Research on Counter-terrorism based on Big Data

Sanjun Nie, Duoyong Sun
College of Information System & Management
National University of Defense Technology
Changsha, China
e-mail: m18274849704@163.com

***Abstract*—In the area of big data, people have a new perspective on counter-terrorism research. In this paper, we have carried out a systematic research on the applications of big data in counter-terrorism field by using quantitative analysis method. And then we have demonstrated effect of big data on counter-terrorism research from data collection and preprocessing, data mining and analysis, monitoring and warning three aspects. After that, we have used all the data of The Times and The New York Times about WUC in 2012 for analysis and argument. Finally, it concludes the deficiencies of the research on the territory of counter-terrorism by using big data and the problems worth studying in the future.**

Keywords—big data; Counter-terrorism; Data mining

I. INTRODUCTION

After World War Ⅱ, the terrorism have appeared in the world, as the world entered the information age, the terrorist attacks have become one of the major threats to the state, society and the public security [1]. After the 1990s, terrorism has occurred substantially change, the structure of organization has become more networked [2]. There are many experts who pointed out that, after 9/11 attacks, the terrorist organizations have become more decentralized, as the international terrorist organization networks which centre around 'al-Qeaeda' have spread over 60 countries worldwide [3]. Therefore, more and more terrorists have made a series of activities through online media, such as contact with other terrorists, release the terrorist information, raised funds, recruited members, and engaged in illegal activities and so on. This information of activities will always leave traces in the internet. However, we have been the area of big data, the data in the internet is large in scale, variety and obviously unstructured. Therefore, using traditional techniques and tools to dispose this data often couldn't dig out any useful information. On the other hand, the terrorists also required a lot of preparatory work before the terrorist attacks, so the massive amounts of data produced by everyday work of people also may have carried terrorism clues. So it is necessary to study counter-terrorism by using data mining and analysis techniques.

In fact, as early as after 9/11 attack, Krebs has drawn the task network of 9/11 attack by using open source data, and concluded the structural feature of terrorist organization [3]. Peter Katona, Michael D. Intriligator and John P. Sullivan (2006) [4] proposed that it is necessary to build and improve the global information network of counter-terrorism through data mining. Benxian Li [5] have drawn the East Turkistan terrorist organization network by using data mining technology, and also the data comes from the three batches of terrorist data released by China Ministry of Public Security, and then they found the network of East Turkistan terrorist organization obeys power-law distribution. Hai Zhang and Duoyong Sun [6] have extracted character and organizational attributes about East Turkistan terrorist organization by using text mining method from the internet, and then they used social network analysis method to analyze the network. However, there are few literatures which focus on the systematic research. The existing researches on counter-terrorism through big data mostly study it from one aspect. This paper aims to break through the current situation.

The rest of this paper is organized as follows. In section two, we demonstrated effect of big data on counter-terrorism research from data collection and preprocessing, data mining and analysis, monitoring and warning three aspects. Then we have used all the data of The Times and The New York Times about WUC in 2012 for analysis and argument. Section four and five concludes the deficiencies of the research on the territory of counter-terrorism by using big data and the problems worth studying in the future.

II. THE APPLICATION OF BIG DATA IN COUNTER-TERRORISM

Figure 1 depicts the application framework of big data in counter-terrorism work, and the rest of this section organized as follows:

A. Data Collection and Preprocessing

Intelligence is the lifeline of counter-terrorism work. Mastering reliable information in time can play an active role when fighting against terrorism, and then it can effectively curb the spread of terrorism. In recent times, big data has become a major source of anti-terrorist intelligence. However, in many cases, the big data related to terrorist attacks neither sufficient nor open. Therefore, we must wide the vision and channel for the collection of counter-terrorism intelligence data, and then find out the relevant data sources about terrorism from all sides. So far as the status, there are three aspects of data sources for terrorist intelligence: the

Authorized licensed use limited to: Monash University. Downloaded on March 22,2021 at 07:33:34 UTC from IEEE Xplore. Restrictions apply.

internet data of open resource, at present, most of counter-terrorism researches is based on it, the most classic of which is Krebs and his research [3], in which Krebs have drawn 9/11 task network based on internet data of open source. The second is social media data, which is closer to life compared to internet data of open source. A variety of mobile data, communication data all come from people's daily life. Such data may carry some traces of terrorist acts, which can be better used for early warning and tracking. And the third kind of data is artificial collected data. However, these data acquisition may produce high cost, and also may be often accompanied by great risk, even likely to cause casualties. In spite of this, it is difficult to ensure the timeliness of the intelligence.

