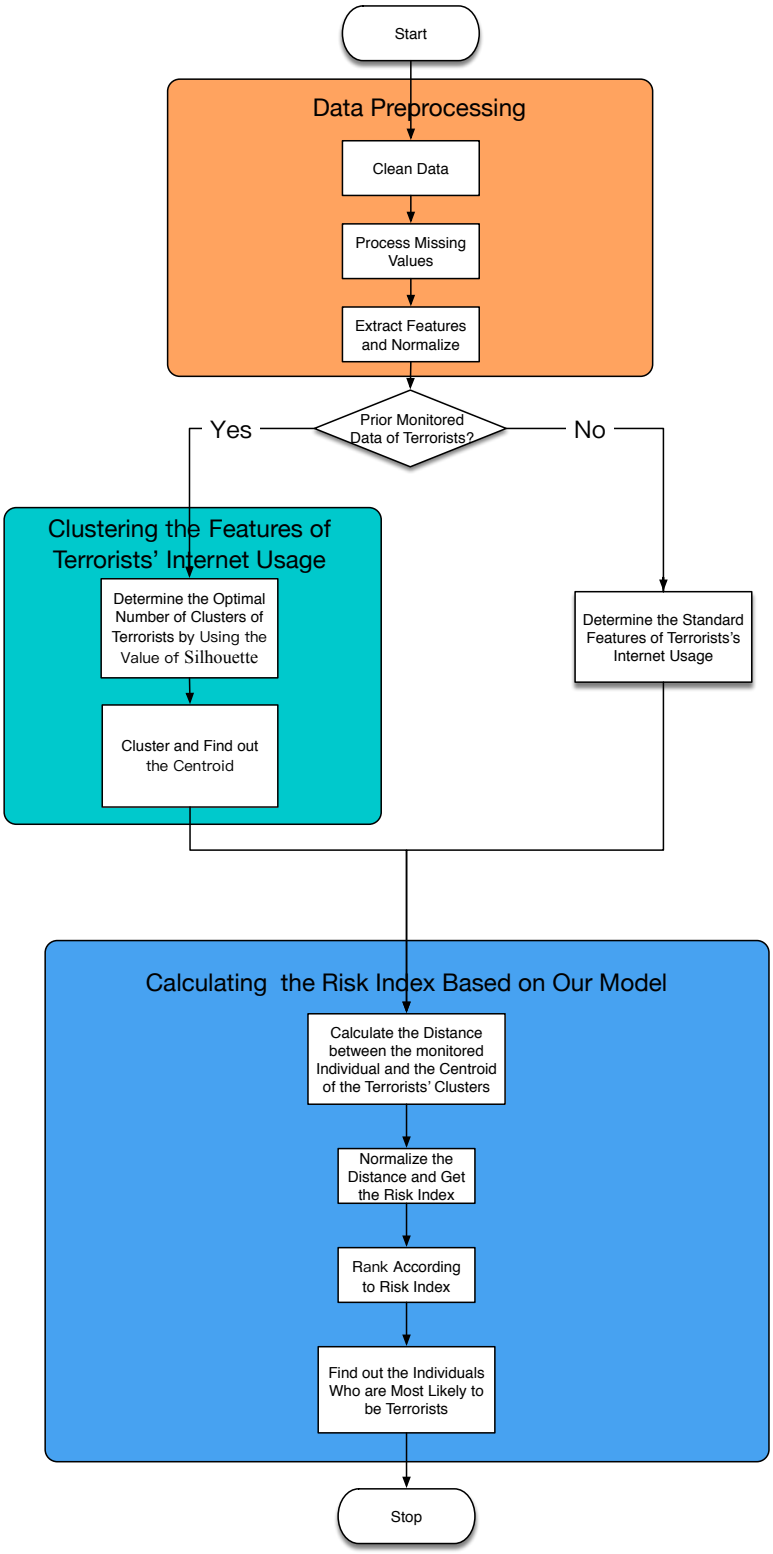

Figure 1: The process of modeling

4.1.1 Model One – Risk Index Evaluation Based on Prior Data

4.1.1.1 Problem Analysis

According to the passage above, we have mentioned that people turn to terrorism

by a series of " push and pull " factors. But in final analysis, it is because these outside factors cause the changing of people's psychological characteristics. We can evaluate psychological characteristics by monitoring people's internet usage mechanics. A recent research conducted by a team of computer scientists, engineers, and psychologists suggests that the monitoring internet usage data could show the tendency to experience depression. By analyzing existing research studying the terrorists' psychological characteristics, we think that a person who associates with strong depressive symptom is highly likely to have big issues such as extremism, bad temper and so on, which have comparative relations with terrorism. Thus, we see depression as the decisive factor of terrorism tendency. In Model One, we are going to solve as follows:

- Based on our analysis above, we build a risk index evaluation model based on depression tendency. In consideration of the realistic situation that different terrorists may have different internet using habits due to their different positions and identities, we firstly build multi-anchor clustering analysis to solve the following problems:
- With prior knowledge of terrorists, we need to cluster terrorists' internet using habits by six indexes. And then we can get some standard pattern categories of terrorists' internet usage.
- Calculate the risk index based on multi-anchor clustering analysis.

