

Risk Assessment of Terrorism in Asia based on K-Means Clustering Analysis

Lingzhi Jiang^{1,*}, Qiwu Wu², Zhuoxu Xie³

¹ College of Information Engineering, Engineering University of PAP, Xi'AN, Shaanxi, China

² College of Equipment Management and Support, Engineering University of PAP, Xi'AN, Shaanxi, China

³ College of Information Engineering, Engineering University of PAP, Xi'AN, Shaanxi, China

* Corresponding author: Lingzhi Jiang

Abstract: This article obtains terrorist attack data from some regions of the global terrorism database, and uses Python language to preprocess and describe the data. Cluster analysis is performed on the preprocessed data using the K-means algorithm. During the analysis process, the focus is on the two key variables of casualties and economic losses, in order to find the patterns and trends behind the data. Based on the clustering results, the level of danger in each region was scientifically graded to assist counter-terrorism forces in gaining a deeper understanding of the core information of terrorist attacks in these areas, thereby enhancing their awareness of terrorist attacks and enhancing their ability to prevent them.

Keywords: K-Means; Cluster Analysis; Terrorism; Risk Assessment.

1. Introduction

Since the 9•11 attacks, terrorist activities have become an important issue in the global security field, exerting profound impacts on the economic, social, and political stability of countries and regions. With the accumulation of global terrorism data, in order to help the public better understand the risk level of terrorist attacks and provide decision-makers, government agencies, and the public with an objective and actionable means to better understand and respond to the threat posed by terrorism, this article constructs a terrorist attack risk level assessment model based on k-means clustering analysis to evaluate and quantify the risk level of terrorist attacks. This article mainly combines relevant analysis and chi square test to construct a terrorist attack risk evaluation index system. Principal component analysis is used to reduce the dimensionality of the features that affect terrorist attack events. k-means clustering is used to cluster the characteristics of terrorist attack events. Similar events are classified into the same level, and the degree of harm of terrorist attack events from 1970 to 2021 is divided into three levels. Visualization technology is used to more intuitively display the risk levels of countries in the Middle East, Central Asia, and South Asia.

2. Organization of the Text Construct Risk Level Evaluation Indicators

The influencing factors of the risk level of terrorist attacks are multifaceted and complex, covering multiple fields such as politics, society, economy, geography, etc. These influencing factors are intertwined and affect each other, and comprehensive consideration is needed to more accurately evaluate and quantify the risk level of terrorist attacks. By consulting relevant literature, the influencing factors of the risk level of terrorist attacks mainly include five aspects: number of injuries, number of deaths, economic losses, social impact, and public psychology. Among them, number of injuries, number of deaths, and economic losses are the main influencing factors of the risk level of terrorist attacks.

Therefore, without considering the characteristic variables in the GTD dataset that have no or low impact on the risk level of terrorist attacks, such as year, month, and whether it is an extended event, 19 characteristics closely related to casualties and economic losses, such as the number of murderers (nperps), the number of captured murderers (nperpcap), the total number of deaths (nkill), and suicide attacks (suicide), will be preliminarily included in the risk level evaluation index system of this article. In addition, in order to further study the risk levels of terrorist attacks in various countries in the Middle East, Central Asia, and South Asia, the two characteristic attributes of country and region were retained.

3. Building a Quantitative Analysis Model for Terrorist Events

PCA was used to reduce the dimensionality of the constructed terrorist attack risk indicator variables and a gravel plot was used to help select the number of principal components, as shown in Figure 1. Using the cumulative variance contribution of principal components to further select the number of principal components, the variance contribution rate graph is shown in Table 1.

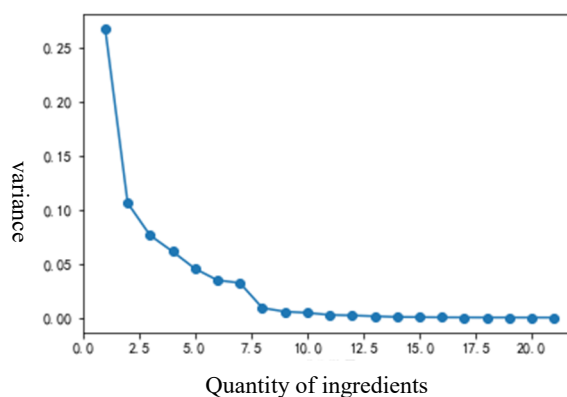


Figure 1. Crushed Stone Map

(1) Select the "inflection point" where the variance sharply

decreases in the gravel plot as the number of principal components. By observing Figure 1, it can be seen that when the number of principal components is 7, there is a clear inflection point in the gravel plot. Therefore, the number of principal components in this article is preliminarily set to 7.