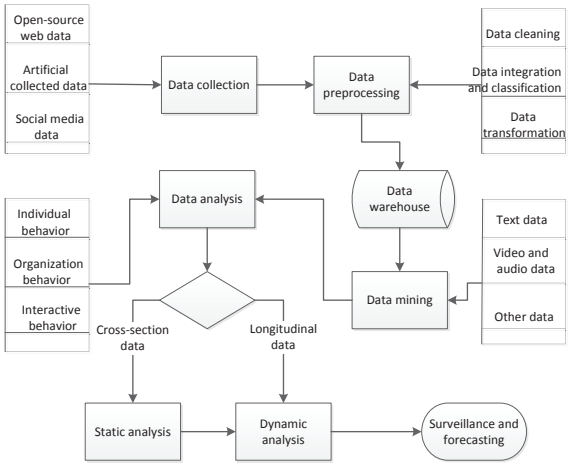

Figure 1. The application framework of big data in counter-terrorism work.

1) *Data Collection*

For open-source web data, the anti-terrorism department can search a large number of original data by related software of Web Crawlers. And for social media data, the anti-terrorism department can take the expanded intelligence gathering method of 'three-dimensional rolling expanding'. Which means that on the premise of observe the law, the related workers can proceed from a little of existing cues (such as nickname, phone number, etc), and then collect the valuable data from the real space through various social sectors, departments, enterprises, and even individual citizens data center. Finally, the anti-terrorist department can screen out the data that reflects the ideological trend about terrorism, the case clues and the behavior trace of terrorist organizations and members by using default data model. Also, the traditional manual collected data has its unique advantages; it is very accuracy though less information and high cost. The collected terrorism-related data through these three channels can be stored in large data terrorism data warehouse after pre-processing.

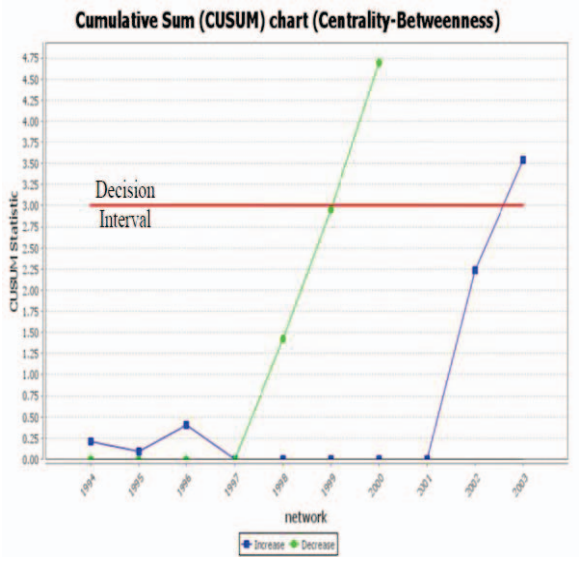

Figure 2. The SPC chart for parameters of terrorist organization [11].

2) *Data Pre-processing*

There are three aspect in pre-processing the big data related to rerrorism: data cleaning, data integration, data transfomation and classification. Removing the noise data and the blank data area belong to data cleaning. Many kinds of data analysis software have build-in data cleaning function, people can also define the noise data based on their own research. Data integration means that gathering the relevant data from different data sources, while data classification means that sorting the different kinds of data. Data conversion refers to that converting the terrorism-related collected data into the required format to carry out the next step.

B. *Data Mining And Analysis*

This part is focus on the correlation analysis and rule extraction after the terrorism-related data pre-processed. It aims that analyzing and finding out the temporal and spatial characteristics of terrorism attacks, the behavior patterns of terrorist organizations and members and so on, and then providing theoretical basis for the early warning and monitoring of counter-terrorism. Currently, the useful and effective methods for data mining and analysis are SNA (Social network analysis).

