Risk assessment of terrorism related to sports events based on K-Means clustering

Liu Minghui

(School of Criminal Investigation and Counter Terrorism, People's Public Security University of China, Beijing 100038)

**Abstract** It is enormous of the impact of terrorist attacks on sports events, and it is of great significance for the analysis of the risk. An effective and stable method for risk assessment is desirable, so that the corresponding countermeasures will be taken accordingly. Using the GTD[[1]](#footnote-1) data of terrorist attacks from 1970 to 2016, this article gives the rank of risk objectively by the K-Means clustering method. The risk of different kinds of attack is given, and the computing process does not need manual intervention or scoring. And this article gives two more examples using K-Means clustering, one is to assess the risk of terrorism on the civil aviation system, and the other is to give new method to compute the rank of global terrorism. The results show that K-Means clustering is effective for risk assessment.

# Introduction

## 1.1 The police institution organization of China

The ministry of public security (MPS) is in charge of public security work throughout the country, and is the highest leading and commanding organ of the national public security work. The MPS is organized into the following departments:

Central Office,

Supervision,

Personnel & Training,

Public Relations,

Economic Crime Investigation,

Public Order Administration,

Border Control,

Criminal Investigation,

Exit & Entry Administration,

Fire Control,

Security Protection,

Public Information Network Security Supervision,

Penitentiary Administration,

Traffic Control,

Legal Affairs,

International Cooperation,

Equipment and Finance,

Drug Control,

Science & Technology,

**Counter-terrorism**, and

Info-communications.

Railway, navigation, civil aviation, forestry and anti-smuggling public security departments are under the dual leadership of their superior administration and the MPS.

## 1.2 The organization of this article

In this article, we introduce the main features of terrorist attacks related to sports events in Section 2, where a basic statistics about the incidents are given. The number of incidents, fatalities and injured per year, and that in corresponding countries, and that on attack types are analyzed in detail in this section. In Section 3, we describe the main idea of the K-Means method and apply it to risk assessment for sports events, where the detailed method is given and the results are shown. In the next section, we give two more applications of the K-Means clustering for risk assessment, one is to the civil aviation system, and the other is to compute the rank of global terrorism. All the applications reveal that this method is effective.

# The Features of terrorist attacks related to sports events

In GTD, the attack types contain the categories: assassination, hijacking, kidnapping, barricade incident, bombing/explosion, armed assault, unarmed assault, facility/infrastructure attack, and unknown. The recorded attack types for incidents related to sports events or sports stadium consist mainly the following: armed assault, assassination, bombing/explosion, facility/infrastructure attack, hijacking, and etc.

The records of incidents related to sports in GTD from 1970 to 2016 are in all 47, so the amount of incidents per year is not much, but the impact is enormous. The total number of incidents, fatalities and injured per year are shown in Table 1 and Figure 1.

Table 1: Number of incidents, fatalities and injured per year

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Total number of incidents | Total number of fatalities | Total number of injured |
| 1978 | 1 | 0 | 0 |
| 1980 | 2 | 4 | 0 |
| 1981 | 1 | 1 | 0 |
| 1982 | 2 |  |  |
| 1985 | 3 | 0 | 2 |
| 1986 | 1 |  |  |
| 1987 | 3 | 0 | 0 |
| 1991 | 2 | 1 | 0 |
| 1992 | 4 | 18 | 2 |
| 1994 | 2 | 0 | 3 |
| 1995 | 2 | 0 | 0 |
| 1996 | 3 | 0 | 0 |
| 2000 | 1 | 0 | 6 |
| 2001 | 2 | 10 | 1 |
| 2005 | 1 | 0 | 0 |
| 2008 | 6 | 11 | 52 |
| 2009 | 1 | 1 | 16 |
| 2010 | 4 | 2 | 12 |
| 2011 | 1 | 0 | 0 |
| 2012 | 0 | 1 | 0 |
| 2013 | 1 | 2 | 9 |
| 2014 | 4 | 1 | 0 |
| Total | 47 | 52 | 103 |

Figure 1: Number of incidents, fatalities and injured per year, where “Times” represents total number of incidents “nkill” represents that of fatalitiesand “nwound” represents that of injured.