4.1.1.2 Assumption

- We can get the data of terrorists' internet usage characteristics.
- Terrorists' internet usage features have some certain distribution patterns.
- The individual depression has a decisive impact on terrorism tendency.

4.1.1.3 The Foundation of Model

Step 1, we will extract the individual internet usage feature. Based on these assumptions and the research of *' Associating Internet Usage with Depressive Behavior among College Students '*[5], we derive a number of internet usage features divided into three broad categories-the aggregate category captures raw aggregates of internet usage; the application usage category captures application specific internet usage features; and the entropy based features captures randomness in internet usage. After a series of statistical analysis, the research summarizes six practical interpretations to their findings between depressive symptoms among users and internet usage.

Average packets per flow: Larger number of packets per flow is typical under internet streaming and downloading when watching videos and gaming, which are common symptoms of internet addiction that have been shown to associate with depressive symptoms [6,7,8].

Peer-to-Peer usage: Sharing files like music, movies, photos etc. are primary reasons for using peer-to-peer services. Students are prone to be addicted to such kinds of content, which may explain this trend.

Chatting: Excess online chatting can affect the psychology of young people in terms of causing social isolation and loneliness in the real world, potentially leading to depressive symptoms [9,10].

Email: As in [10], frequent email checking may relate with high levels of anxiety, which in-turn correlates with depressive symptoms.

Flow Duration Entropy: When Flow Durations have high entropy, it is likely a result of frequent switching among multiple Internet applications. Frequent switching may also reflect an attempt to elevate feelings in the face of Anhedonia, when there is desperation to find something - an interesting article, an e- mail, a pleasing video, etc., to derive a momentary spark of pleasure and elevate mood. We capture randomness in Internet usage via *Shannon Entropy (H)*. Intuitively, entropy estimates the average uncertainty of a series of discrete events. Given a discrete random variable *X*, *Shannon entropy H(X)* is:

$$H(X) = -\sum_x P(x)\ log(P(x)) \tag{1}$$

where, $P(x)$ is the probability that *X* is in state *x*.

Ftp and Remote File usage: since excess ftp usage and remote file octets are indicative of excess file transfers, this could indicate addiction to certain types of files that may associate with depressive symptoms.

After obtaining data of the six targets, we normalize and vectorize these data and define the internet usage characteristic F for the convenience of later data processing.

Step 2, we will cluster the terrorists' internet usage features. Assuming that we already have the data of terrorists' internet usage features, we can use K-means to cluster these data of any two group according to the Euclidean distance of their centers, until the number of groups reaches our expectation.

$$d_{ij} = \sqrt{\sum_f \left(x_{if} - x_{jf}\right)^2} \tag{2}$$

The equation above needs an expected the number of categories. However, we may not have enough empirical knowledge to confirm this number, in fact. Consequently, we adopt Silhouette's value to confirm the most optimal number [15].

$$s(P_i) = \frac{\min\limits_{k|P_i \notin C_k} \left(b(P_i,k)\right) - a(P_i)}{max\left(a(P_i), \min\limits_{k|P_i \notin C_k} \left(b(P_i,k)\right)\right)} \tag{3}$$

Among them, $a(P_i)$ is the average distance between P_i and other points in P_i's colony, $b(P_i,k)$ is the average distance between P_i and C_k, where $P_i \notin C_k$. We can find that the value of *s* is between -1 and 1. Only if every point is a colony, $s = 1$; Only if some points belong to different clusters, $s = -1$. The more $s(P_i)$ is close to 1, the more P_i is suitable for current colony. The more $s(P_i)$ is close to -1, the more P_i is unsuitable for current colony.

We can set the initial number of colonies 2,3,4 or other values. In every case, we only calculate the Silhouette's value of each point in colonies which have more than one point. We ignore those colonies with one point because they will have an impact on the calculating of average distance. After changing the number of colonies to obtain the max average Silhouette's value, the colony number under which is the most optimal

accordingly.

Under the background of the internet usage big data of terrorists, this clustering method can calculate the certain patterns of terrorists' internet usage features rapidly and accurately.

Step 3, we'll evaluate the risk index of monitored people. After clustering, we obtain some certain patterns of terrorists' internet usage features. Considering analysis and assumptions above, we set the depression extent as the critical factor in evaluating the terrorism tendency [12].

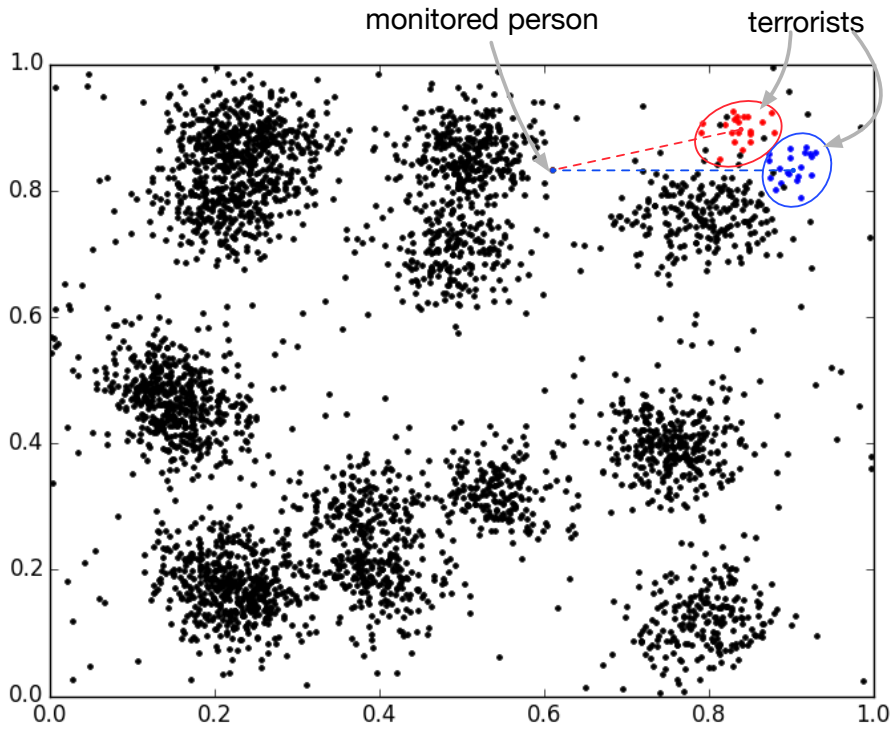

Figure 2: The calculating distance between monitored person and terrorists

In **Figure 2**, we evaluate the monitored person's risk index by multi-anchor analysis. After calculating the individual's distance to all terrorists' colonies respectively and normalizing the results, we set the minimum as his risk index, which reflects the possibility of becoming that kind of terrorists. The calculating equations are defined as:

$$risk\ index = \min_j \left(d_{xo_j}\right) \tag{4}$$

$$d_{xo_j} = \frac{\sqrt{\sum_i (x_{f_i} - o_{j_{f_i}})^2}}{n\sqrt{2}} \tag{5}$$

In the equation, d_{xo_j} represents the normalized distance between monitored individual x and the centroid of terrorist colony j, x_{f_i} represents the monitored person's

features i of the six targets, $o_{j_{f_i}}$ represents features *i* of the centroid of terrorist colony *j*, *n* represents the number of features (in this case, the value of *n* is 6)

We can rank the monitored people according to their risk index. And relevant measures like monitoring and preventing could be taken to those possible terrorists.

4.1.2 Model Two – Risk Index Evaluation without Prior Data

4.1.2.1 Problem Analysis

In Model One, we evaluate risk index based on prior knowledge of terrorists' internet usage features. However, in realistic situation, we may not get these prior terrorists' features and not access to the internet usage mechanics of terrorists. As a result, we need to solve the evaluation without prior terrorists' data. The document [5] mentioned the six internet usage features. According to our assumption and the document [5], the higher the risk index is, the more depressive and risky the monitored person is. Based on these analysis, we build a model to evaluate the risk index without prior data.

4.1.2.2 Assumptions

- The extent of individual depression has a decisive impact on risk index.
- Terrorists' internet usage features have some certain distribution patterns.

4.1.2.3 The foundation of model

Based on the research [5], we set terrorist's six features value 1. By calculating the Euclidean distance between the monitored person and the standard feature, we can obtain the risk index after normalization. The calculating equation is defined as:

$$risk\ index = \frac{\sqrt{\sum_i (x_{f_i} - s_{f_i})^2}}{n\sqrt{2}} \tag{6}$$

The risk index under different types of features is shown in **Figure 3.**

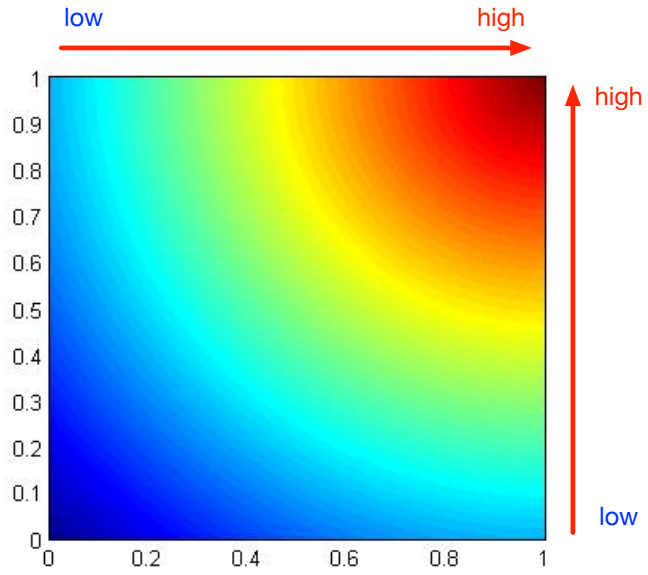

Figure 3: Risk index of different features

Meanwhile, we also use the radar map to visualize the model output (See **Figure 4** and **Figure 5**).