SNA is built on the interactive of actors, in which actors are considered as points, the cennection between points indicat the relationships between actors [7]. After 9/11 attack, terrorism has become one of the major threats to world peace and security. The related scholars have began to conduct anti-terrorism research by using SNA. Rodriguez [8] constructed the terrorist network about Madrid bombings in 2004, and then found the decentralized network based on weak ties. Memom [9] have constructed the terrorist network of London bombings and found out the key persons of network and the covert organization structure. Most of

Authorized licensed use limited to: Monash University. Downloaded on March 22,2021 at 07:33:34 UTC from IEEE Xplore. Restrictions apply.

these studies have combined SNA and organization behavior to locate the key persons and to explain and evaluate the behavior patterns of terrorist organizations and members.

The core measures in SNA are degree centrality, closeness centrality and betweenness centrality:

1) Degree centrality: The degree centrlity is a measurement of connection degree from one node to other nodes. If the degree centrality value of one node is great, the node is considered active, which located in the center position, probably have great power. The degree centrality of point i can be divided into two categories, namely absolute degree centrality and relative degree centrality. The former is the number of nodes who derectly conect with the point, with $C_{AD}(i)$ representation. The latter is aims to facilitate comparison of different size networks, which is the standardized form of former, the mathematical expression is:

$$C_{RD}(i) = C_{AD}(i)/(n-1) \qquad (1)$$

2) Closeness centrality: The closeness centrality is a measurement of difficulty level for one node to other nodes, which reflecing surveillance and information capabilities. The absolute closeness centrality of point i refers to the sum of the shortest distances from the node to other nodes. Its expression is:

$$C_{AP}^{-1}(i) = \sum_{j=1}^{n} d_{ij} \qquad (2)$$

For comparing the closeness centrality of nodes who come form different network, it is necessary to calculate the relative closeness centrality, whose mathematical expression is:

$$C_{RP}^{-1}(i) = C_{ap}(i)/(n-1) \qquad (3)$$

3) Betweenness centrality: The betweenness centrality is a measurement of the degree of one node which located in the middle of other nodes. A point that the betweenness centrality is great plays a important "intermediary" role, which is most likely play a "gatekeeper", "liaison" role. The absolute betweenness centrality of point i is calculated as:

$$C_{AB}(i) = \sum_{j}^{n}\sum_{k}^{n} b_{ij}(i), j \qquad (4)$$

For comparing the betweenness centrality of nodes who come from diffrent network, it is necessary to calculate the relative betweenness centrality, whose mathematical expression is:

$$C_{RB}(i) = 2C_{AB}(i)/n^2 - 3n + 2 \qquad (5)$$

C. Surveilance And Forecasting

The application of big data in counter-terrorism research aims to mine and analyze the the related factors about terrorism attacks, and find out the temporal and spatial characteristics to prevent terrorism attacks. Dynamic social network analysis can effectively monitor abnormal of terrorist organizations. Actually, as early as in 1977, Holland and Leinhardt proposed Markov models and Multi-agent simulation models to analysis longitudinal social networks [10]. While Carly and his team first put forward that longitudinal social network analysis can effectively monitor the abnormal when the parameters of terrorist organization changed. They used SPC (statistical process control) charts to monitor the parameters of Al-Qaida organizations [11]. As shown in figure 2, we can see that the year of 2000 is an outline (beyond the warning line), it should be alert in this time, and the subsequent events have proved the effectiveness of this method (Al-Qaeda have launched 9/11 attack event in 2001). Since then, many scholars have made improvement and innovation based on it [12], [13], which greatly promoted the development of longitudinal social network analysis in the field of counter-terrorism.

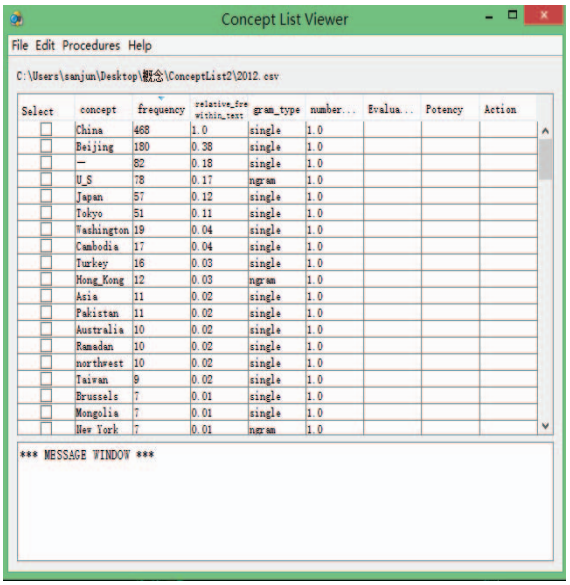

Figure 3. Concept list.

III. Positive Analysis

In this section, we have written a Web Crawlers program to gather all the data of The Times and The New York Times about WUC in 2012,after that we have pre-processed the data and extracted the related properties by using AUTOMAP software. Finally we have used ORA to analyze it.

D. Preparatory Work

AUTOMAP is a special text analysis tool for networks, it can extract vocabulary of text and their correlations, constitute networks through them. The network diagram is 'mind mapping' of author, we can analyze the network diagrams and find out the rules to effectively understand the authors' ideas. In other words,the text network analysis based on the assumption the language and the knowledge can

Authorized licensed use limited to: Monash University. Downloaded on March 22,2021 at 07:33:34 UTC from IEEE Xplore. Restrictions apply.

be made up to network models through the vocabulary which made of language and their relationships.

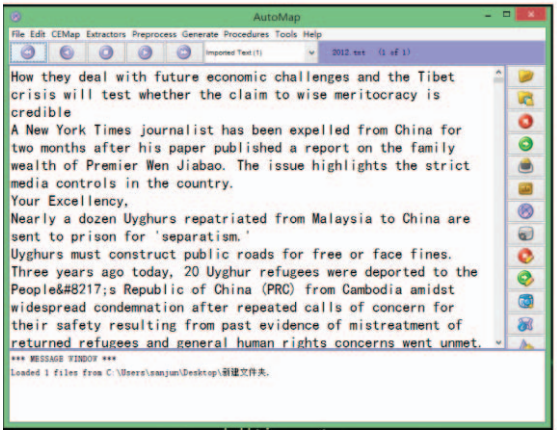

Figure 4. Importing the data crawled to AUTOMAP.

AUTOMAP can extract the data of several categories: agent, organization, knowledge, role, task, location and so on. In this section, we mainly analyzed the hot area where WUC terrorist organization focus on. Therefore, we convert the data crawled into UTF-8 format which AUTOMAP need, and then import it as shown in figure 4. After that, we have used AUTOMAP to extract the location property, as shown in figure 3, in which there is a concept list of locations.

Figure 5. The top 10 ranked locations

E. Building Netwrok Models

Then we have made the concepts constructed a semantic network and imported it to ORA, as shown in figure 5.

We have used ORA to analyze all measures for the locations, such as degree centrality, betweenness degree centrality, closeness degree centrality and so on. Then we listed the top ten locations to analyze in table 1 and figure 6.

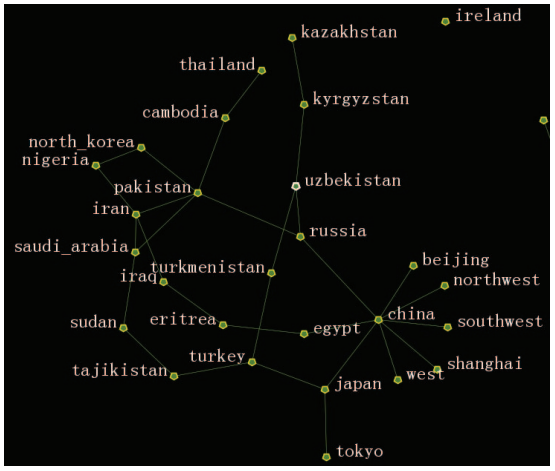

Figure 6. The degree centrality in national network models over time.

IV. DISCUSSION

In social network analysis, the degree centrality is a measure index of active degree for actor, which reflecting the degree of concern for region by the organization here. The betweenness degree centrality indicates that an actor of group whether easy get to other actors, which also reflecting the active level of organization.

In figure 6, we can see the top 10 hotspots related to WUC terrorist organization, wherein the top three are China, Japan and Pakistan, Russia Saudi Arabia, Afghanistan and son are also the counties which the WUC terrorist organization focus on.

TABLE I. THE TOP 10 LOCATIONS OF INDEXES

Location	Degree centrality	Location	Betweennes s centrality
China	0.993	Beijing	0.954
Pakistan	0.777	China_northwe st	0.636
Iran	0.500	Russia	0.338
Japan	0.223	Japan	0.338
Russia	0.223	Egypt	0.338
Saudi_arabi a	0.223	Shanghai	0.234
Turkey	0.223	China_southwe st	0.234
Uzbekistan	0.114	China_west	0.124
Cambodia	0.076	Turkmenistan	0.088
Egypt	0.076	Saudi_arabia	0.088

In table I, it is indicated that in terms of degree centrality, China has been the country which the WUC first focus on, and then Pakistan and Iran have been the second and third. The WUC is a wholly anti-China organization, they have been tried in vain to split Xinjiang from China, and the Al-Qaeda support it behind. Therefore, except China, these degree centralities of Middle Eastern countries like Pakistan and Iran which border with Afghanistan have large values.

Authorized licensed use limited to: Monash University. Downloaded on March 22,2021 at 07:33:34 UTC from IEEE Xplore. Restrictions apply.

For betweenness centrality, Beijing ranked first, the second is the northwest of China, namely Xinjiang, which also indicates Beijing and the northwest of China will be the key areas for activities of the WUC. It is worth noting that the WUC have designed the Beijing Golden Water Bridge event, which has proved the conclusion.

V. Conclusion and Future Research

In conclusion, the activities of various organizations in the world have become more and more frequent, and the harm also has become bigger and bigger. Therefore, the anti-terrorism departments must pay more attention to counter-terrorism intelligence, and strengthen the monitoring of terrorist organizations by using big data to avoid terrorist activities.

However, the application of big data in counter-terrorism still exist some problems, and which also are worth studying in the future. First, studying counter-terrorism based on big data may involve a lot of privacy; even sometimes violate the privacy of citizens. Second, in terms of the worldwide counter-terrorism, the cooperation is limited to information on scale; it is difficult to share the active value of big data, as terrorism is often related to the background of deep international politics and national history. Finally, the capabilities of data pre-processing and analysis remain to be improved. How to eliminate the redundancy part of big data, and handle multiple types of data, are still the difficulties which people wish to resolve.

Acknowledgment

This research is supported by the National Natural Science Foundation of China (Grant No.71473263) and the Specialized Research Fund for Doctoral Program of Higher Education of China (Grant No.20134307110020).

References

[1] Yizhou Wang, etc. Origins of Terrorism: Perception of Chinese [M]. Social Science Academic Press, 2010.

[2] Yirong Guo. The new trends of terrorism offenses and the coping mechanisms in our country [J]. Journal of Economic and Social Development research, 2014:8-2.

[3] Krebs V. Mapping Networks of Terrorist Cells[J]. Connections, 2001, 24 (3): 43-52.

[4] Peter Katona, Michael D Intriligator, John P Sullivan. Counte-ring Terrorism and WMD: Creating a Global Counter-terrorism Network[M]. New York: Routledge, 2006

[5] Benxian Li, Mengjun Li, Duoyong Sun, Yan Chi. Mapping "East Turkistan" Terrorism Network in China[J]. Journal of the china society for scientific and technical information, 32(2), 171-189 , 2013.

[6] Hai Zhang, Duoyong Sun. The research on covert networks of terrorist organization based on social network analysis [D]. National University of defense Technology. 2012.

[7] Lin Juren. Social network analysis: Theory, methods and applications[M]. Beijing: Beijing Normal University Press, 2009.

[8] Rodr__guez J A. The March 11th terrorist network: In its weakness lies its strength[C]// XXV.International Sunbelt Conference, Los Angeles, 2003.

[9] Memon N, Harkiolakis N, Hicks D L. Detecting high-value individuals in covert networks: 7/7 London bombing case study[C]// AICCSA 08-6th IEEE/ACS International Conference on Computer Systems and Applications,2008: 206-215.

[10] P. W. Holland and S. Leinhardt, "A dynamic model for social networks " Journal of Mathematical Sociology, vol. 5, pp. 5-20, 1977.

[11] Ian McCulloh. Detecting Changes in a Dynamic Social Network[C]. Institute for Software Research School of Computer Science Carnegie Mellon University. 2009.

[12] Zhao Jian. Improvement of Social Network Change Detection Based on Time Series Analyze [D]. Xidian University. 2011

[13] Li Ze. The research on Methods for Detecting Dynamic Change of Terrorist Network Based on SPC Control Chart [D]. National University of Defense Technology.

Authorized licensed use limited to: Monash University. Downloaded on March 22,2021 at 07:33:34 UTC from IEEE Xplore. Restrictions apply.