(2) Based on (1), by observing Table 1, it is found that when

the number of principal components is 7, the cumulative variance contribution rate is 95.96%, which means that 7 principal components can explain 95.96% of all indicator variables. Therefore, the number of principal components in this article is determined to be 7.

Table 1. Principal Component Information

principal component	characteristic value	Variance contribution rate%	Accumulated variance contribution rate%
Principal Component 1	0.267358386	41.210387	41.210387
Principal Component 2	0.106014719e	16.341016	57.5514
Principal Component 3	0.0760631772	11.724311	69.2757
Principal Component 4	0.0611820649	9.430550	78.7062
Principal Component 5	0.0452645013	6.977031	85.6833
Principal Component 6	0.0346028836	5.333658	91.0169
Principal Component 7	0.0320668753	4.942760	95.9597

(3) Solve the principal component load and comprehensive score. The scores of each principal component and the

comprehensive score of the sample are shown in Table 2.

Table 2. Principal Component Score and Sample Comprehensive Score

Eventid	Principal Component 1	Principal Component 2	Principal Component 7	Score
197002280001	-0.371243458	0.054786202	-0.411553605	-1.254239946
197003140002	-0.373923556	-0.090034608	0.05805582	-0.926703805
197003290002	-0.364330344	0.04330642	-0.357646064	-1.02993341
.....
202106300016	0.465863753	-0.158740967	0.006467421	0.137994751
202106300020	-0.582897608	-0.008097875	-0.286957795	-0.377680903
202106300023	-0.596629791	0.014771726	-0.394786528	-0.824903889

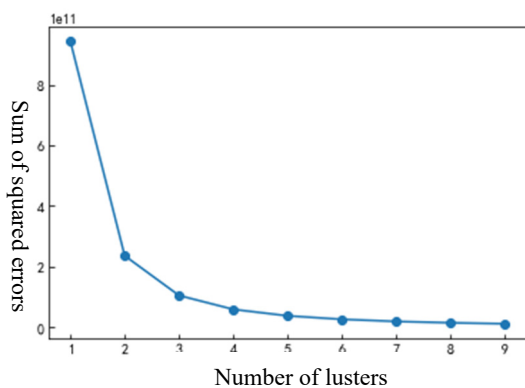


Figure 2. Elbow Rule Diagram

Input the characteristic data of each principal component score and sample comprehensive score obtained through principal component analysis into the K-means algorithm, and divide the data into different clusters. Select the appropriate number of terrorist attack risk levels through the

elbow rule, as shown in Figure 2.

Observing Figure 2, it can be seen that the optimal number of clusters for establishing a terrorist event classification model using k-means clustering is 3. Therefore, this article divides the risk level of terrorist events into three levels, and uses k-means clustering algorithm to cluster the scores of each principal component and the comprehensive score characteristics of the samples into three levels, representing the different risk levels of terrorist attacks from high to low, divided into one to three levels. The specific process is as follows:

(1) Initialization: First, randomly set three data points as the initial cluster centers.

(2) Assign data points: For all data points, calculate their distances to each centroid using Euclidean distance. According to equation (1), the specific Euclidean distance calculation formula can be determined, and then divide the data point into the cluster where the nearest cluster center is located.

$$d(o_i, o_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2} \quad (1)$$

(3) Update cluster center: Calculate the mean of all data within each specific cluster and set it as the new cluster center position.

(4) Repeat steps (2) and (3) until the cluster center no longer moves or reaches the maximum number of iterations

(5) Output result: Finally, 3 clusters were obtained, and the risk level was divided into 3 categories.

After k-means clustering, the number and proportion of events included in three levels are shown in Table 3.

Table 3. Number and proportion of events at different levels

Risk level	Number of events	Event proportion
1	10283	45.7%
2	2541	11.3%
3	9630	42.8%

List two terrorist attacks with high, medium, and low risk levels, as shown in Table 4.