The countries involving in the incidents in GTD records are shown in Table 2. We can see that Colombia, France, Spain, and United Kingdom suffer the largest number of incidents, and the number of fatalities in Colombia, Algeria and Ethiopia are the most.

Table 2: Total number of incidents, fatalities and injured in corresponding countries

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total number of incidents | Total number of fatalities | Total number of injured |
| Algeria | 1 | 9 | 0 |
| Austria | 1 | 0 | 0 |
| Chile | 1 | 0 | 0 |
| China | 1 | 0 | 0 |
| Colombia | 5 | 19 | 0 |
| Egypt | 1 | 0 | 2 |
| El Salvador | 2 | 0 | 0 |
| Ethiopia | 1 | 7 | 24 |
| France | 5 | 0 | 0 |
| Germany | 3 | 0 | 3 |
| Guatemala | 1 | 2 | 0 |
| India | 1 | 0 | 0 |
| Iran | 1 | 0 | 6 |
| Iraq | 3 | 2 | 24 |
| Israel | 2 | 0 | 0 |
| Pakistan | 2 | 4 | 13 |
| Peru | 2 | 0 | 0 |
| Somalia | 0 | 1 | 0 |
| Spain | 4 | 0 | 2 |
| Syria | 1 | 1 | 0 |
| Thailand | 2 | 1 | 16 |
| Turkey | 1 | 2 | 0 |
| United Kingdom | 4 | 1 | 1 |
| Yemen | 1 | 2 | 12 |
| Yugoslavia | 1 | 1 | 0 |

The number of incidents on attack types is shown in Table 3, where it is readily seen that bombing/explosion is used most frequently in terrorist attacks. The total number of incidents, fatalities and injured on attack types are shown in Figure 2-Figure 4.

Table 3: Number of incidents, fatalities and injured on attack types

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total number of incidents | Total number of fatalities | Total number of injured |
| Armed Assault | 7 | 16 | 3 |
| Assassination | 3 | 8 | 13 |
| Bombing/Explosion | 30 | 28 | 84 |
| Facility/Infrastructure Attack | 6 | 0 | 3 |
| Hijacking | 1 | 0 | 0 |
| **Total** | **47** | **52** | **103** |

Figure 2: Number of incidents on attack types

Figure 3: Number of fatalities on attack types

Figure 4: Number of injured on attack types

# K-Means clustering model for risk assessment

## 3.1 The main idea of K-Means clustering

K-Means clustering aims to partition *n* observations into *K* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The results of cluster can be criteria for the ranking of risk.

Given a set of observations, where each observation  is a *d*-dimensional vector, K-Means clustering aims to partition the *n* observations into *K* sets so as to minimize the within-cluster sum of squares. Formally, the objective is to find the cluster, such that

,

where  is the mean of points in .

## 3.2 Application to risk assessment for sports events

Though the data set for terrorist attack on sports events is not very large, we present a computing model to illustrate the K-Means method for risk assessment.

**Step 1:** Specify the set of observations. There are 5 kinds of attack types, so it is readily seen that *n*=5, and each observation  is a 3-dimensional vector with the form (Total number of incidents, Total number of fatalities, Total number of injured). Because the number of incidents, fatalities and injured are not in the same scale, we change the data to the same scale by

,

where  denotes the *j*-th element of vector . So the new  represents the relative number of incidents, fatalities and injured, and the results are shown in Table 4.

Table 4: Relative number of incidents, fatalities and injured

|  |  |  |  |
| --- | --- | --- | --- |
|  | *j*=1 (Relative number of incidents) | *j*=2 (Relative number of fatalities) | *j*=3 (Relative number of injured) |
| *i*=1 (Armed Assault) | 0.1489 | 0.3077 | 0.0291 |
| *i*=2 (Assassination | 0.0638 | 0.1538 | 0.1262 |
| *i*=3 (Bombing/Explosion) | 0.6383 | 0.5385 | 0.8155 |
| *i*=4 (Facility/Infrastructure Attack) | 0.1277 | 0 | 0.0291 |
| *i*=5 (Hijacking) | 0.0213 | 0 | 0 |

**Step 2**: Determine the number of clusters. The risk level is classified into high, medium and low, so the number of clusters should be 3, that is *k*=3.

**Step 3**: Select 3 vectors randomly as the center of the initial cluster.

**Step 4**: Calculate the distance from all the observation points to the centroids of each cluster. The centroid of cluster  is defined by the mean

 .

**Step 5**: Assign each observation point to the nearest cluster (the distance between the observation points to the cluster is defined as the distance to the centroid of the cluster).

**Step 6**: Calculate the new means to be the centroids of the observations in the new clusters.

**Step 7**: Repeat steps 4 to 6 until the clusters no longer change or the maximum number of iterations is reached.

Using the steps 1~7, the data ~are clustered into 3 groups, as is shown in Table 5.

Table 5: Results of clusters

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| Cluster number | 3 | 3 | 2 | 1 | 1 |

The centroids of the three clusters are (0.0745,0,0.0146), (0.6383,0.5385, 0.8155) and (0.1064,0.2308,0.0777), respectively. And the distances from the origin are 0.0759, 1.1672 and 0.2657, respectively, so they correspond to risk level low, high and medium. The results are shown in Figure 5 and Table 6.



Figure 5 Risk assessment of terrorism related to sports events, where “\*” represents low level, “+” represents medium level, and “o” represents high level.

Table 6: Risk assessment of terrorism related to sports events

|  |  |
| --- | --- |
|  | Rank of Risk |
| Armed Assault | 3/Medium |
| Assassination | 3/ Medium |
| Bombing/Explosion | 2/High |
| Facility/Infrastructure Attack | 1/Low |
| Hijacking | 1/Low |

# More applications

In this section, we give two more applications of the K-Means clustering for risk assessment, one is to the civil aviation system, and the other is to compute the rank of global terrorism.

## 4.1 Application to civil aviation system

Using the data from 1992 to 2015 in GTD, we can apply the K-Means clustering to assess the risk of civil aviation system. Firstly, we specify the set of observations. There are 8 kinds of attack types, so *n*=8, and each observation  is a 3-dimensional vector with the form (Total number of incidents, Total number of fatalities, Total number of injured). Because the number of incidents, fatalities and injured are not in the same scale, we change the data to the same scale by

,

where  denotes the *j*-th element of vector . So the new  represents the relative number of incidents, fatalities and injured, and the results are shown in Table 7. And using the steps 1~7 in the former section, we can obtain the rank of terrorism risk on aircraft.

 The centroids of the three clusters are (0.0256，0.0231，0.0556), (0.1667，0.7660，0.3472) and (0.6543，0.0724，0.2639), respectively. And the distances from the origin are 0.0654, 0.7092 and 0.8574, respectively, so they correspond to risk level low, medium and high. The results are shown in Figure 6 and Table 8.

Table 7: Relative number of incidents, fatalities and injured

|  |  |  |  |
| --- | --- | --- | --- |
|  | *j*=1 (Relative number of incidents) | *j*=2 (Relative number of fatalities) | *j*=3 (Relative number of injured) |
| *i*=1 (Armed Assault) | 0.1235  | 0.0109  | 0.3611  |
| *i*=2 (Assassination) | 0.0062  | 0.0000  | 0.0278  |
| *i*=3 (Bombing/Explosion) | 0.1667  | 0.7660  | 0.3472  |
| *i*=4 (Facility/Infrastructure Attack) | 0.0062  | 0.0000  | 0.0000  |
| *i*=5 (Hijacking) | 0.6543  | 0.0724  | 0.2639  |
| *i*=6 (Hostage Taking (Barricade Incident)) | 0.0062  | 0.1484  | 0.0000  |
| *i*=7 (Hostage Taking (Kidnapping)) | 0.0185  | 0.0000  | 0.0000  |
| *i*=8 (Unarmed Assault) | 0.0000  | 0.0000  | 0.0000  |



Figure 6 Rank of terrorism risk on aircraft, where “\*” represents low level, “o” represents medium level, and “+” represents high level.

Table 8: Rank of terrorism risk on aircraft

|  |  |
| --- | --- |
| Armed Assault | Low |
| Assassination | Low |
| Bombing/Explosion | High |
| Facility/Infrastructure Attack | Low |
| Hijacking | Medium |
| Hostage Taking (Barricade Incident) | Low |
| Hostage Taking (Kidnapping) | Low |
| Unarmed Assault | Low |

## 4.2 Application to rank of global terrorism

The Global Terrorism Index (GTI) is a report published annually by the Institute for Economics and Peace (IEP). It produces a composite score in order to provide an ordinal ranking of countries on the impact of terrorism using the method in Appendix. In this article, we give another model for computing the rank of terrorism risk using the K-Means clustering.

Firstly, we specify the set of observations. There are 221 countries in GTD records until the year 2014, so *n*=221, and each observation  is a 3-dimensional vector with the form (Total number of incidents, Total number of fatalities, Total number of injured).

To summarize the incidents in the past four years, we define

 .

For the’s differ very much, we define the new  as

 

and apply the K-Means method to the new ’s.

The ranking of the risk is represented by A~E from the highest to the lowest level, and some of the results are shown in Table 9. Besides, the countries ranking “A” are shown in Table 10.

Table 9: The terrorism ranking of some countries

|  |  |
| --- | --- |
| Afghanistan | A |
| Albania | D |
| Algeria | B |
| Andorra | E |
| Angola | E |
| Antigua and Barbuda | E |
| Argentina | D |
| Armenia | D |
| Australia | C |
| Austria | D |
| Azerbaijan | D |
| Bahamas | E |
| Bahrain | C |
| Bangladesh | B |
| Barbados | E |
| Belgium | C |
| Belize | E |
| Benin | E |
| Bermuda\* | E |
| Bhutan | D |

Table 10: The terrorism ranking with level “A”

|  |  |
| --- | --- |
| Egypt | A |
| India | A |
| Iraq | A |
| Libya | A |
| Nigeria | A |
| Pakistan | A |
| Philippines | A |
| Somalia | A |
| Syria | A |
| Thailand | A |
| Ukraine | A |
| Yemen | A |

# Conclusion

This paper analyzes briefly the main attack types and casualties of terrorism related to sports events, and makes quantitative and objective assessment of the risk by using the K-Means clustering method. The assessment results are reliable. The analysis process of the method does not need manual intervention. It can cluster according to statistical data intelligently and can be applied to other problems. We present two more applications in the end, and the method performs effectively.

# Appendix: GTI by the institute for economics & peace

The GTI ranks the countries based on four indicators weighted over five years. A country’s annual GTI score is based on a unique scoring system to account for the relative impact of incidents in the year. The four factors counted in each country’s yearly score are:

total number of terrorist incidents in a given year

total number of fatalities caused by terrorists in a given year

total number of injuries caused by terrorists in a given year

a measure of the total property damage from terrorist incidents in a given year.

To assign a score to a country each incident is rated according to the four measures. The measures are then multiplied by their weighting factor and aggregated:

|  |  |
| --- | --- |
| Dimension | Weight |
| Total number of incidents | 1 |
| Total number of fatalities | 3 |
| Total number of injuries  | 0.5 |
| Sum of property damages measure | