Research on Multi-factors Terrorist Attacks in China Based on K-Apriori Algorithm Research

Shang Liu*, Hao Chen and Yan Yu
Institute of Science and Technology, Tianjin University of Finance and Economics, Tianjin, China

*Corresponding author email: liushangw@tjufe.edu.cn

Abstract. Multi-factor research on terrorist attacks is of great practical significance. In order to explore and analyse the correlation between important factors in Terrorist attacks in China, a K-Apriori algorithm based on clustering was proposed. Aiming at the problem that only the attribute of casualties is considered in the general method, this paper firstly adopts the K-Means method to combine the attribute of death and the attribute of injury, so as to make full use of the information of casualties in terrorist attacks. Then, the Apriori algorithm is used to mine the pre-processed data set to find frequent item sets, so as to find the association rules between multiple factors of terrorist attacks. On the basis of the selected association rules, combined with the practical basis, the realistic meaning hidden under the appearance of association rules is mined to provide the intelligence reference for the anti-terrorism department.

1. Introduction
The global terrorist attacks have gradually increased since 1990 [1]. After the ‘9.11’ terrorist attack, the US government directly regarded terrorism as the most serious threat. After the ‘11.13’ Paris terrorist attack [2], the government announced that France was in a state of emergency. In recent years, the number of terrorist attack is growing in China. The uncertain and powerful destructiveness of terrorism seriously threaten China’s national security and social stability. With the continuous change of space-time factors of terrorist attacks, the targets of terrorist attacks also show a trend of diversification. So, this paper aims to find the relation among time, location, and target of attacks, which by studying the multi-factors of terrorist attacks in China. Our research can help to reduce terrorist attacks and protect people's lives and property.

Li, Sui used statistical multivariate statistical analysis methods to analyse terrorism and explored the impact of single factors such as time on death levels [3,4]. Li used decision trees, prior principles, and other methods to mine terrorism intelligence information on virtual sample data sets for daily terrorism intelligence analysis [5,6]. Shu used descriptive statistical analysis methods to analyse the characteristics and trends of terrorist attacks [7]. Cheng used the spatial hotspot analysis method to excavate the hotspots of terrorist attacks to assess the security situation of terrorist attacks in various regions [8]. However, these methods only consider the death attribute or injury attribute, while the innovation of this method lies in the adoption of k-means method to combine the death attribute and the injury attribute. The clustering of these two can fully reflect the damage of people in terrorist attacks, and better reflect the impact of people in the event. And other research methods are not suitable for early warning of real scenes, because they used virtual intelligence data sets. We make the following contributions:
1) The clustering of the death attribute or injury attribute—We introduce the method of death and injury clustering to process the data.

2) Association rules for real data—We choose association rules combined with realistic basis to analyse realistic data set.

This paper selects the Global Terrorism Database (GTD), uses the K-Means clustering [9] algorithm to pre-process the casualty attributes of the original data set, and then selects the Apriori algorithm to find frequent itemset [10], and finally mines the association rules about six factors (time, location, attack method, attack target, weapon type, casualties).

2. Association Analysis based on Apriori Algorithm

2.1. Correlation Analysis

Association rules are commonly used and concrete manifestations of association analysis. They intuitively reflect the interrelationship and dependence between things. Association analysis was proposed for the shopping basket problem, and its purpose was to find out the correlation among the goods. The association rules are defined as follows: Suppose \( I = \{I_1, I_2, \ldots, I_m\} \) is a collection of item [11]; given a transaction database \( D \), where transaction \( t \) is a non-empty subset of \( I \); let \( X \) be an item set, and \( t \) contains \( X \); the association rule is \( X \Rightarrow Y \), where, \( X \subset I, Y \subset I, X \neq \emptyset, Y \neq \emptyset, X \cap Y = \emptyset \). The supporting formula is as in (1),

\[
\text{Support}(X \Rightarrow Y) = \frac{n(X,Y)}{n(\text{AllSamples})}
\]

(1)

where, \( n(X,Y) \) represents the number of times of a transaction that contains \( X \) and contains \( Y \) in \( D \), and \( n(\text{AllSamples}) \) represents the total number of transactions in \( D \). The confidence formula is as in (2),

\[
\text{Confidence}(X \Rightarrow Y) = \frac{P(X,Y)}{P(Y)}
\]

(2)

where, \( P(X,Y) \) represents the probability that a transaction contains \( X \) and \( Y \), and \( P(Y) \) represents the probability that a transaction contains \( Y \). If the minimum support threshold and the minimum confidence threshold are met, the association rules are considered meaningful [12]. The Apriori algorithm is a classical algorithm for mining association rules.

2.2. Space Considerations

According to the actual situation of China's terrorist attack, this paper gives a new method to do the Multi-factors analysis of terrorist attacks. This method is shown as Figure 1. The main process is as follows.

a) Data collection and pre-processing. It selects 6 factors about terrorist attacks in China from the Global Terrorism Database (GTD). The 6 factors are time, location, attack method, attack target, weapon type, and casualties. It gets the analysis data set base on the 6 factors, we call it as ‘Chinese terrorist attack association analysis dataset’. And we pre-processing on this dataset, such as missing value filling and denoising.

b) Discretization of casualty data. It used the number of casualties to clustering base on K-Means method. By processing, the number of casualties is converted into discrete data, and we got a standard Chinese terrorist attack association analysis dataset.

c) Association rules analysis. We do correlation analysis based on the standard Chinese terrorist attack correlation analysis dataset by using Apriori algorithm to obtain the terrorist attack correlation results.

![Figure 1. Multi-factors terrorist attacks analysis method](image-url)
3. Multi-factors Analysis on China’s Terrorist Attack

3.1. Data Collection and Pre-processing
GTD is a global open source database that records information about global terrorist incidents occurred since 1970. It includes GTD ID, incident time, incident location, attack method, attack target, weapon type, and casualties and so on. This paper selects data from 1970 to 2017. There are 302 records about China. After prepossessing, we get 292 records. And we select 6 important factors to analyse, which include time, incident location, attack method, attack target, weapon type, and casualties.

3.2. Discretization of Casualty Data
After we process the data set, there are still many missing values in the attribute of casualties. Because the number of casualties is related to the attack method, the death and injury fields are classified by attack method in this paper. It uses the average value of the death and injury to replace the missing value in the same attack method.

According to the "Emergency Response Law of the People's Republic of China", emergency warning levels are divided into 4 levels: extremely serious (level I), serious (Level II), major (level III) and general (level IV) [4]. In "Production Safety Accident Reporting and Investigation and Handling Regulations", it also classifies the emergency warning into 4 levels: 1-2, 3-9, 10-29 and >=30. So, this paper divides the terrorist attacks data into 4 levels based on casualties. To ensure the objectivity and accuracy, this paper using K-Means clustering method to divide the data.

In the dataset, some injured attributes are too large. To avoid the impact on the clustering results, 13 such records are classified as level I. Then we use K-means algorithm to clustering. The process is as follows:
a) Select 4 objects from death and injury data set \( X = \{x_i | i = 1, 2, \ldots, n\} \) as the cluster center points, i.e. \( \{\mu_1, \mu_2, \mu_3, \mu_4\} \);
b) Calculate the distance from each object to the 4 cluster center in the data based on the Euclidean distance, and Euclidean distance between two data object like \( x_i \) and \( x_j \) is \( D(x_i, x_j) \). The distance formula is as in (3),

\[
D(x_i, x_j) = \sqrt{(x_i - x_j)^2}
\] 

(3)

Then, it classifies the objects into the minimum distance cluster, and \( C_k (k = 1, 2, 3, 4) \) represents the center of mass of the cluster,

\[
C_k = \frac{1}{n_k} \sum_{x_i \in C_k} x_i
\] 

(4)

where, \( n_k \) is the number of samples of cluster \( k \), and \( x_i \) is the sample belonging to class \( C_k \).
c) Update each cluster center point, i.e. \( (C_1, C_2, C_3, C_4) \);
d) If the maximum number of iterations is reached or the 4 cluster center no longer change or the sum of the squared errors SSE is within a given range, terminates the algorithm; otherwise, continues to do steps b and c. SSE formula is as in (5),

\[
E = \sum_{j=1}^{4} \sum_{i \in C_j} |x_i - C_j|^2
\] 

(5)

The clustering results are shown as Fig. 2. In the Figure, green points represent level I, blue points represent level II, purple points represent level III, and yellow points represent level IV. Based on the clustering results, a "standard Chinese terrorist attack correlation analysis data set" is obtained, as shown as Table 1.
Table 1. Sample data set of terrorist attacks.

| Month | Location | Attack     | Target                 | Weapon | Level  |
|-------|----------|------------|------------------------|--------|--------|
| 4     | Zhejiang | Hijacking  | Airports & Aircraft    | Explosives | Level IV |
|       |          | Bombing/E  |                        |        |        |
| 6     | Shanghai | Bombing/E  | Transportation          | Explosives | Level I |
| 12    | Beijing  | Hijacking  | Airports & Aircraft    | Others | Level IV |
| 12    | Sichuan  | Bombing/E  | Transportation          | Explosives | Level III |
|       |          | Bombing/E  | Government              |        |        |
| 1     | Beijing  | Bombing/E  | (Diplomatic)            | Explosives | Level IV |

Figure 2. Clustering results of casualties.