Table 4. Examples of events at different levels

Risk level	Eventid	Event Summary
1	202105010051	An explosive device on a tanker truck exploded on a highway in Shakardara District, Kabul, Afghanistan. The explosion caused at least 7 deaths, 14 injuries, and 5 missing people. No group claims responsibility for this incident.
	201801270001	A suicide bomber detonated an ambulance filled with explosives at a police checkpoint outside Jomhuryat Hospital in Kabul, Afghanistan. Except for the suicide bomber, the explosion resulted in 103 deaths and 235 injuries. The Taliban claimed responsibility for this incident. In addition, the authorities also blamed the attack on the Haqqani network.
2	202105290010	The attackers attacked a security checkpoint in Analjoy, Kapisa Province, Afghanistan. In the subsequent conflict, the attackers fired a mortar and landed on a wedding. The attack resulted in at least 6 civilian deaths and 4 injuries. No group claims responsibility for the incident; However, sources attribute the attack to the Taliban, who deny any involvement.
	202004300014	A roadside explosive device was detonated in Khmeime, Homs, Syria, and attackers opened fire on a Syrian military bus. Six soldiers were killed in the attack. Iraq and the Islamic State of Levant (ISIL) claimed responsibility for the incident.
3	202106290004	In Agandi, Kabul, Afghanistan, attackers opened fire on a police car. At least two police officers were killed in the attack, and another person was injured. No group claims responsibility for this incident.
	202106180006	The attacker fired a rocket and landed on a house in the Jazzhan Akcha district of Afghanistan. The explosion resulted in two civilian deaths and two others being injured. No group claims responsibility for the incident; However, sources attributed the attack to the Taliban, who denied involvement and accused the government of carrying out the attack.

From the data shown in Table 4, it can be found that the differences in the risk levels of terrorist attacks are mainly reflected in personnel injuries, the number of deaths, and the degree of property damage. In addition, there is also a certain relationship with other factors, such as the target and location of the attack, the weapons used, and the means used. If the target of the attack is an important government agency, military base, or densely populated area, the risk level may be relatively high. For example, the attack of explosives outside hospitals in Afghanistan, as well as the attack on Syrian military buses, all have significant importance and therefore have a higher risk level. The weapons and methods used by attackers can also affect the risk level, for example, suicide bombing attacks are usually more destructive than typical

explosions or shootings, so the risk level may be higher. From Table 4, it can be seen that Afghanistan and Syria are currently facing high-risk terrorist attacks, which highlight the turbulent situation and ongoing security challenges in regions such as Afghanistan and Syria. At the same time, it indicates that the political and security environment of the region is also an important factor in assessing risk levels. If the region is in a state of turmoil, political instability, or intense conflict, the risk level of terrorist attacks is usually higher. For example, both Afghanistan and Syria are in a state of political turmoil and conflict, which increases the risk of terrorist attacks.

The data and proportion of terrorist attack risk events at all levels in the Middle East, Central Asia, and South Asia are shown in Table 5.

Table 5. Risk events at all levels in the region

region	First level quantity	First level proportion	Secondary quantity	Secondary proportion	Third level quantity	Third level proportion
Middle East	11254	90%	1136	10%	0	0%
Central Asia	29	100%	0	0%	0	0%
South Asia	0	0%	1405	12.8%	9630	87.2%

According to Table 5, there are a large number of high-risk events in the Middle East, indicating a higher risk of terrorist attacks in the region. This may be closely related to the long-standing political turmoil, conflicts, and terrorist activities in

the region. Terrorist organizations continue to operate in multiple countries in the Middle East, including ISIS and Al Qaeda. The scale of the attack may be large, with a high number of casualties. Although the number of high-risk

events in Central Asia is relatively small, this does not mean that the risk of terrorist attacks in the region is low. There may be potential terrorist activities in Central Asia, such as organizations associated with Islamic extremists, which could pose a threat to regional security. Political turmoil and ethnic conflicts are potential breeding grounds for terrorist activities. There are many intermediate risk events in South Asia, which are related to the political and religious turmoil in the region. Common terrorist organizations in South Asia include the Taliban, Indian terrorist organizations, etc. Especially countries such as Pakistan, Afghanistan, and India are often affected by terrorist attacks. South Asia is also one of the important hotspots of global terrorism activities.