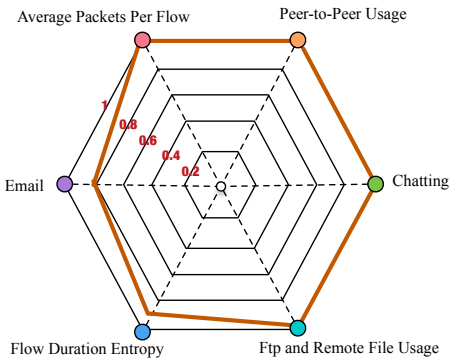

Figure 4:
Radar map of high risk index person

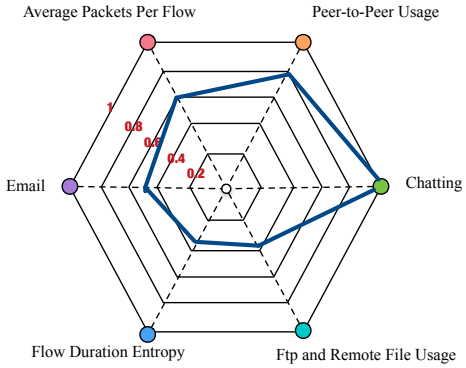

Figure 5:
Radar map of low risk index person

As it shown above, the radar map of **Figure 4** represents a person who has a high risk index, while **Figure 5** represents one who has a low risk index.

4.1.3 Missing data processing

There may exist missing data in reality. Before inputting data to our model, we should deal with the missing data by filling them with mean value, or obtain the corresponding value by using interpolation method.

4.1.4 Strengths and weaknesses

4.1.4.1 Strengths

- We take the availability of prior data into consideration and build two models respectively.
- We consider the missing data in reality and propose corresponding treatments.
- We consider different types of terrorists' internet usage features to make our results more accurately.
- We propose an effective quantized method of internet usage features. The number of the six features we set in the model can be adjusted according to practical data.
- Our risk index evaluation model is in high efficiency and convenient and can be adaptable to big data.

4.1.4.1 Weaknesses

- We set terrorists' internet usage features as some certain patterns without prior data, which is subjective to some extent.
- We cannot test our models due to a lack of practical data.
- The internet usage features of monitored people may change largely due to some uncertain factors in reality, which may cause great errors in our models.

4.2 Task Two

4.2.1 Problem Analysis

Experts use the expression big data to indicate huge amounts of information. Assuming that we have already got a lot of monitoring data, we need to turn these data into useful information [4]. After transforming data format, we can use the models we built in task one to evaluate risk index. In this chapter, we use a fast clustering algorithm mentioned in [12] to categorize these big data in an effective, fast and automatic manner. The problems needed to solve include:

- Cluster the monitored data, we can get different types of internet usage patterns colonies.
- Evaluate the risk index based on Model One or Two.

4.2.2 Assumptions

- The cluster centers are characterized by a higher density than their neighbors.
- The cluster centers have a relatively large distance from points with higher densities [12].
- There exist certain patterns of monitored people's internet usage features.

4.2.3 Model Three – Fast Clustering and Risk Index Evaluation Model

On the basis of model one, we add rapid cluster to monitored data. We use the improved clustering algorithm [12], and calculate risk index of every category to judge the terrorism colonies. The clustering algorithm goes as follows:

Step1, Based on assumptions, we calculate two values for every point *i* : The local density ρ_{i} and the distance δ_i between *i* and other points which have higher local density. The two values both depend on the distances d_{ij} between two points. The local density of point *i* is defined as:

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \tag{7}$$

If $x<0$, then $\chi(x) = 0$. d_c is a interceptive distance. Basically, ρ_i is equal to the number of points whose distance with *i* is less than d_c.

δ_i is the minimum of the distance between *i* and any point whose local density is greater than *i*. It's defined as:

$$\delta_i = \min_{j:\rho_j>\rho_i}(d_{ij}) \tag{8}$$

To the point whose local density is the largest, we can know:

$$\delta_i = \max_j(d_{ij}) \tag{9}$$

Step 2, after calculating ρ_i and δ of every point, we set the point with both high ρ_i and high δ as the clustering center.

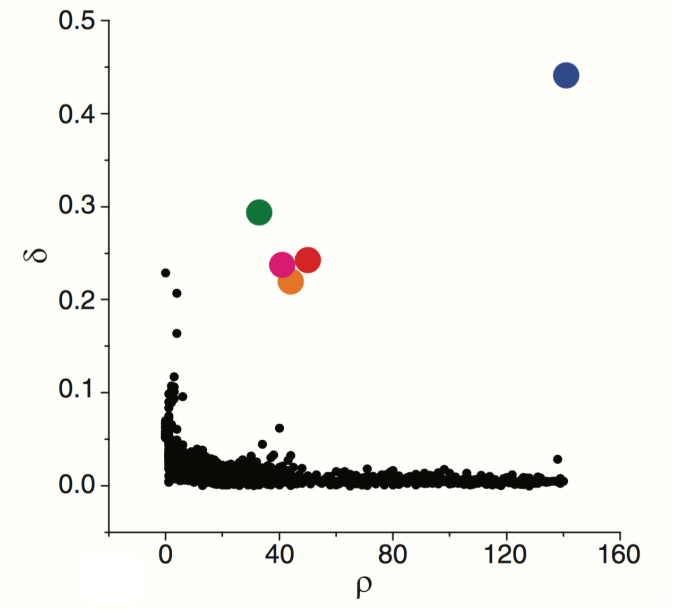

Figure 6: Centers of cluster obtained by the algorithm [12]

As we can see in **Figure 6**, the colorful points are clustering centers. In this algorithm, these clustering centers are obvious and easy to calculate.

Step 3, after the clustering centers have been found, each remaining point is assigned to the same cluster as its nearest neighbor of higher density. The cluster assignment is performed in a single step, in contrast with other clustering algorithms where an objective function is optimized iteratively [12].

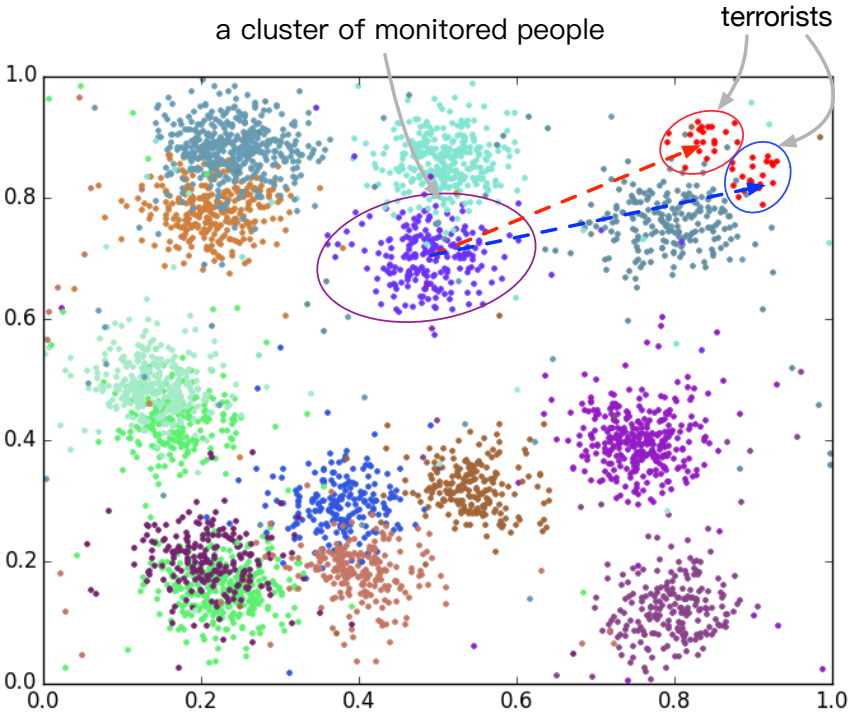

Figure 7: The result of clustering

After clustering, we use model one to evaluate the colony's risk index (See **Figure 7**). What is different from model one is we extend the individual evaluating method to colony.

4.2.4 Strengths and weaknesses

4.2.4.1 Strengths

- We categorize data by using unsupervised-clustering algorithm.
- The clustering algorithm we adopt is effective, fast and automatic, which is of strong adaptability to big data.
- After clustering, we calculate colonies' risk index by employing our model one and model two.

4.2.4.2 Weaknesses

- We cannot test our models due to a lack of practical data.
- Some parameters in the clustering algorithm need manual adjustments to attain the superior result.

4.3 Task Three

4.3.1 Problem Analysis

In task 3, our solutions are carried out from four aspects, Here is our analysis (See **Figure 8**).

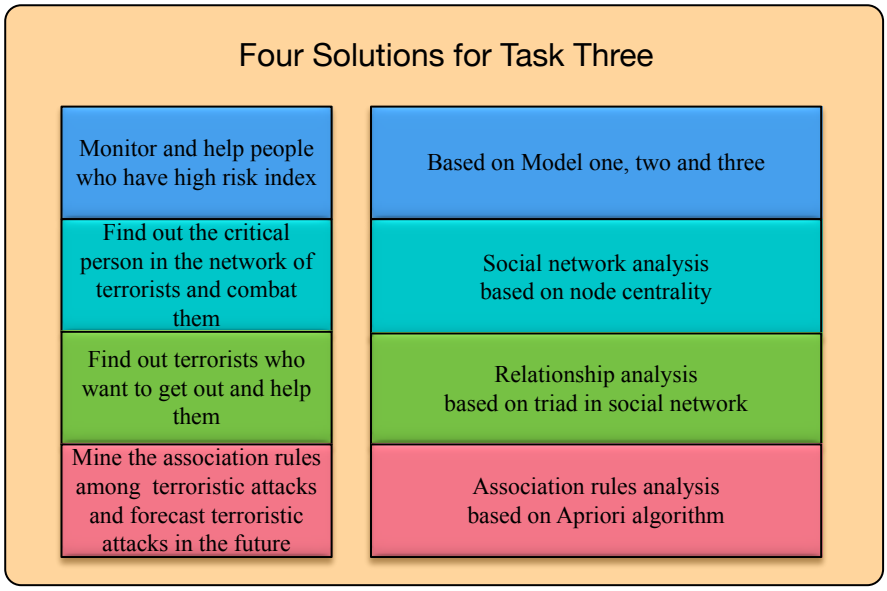

Figure 8: Four solutions for Task Three

First, we have built Model One, Two and Three to cluster monitored big data effectively and evaluate the risk index of monitored people. Based on the three models, we can take some measures like monitoring and prevention if his evaluating risk index is abnormal, so as to eliminate terrorism fundamentally. The second aspect starts from striking current terrorism. We can find the leader through social network analysis to carry out correct anti-terrorism strategic. Thirdly, we find out the terrorists who want to