4. Experimental Results and Analysis

In the association analysis, this paper sets the minimum support as 0.02 and the minimum confidence as 0.5, and the Apriori algorithm is used to obtain the association rules.

1) Table 2. shows the association rules of factors affecting the level of casualties. After analysing these rules, we get the following results.

a) From the association rules of location, month and casualty level, we can get three rules of terrorist attacks in Xinjiang. First rule is that the terrorist attacks in September are related to the level of special heavy casualties; Second rule is that the terrorist attacks in June is related to the level of large casualties; Third rule is related to the level of general casualties. According to relevant information, China Asia Europe Expo is conducted at September in Xinjiang. Meanwhile, the flow of people and traffic is huge. In order to attract large attention, terrorist attacks are likely to be carried out during the World Expo. Therefore, Xinjiang residents should improve their safety awareness and pay attention to strengthen their own safety protection in September.

b) From the association rules of attack and casualty level, it can be concluded that armed attack by incendiary weapons causes greater casualties than the assassination attack method with weapons as melee. Because incendiary bombs have large lethality, wide range of damage, and untargeted. Attacks with incendiary weapons are easy to carry out. But the assassination attacks are only for specific people or groups, and the target of assassination is not clear. So, relevant institutions should strengthen the control of raw materials for making incendiary weapons.

c) From the association rules of weapon, location, and casualty level, it can be concluded the conclusion: Even by the same weapon in an incendiary attack, the casualty level in Xinjiang is much higher than that in Taiwan. The reason may be that the climate in Xinjiang is dry and the climate in Taiwan is humid, and the damage caused by incendiary bombs lasts longer.

d) The association rule of location, target and casualty level shows that the terrorist attacks in Xinjiang will cause serious casualties. So, Xinjiang is the key location of anti-terrorism.
Table 2. Association rules affecting the level of casualties.

| Association rules of location, month, and casualty level          | {5, Xinjiang} —> {Level IV} |
|-----------------------------------------------------------------|-----------------------------|
| Association rules of attack and casualty level                  | {Armed Assault, Incendiary} —> {Level III} |
| Association rules of weapon, location, and casualty level       | {Incendiary, Xinjiang} —> {Level III} |
| Association rules of location, target, and casualty level       | {Private Citizens & Property, Xinjiang} —> {Level I} |

2) Table 3. shows the association rules related to time and location factors, and the following constructive conclusions can be obtained by analysing these rules.

a) From the analysis of the association rules for weapon types, time and location, it can be concluded that the explosive attacks are prone to occur in December of Guangzhou and February, May and July of Xinjiang. So, it should strengthen the inspection of explosive attack weapons in these months of Guangdong and Xinjiang. In conclusion, explosive terrorist attacks occur frequently, which may be related to the low cost and easy manufacture.

b) From the association rules of time, location and weapon, we can draw the following conclusion: the explosion attack in Xinjiang changes with time. The attacks are likely to take place at police stations in May and occur on public transport in July. A terrorist attack in these two locations could have a big impact. Therefore, Xinjiang security departments must strengthen the safety inspection of explosives near police stations in May and public transportation in July.

c) From the association rules of time, location, weapon and attack method, we can draw the following conclusion: most terrorist attacks used incendiary weapons during June in Xinjiang, while terrorists are more likely to use melee attacks during July. Therefore, it must be vigilant against large-scale crowd gathering to prevent organized terrorist attacks during June and July in Xinjiang.

Table 3. Rules related to time and location

| Association rules for time, location and weapon type           | {12, Guangdong} —> {Explosives} |
|----------------------------------------------------------------|-----------------------------|
| Association rules of time, location, weapon and target        | {5, Xinjiang} —> {Explosives} |
| Association rules of time, location, weapon and target        | {7, Xinjiang} —> {Explosives} |
| Association rules of time, location, weapon and target method | {5, Explosives, Xinjiang} —> {Police} |
| Association rules of time, location, weapon and target method | {7, Explosives, Xinjiang} —> {Transportation} |
| Association rules of time, location, weapon and attack method | {6, Incendiary, Xinjiang} —> {Armed Assault} |
| Association rules of time, location, weapon and attack method | {7, Melee, Xinjiang} —> {Armed Assault} |