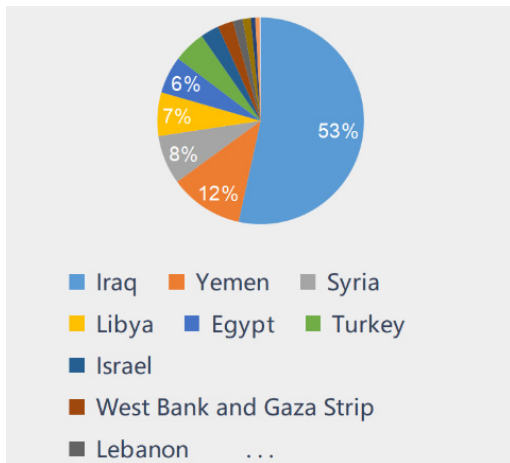


Figure 3. Distribution of the proportion of low-risk events in countries

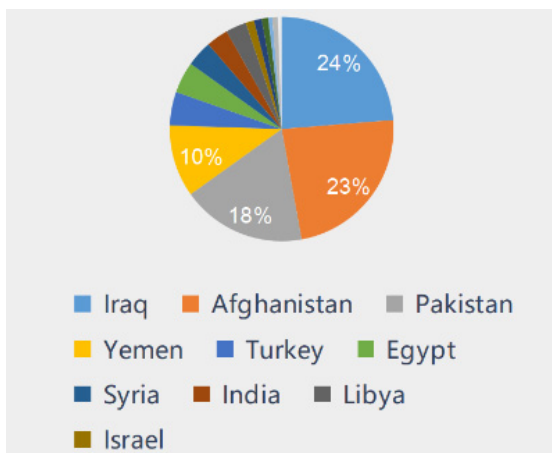


Figure 4. Distribution of the proportion of countries with intermediate risk events

The proportion of countries with low, medium, and high levels of terrorist attack risk is shown in Figures 3 to 5, respectively. Analyzing Figures 3 to 5, it can be seen that Afghanistan, Pakistan, and India are the main countries with a relatively high proportion of high-level risk events. This may be related to the long-standing political turmoil, conflicts, and terrorist activities in these countries. These countries often become the activity sites of various terrorist organizations, such as the Taliban and Al Qaeda. The long-term conflicts and political instability within the region have increased the likelihood of terrorist attacks occurring. Iraq, Afghanistan, Pakistan, and Yemen are the main countries for intermediate risk events. The geopolitical environment of

these countries is complex, experiencing long-term wars, internal turmoil, and terrorist attacks, and terrorist organizations have a large space for activity in these regions. Iraq, Yemen, Syria, Libya, and Egypt are the main countries for low-risk events. Although these countries still face challenges in terrorist attacks, their terrorist activities are relatively limited compared to high-level and intermediate risk countries. However, this does not mean that these countries do not have terrorism issues, but may be due to other factors such as government response to terrorism, the degree of regional conflict, etc.

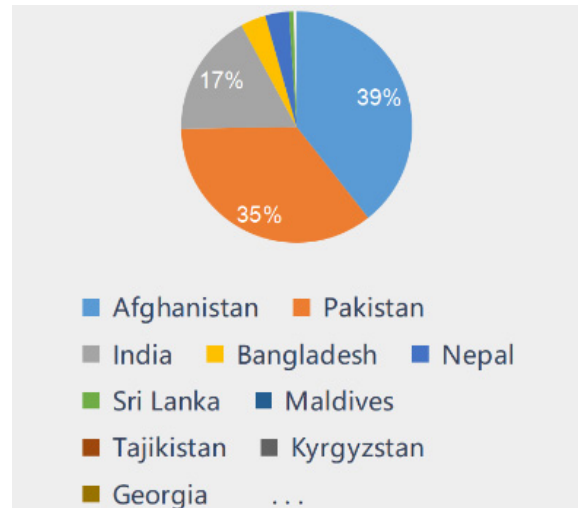


Figure 5. Distribution of the proportion of countries where high-level risk events occur

4. Summary

the risk level of terrorist attacks varies among different countries and is influenced by various factors, including political stability, geopolitical environment, existing terrorist organization activities, historical conflicts, etc. These factors collectively determine the level of terrorist attack risk that a country faces. The government and the international community need to closely monitor these issues and take effective measures to respond to and prevent terrorist activities to ensure regional and global security and stability.

References

- [1] Wu Zhenyu. Quantitative Analysis of Terrorist Attacks Based on Data Mining [D]. Xi'an University of Electronic Science and Technology, 2020 DOI: 10.27389/dnki.gxadu.2020.000585.
- [2] Liu Shuo, Li Yucong, Guo Weifeng. Research on the characteristics of terrorist attack targets along the "the Belt and Road" based on GTD database [J]. Intelligence Journal, 2020, 39 (07): 18-22.
- [3] Xu Kaiyang, Xin Wenfang. Quantitative Analysis of Terrorist Attack Record Data Based on Big Data [J]. Information Technology and Informatization, 2019, (08): 219-223.
- [4] Li Yongqun, Ying Wanming, Yuan Fei, etc Global terrorism database data analysis based on data mining [J]. Economic Mathematics, 2019, 36 (02): 91-94 DOI: 10.16339/j.cnki.hdjjx.2019.02.014.
- [5] Jiang Libao, Chen Yufan, Yu Lu, etc A Clustering Based Anti Terrorism Data Analysis Method [J]. Intelligence Exploration, 2019, (06): 74-77.