get out through triad analysis to rescue the terrorists who want to escape in time. Last but not least, we think some prevention measures of terrorism attacks should be taken. We can obtain rich data about terrorism from the website [13], which include the initiators, dates, cities, the numbers of the injured, target type etc. With these information, we can mine the association rules among different terroristic attacks to forecast terroristic attacks in the future and make full preparation for anti-terrorism.

4.3.2 Assumptions

- We can get richer data, which includes not only internet usage features but also individual social relationships.
- We can learn about detailed inner network data in terroristic organization.
- In a social network, the one with the highest centrality index plays a critical role.

4.3.3 Solution One

We can calculate individual or group's risk index using Model One, Two or Three. To those with high risk index and potential terrorism tendency, we should monitor them and help them out of abnormal psychology to eliminate terrorism tendency from origin.

4.3.4 Solution Two

There are three methods to evaluate the node's centrality, including, degree centrality $D_C(P_J)$, betweenness centrality $C_B(k)$ and closeness centrality $C_C(j)$. Here we build a general model to calculate the importance of nodes in social network. The equation is defined as:

$$D_C(P_J) = \sum_{i=1}^{n} d(p_i, p_j), \text{ where } \mathrm{d}(p_i, p_j) = \begin{cases} 1, & p_i \\ \mathrm{x}, & \mathrm{x} \geq 0 \end{cases} \tag{10}$$

$$C_B(k) = \sum_{i \neq j \neq k \in V} \frac{\sigma_{ij}(k)}{\sigma_{ij}} \tag{11}$$

$$C_C(j) = [\sum_{i=1}^{n} d(i,j)]^{-1} \tag{12}$$

$$centrality\ index = \alpha D_C(P_J) + \beta C_B(k) + \gamma C_C(j) \tag{13}$$

In equation 13, α, β, γ is the weight of the three centrality index respectively, which should be adjusted in different kinds of situations. Due to the lack of actual data, here we think the three index is equally important, and their values are all set one. Here is a typical inner network of terroristic organization (See **Figure 9**).

From **Figure 9**, we can see that the critical person is in the center of the network, and has complex network relationship. Document [14] mentioned a concept of *HVT (High Value Target)*, that is to say, a new leader would appear after the old leader was killed in a terroristic organization. Shakarain introduces a concept of *'shaping actions'*[14]. He suggests that before striking the leader, we should take some actions to collapse the ability of creating new leader in a terroristic organization. Then the organization would decline soon as a result of the lack of leader.

Consequently, we can find out the critical terrorist through social network analysis and collapse the terroristic organization effectively.

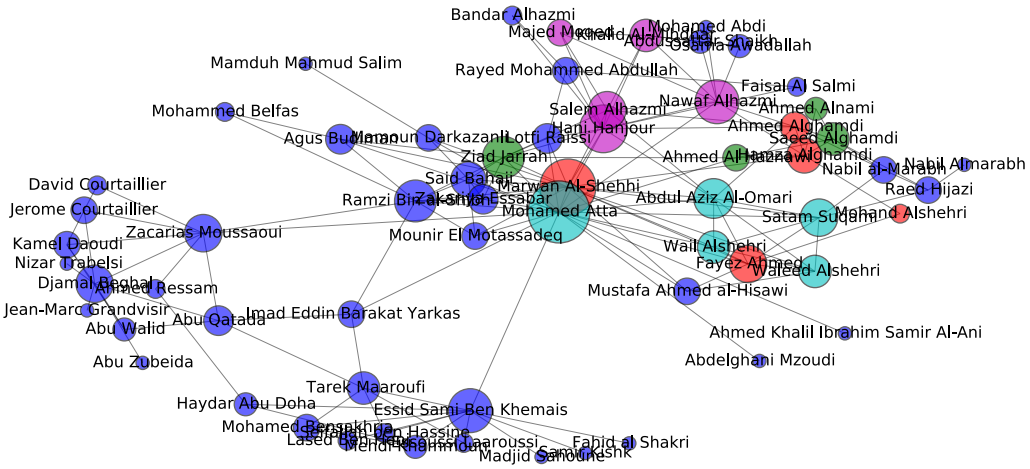

Figure 9: The social network of hijackers in 911 [11]

4.3.5 Solution Three

As mentioned in the subject，‘They report a sense of suffocation - being unable to leave for fear of retaliation (either by the terrorists or by the State) and being equally afraid of their disillusionment being detected by those close to them in the movement. We need to do a better job of providing “off-ramps”’ [4]. Thus, we build an effective triad model to find these people based on social network.

The social relationship can change one’s tendency to a great extent, and the relationship of triad is the most stable with the time coursing [11]. In realistic social network, a triad is the connection among three people in the form of social relationship (See **Figure 10**)

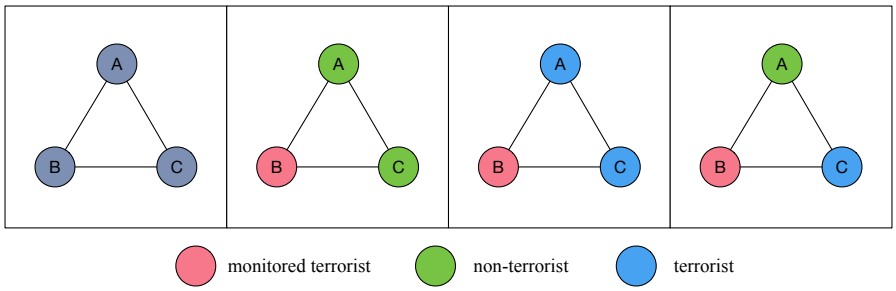

Figure 10: Triad

As we can see in **Figure 10,** the left one is the diagram of triad. The red node represents the monitored terrorists, and the right side of the three pictures shows the three cases of the triad, from left to right, respectively, which are positive triad, negative triad and balanced triad.

Based on our assumptions and analysis, our equation to calculate terrorist’s ‘flee index’ is defined as:

$$flee\ index = tanh\left(\frac{outer+1}{inner+1}\right) \tag{14}$$

In **equation (14)**, $tanh$ is the hyperbolic tangent function, *outer* is the number of positive triad of monitored terrorist, while *inner* is the number of negative one. In addition, we use Laplace Smoothing to avoid the problem due to a divide by zero. We can use this 'flee index' calculating model to identify those terrorists who want to get out, and then can help and rescue them according to the results of investigation.

4.3.6 Solution Four

The act of terrorism is not entirely random, but a rule to follow. Consequently, finding out the rules of terroristic attacks is of great importance to prevent terroristic incidents and combat terrorism. There is plenty of data related to terrorism on the website [13]. We mine the association rules of the data on the website based on Apriori algorithm, which can be used to predict future terrorism incident to take effective anti-terrorism measures.

Here are the specific steps of Apriori algorithm:

Step 1, calculating the support of item sets and finding out frequent item sets meeting the minimum support requirements.

Step 2, finding out potential association rules in frequent item sets and calculating their confidence to identify association rules meeting the minimum confidence requirements.

We have obtained the data of terrorist attacks in Iraq from [13], the attacks are from September 1st to December 31st in 2014. By using Apriori algorithm, we got the following results:

Table 2: The result of association analysis

Current place	Next place	Confidence Degree	Current place	Next place	Confidence Degree
Madain	Baghdad	1.00	Yusufiyah	Baghdad	1.00
Baiji	Baghdad	1.00	Mahmudiyah	Baghdad	1.00
Ramadi	Baghdad	1.00	Baghdad	Mahmudiyah	0.71
Tuz Khormato	Baghdad	1.00	Baghdad	Ramadi	0.71
Mosul	Baghdad	1.00			

In this algorithm, we set the length of sliding time window of one week, the minimum support level of 0.5 and the minimum confidence level of 0.7 in association rules. By analyzing, we get the results (See **Table 2**), from which we can see the association of terroristic attacks in different cities. For example, as we can see, if a terroristic attack happens in Madain in one week, another terroristic attack will be likely to happen in Baghdad in the next week. By mining the association rules among prior information, we can predict the laws of terrorism attacks to prevent them effectively.

4.3.7 Summary

In Task Three, we give four pieces of advice for President Obama in combating terrorism, that is, controlling for terrorism tendency, combating the critical persons in a terroristic organization, rescuing terrorists who want to get out and mining association rules in terroristic attacks. We think the four measures are effective and feasible.

Nowadays science and technology are primary productive force, we should use science to fight terrorism.

5 Robustness Analysis

In Task One, we considerate the accessibility of prior data and build two models corresponding to the two situations-one is with prior terrorists' internet usage features, the other is without prior terrorists' internet usage features. In addition, we deal with the missing data in reality reasonably, which helps our models to deal with practical data accurately. What's more, we normalize the data before putting them into our models, which avoids the abnormal values in calculating risk index.

In Task Two, we adopt a fast and effective clustering algorithm. Different from other algorithms, it can avoid the objective function being optimized iteratively, which makes it very adaptable in big data. However, some of its parameters need annual adjustment, which may have some impacts on clustering results.

In Task Three, we propose four solutions. In identifying critical persons in terroristic organization, we set three values to calculate centrality index and the weights of the three values can be adjusted according to realistic situation, which makes our models more applicable. However, it's a tough job for us to obtain the data of social relationships network of terrorists. In the model of calculating the terrorist's 'Flee Index', we revise the 'flee index' value to a reasonable range by using hyperbolic tangent function (See **Figure 11**).

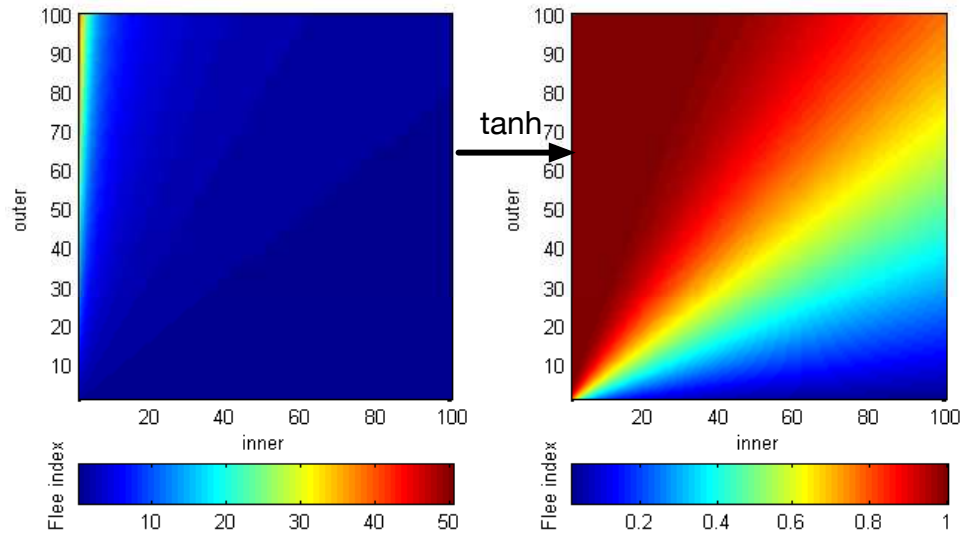

Figure 11: The revision of flee index by hyperbolic tangent function

In associating rules analysis, we test our models with practical data in website [13] and obtain many meaningful conclusions.

6 Conclusion and Future Work

6.1 Conclusion

Science is a powerful weapon and we can solve terrorism through scientific methods in different aspects. In our passage, we firstly build risk index calculating model according to internet usage features to identify potential terrorists. Secondly, we calculate colonies' risk index after rapid clustering. Lastly, we identify critical persons and terrorists with high 'flee index' in terrorists' organization by social network analysis, and mine association rules among terrorism attacks. With our four solutions, we can identify potential terrorists, collapse terrorism organizations, rescue terrorists who want to get out and prevent terrorism attacks in advance. Consequently, we can draw the conclusion that effective scientific models can solve terrorism to a great extent. The government and security officials should pay attention to the application of science in anti-terrorism.

6.2 Future Work

- We need to obtain realistic internet usage features to test and improve our models.
- In mining association rules among terrorism attacks, a more rapid and efficient FP-growth algorithm should be founded to identify frequent item sets.

7 References

[1] "In extended or weakened use: the instilling of fear or terror; intimidation, coercion, bullying" ("terrorism, n.". Oxford English Dictionary (3rd ed.). Oxford University Press. September 2005.

[2] Hoffman, Bruce (1998). Inside Terrorism. Columbia University Press. p. 32.

[3] http://terrorism.about.com/od/causes/a/causes_terror.htm.

[4] Can Science Solve Terrorism? Q&A with Psychologist John Horgan.

[5] Associating Internet Usage with Depressive Behavior among College Students.Raghavendra Kotikalapudi, Sriram Chellappan, Frances Montgomery, Donald Wunsch and Karl Lutzen.

[6] A. Boals. Correlates of Video Game Addiction. PhD thesis, University of North Texas 2010.

[7] J. Kim, R. LaRose, and W. Peng. Loneliness as the cause and the effect of problematic internet use: The relationship between internet use and psychological well-being. CyberPsychology & Behavior, 12(4):451–455, 2009.

[8] J.B. Weaver III, D. Mays, S. Sargent Weaver, W. Kannenberg, G.L. Hopkins, D. Eroglu, and J.M. Bernhardt. Health-risk correlates of video-game playing among adults. American Journal of Preventive Medicine, 37(4):299–305, 2009.

[9] L. Bonetti, M.A. Campbell, and L. Gilmore. The relationship of loneliness and social anxiety with children's and adolescents' online communication. Cyber Psychology, Behavior, and Social Networking, 13(3):279–285, 2010.

[10] J. Morahan-Martin and P. Schumacher. Loneliness and social uses of the Internet. Computers in Human Behavior, 19(6):659–671, 2003.

[11] Maksim Tsvetovat, Alexander Kouznetsov. Social Network Analysis for Startups

(O'Reilly).2011. 978-1-449-30646-5.
[12] Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. Science 344(6191):1492–1496.
[13] http://www.start.umd.edu/gtd/.
[14] http://world.chinaso.com/detail/20151116/1000200032709021447642184801064524_1.html.
[15] Zhongmeng Zhao, Jiayin Wang, Chen Qiao. Analysis and Research on the contest questions of COMAP.