3) In addition to the above rules, the meaningful rules excavated in this paper are as follows. ① {9, Incendiary, Taiwan} —> {Transportation}, ② {9, Facility/Infrastructure Attack, Taiwan} —> {Transportation}, ③ {12, Guangdong, Private Citizens & Property} —> {Bombing/Explosion}. The first rule indicates that incendiary bombs in Taiwan often attack the public transportation system in September. The second rule indicates that attacks on facilities often occur on the public transportation system of Taiwan in September. Therefore, Taiwan must strengthen the prevention and control of public transportation system security in September. The third rule indicates that explosion have occurred during December in Guangzhou. So, relevant departments in Guangdong need to strengthen explosion-proof inspection during the time.
Based on the above association rules, it can be concluded that explosion attacks occur frequently in terrorist attacks. Xinjiang, and Taiwan and Guangdong are the high incidence locations of terrorist attacks. Incendiary bomb is a common terrorist attack weapon. The common attacks target is the public transportation system. Based on the practical significance, the conclusions drawn from the filtered association rules can provide a strong theoretical reference for the anti-terrorism department.

5. Conclusion

Based on the analysis of China's terrorist attacks, this paper gives a multi-factor analysis method of terrorist attacks. Based on the actual attack data, the innovation of this method lies in the adoption of k-means method to combine the death attribute and the injury attribute. The clustering of these two can fully reflect the damage of people in terrorist attacks, and better reflect the impact of people in the event, and it uses data mining algorithm to analyse and obtain multiple association rules of terrorist attacks in China. These rules include the relationship among time, location, attack method, attack target, weapon type, and casualties. The analysis of these rules is described in detail in the 3 part of this paper. These conclusions can provide useful information for anti-terrorism departments. This paper only involves six factors in the multi-factor analysis. There are other factors that can be taken into account in the data set, such as loss of property, terrorist identity, etc., so that more comprehensive results can be obtained. The graph network composed of points and edges of such data sets will be the next research direction.

Acknowledgments

This paper is one of the research results of the General project of humanities and social science research of the Ministry of Education, "Group Behaviour Research and Coping Strategies under the Background of Video Big Data Based on Video Big Data" (project No. : 19YJA630046), and Tianjin Natural Science Foundation project, "Video Big Data Analysis and Research on Passenger Flow of Beijing, Tianjin and Hebei Airports" (project No. : 18JCYBC85100).

References

[1] Chengcong Wang, Yajing Liu, Mingyue Liu. Temporal and spatial evolution and situation analysis of global terrorist attacks [J]. Journal of earth information science,2019,21(11):1710-1720.
[2] Quan Xiao. Terrorist Attacks in Paris on 11·13 and their Impact [J]. Military Digest, 2016(01):21-24.
[3] Guohui Li. Research on temporal and Spatial evolution and Risk Analysis of Global Terrorist attacks [D]. Hefei: University of Science and Technology of China, 2014.
[4] Xiaoyan Sui. Research on temporal and Spatial variation characteristics and Influencing Factors of Terrorist attacks in China [D]. Dalian: Dalian University of Technology, 2017.
[5] Yongnan Li, Donglong Jiang, Jianming Mei. Application of decision tree based on gini coefficient in terrorism-related intelligence analysis [J]. Intelligence journal, 2017,36(04):29-32.
[6] Yongnan Li, Jianming Mei. Application research of Transcendental Principle in terrorism-related Intelligence Analysis [J]. Journal of Intelligence, 2017, 36(08):23-26.
[7] Hongshui Shu, Yanbiao Li. Analysis of Characteristics and Trends of Terrorist Attacks from the Perspective of Big Data [J]. Journal of Intelligence, 2019, 38(11): 87-93, 19.
[8] Cangyang Cheng. Analysis of hot spots of terrorist attacks in countries along the One Belt And One Road route based on GTD [D]. Beijing: people's public security university of China, 2019.
[9] Minghui Liu. Terrorism Risk Assessment of Civil Aviation System based on K-Means Cluster Analysis [J]. Data Analysis and Knowledge Discovery, 2012,2(10):21-26.
[10] Yangyang Zhang. Analysis of abnormal Wells in Oil Fields based on clustering Algorithm and Association rules [D]. Lanzhou University of Technology, 2019.
[11] Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques (3rd edition) [M], Fan Ming, MENG Xiaofeng, Trans. Beijing: Machine Press, 2012:158-159.
[12] Qun Jiang, Liwei Tian, Rongrong Li, et al. Research on Association Rule Mining in Smart City Big Data [J]. Modern Information Technology,2020(2):20-22.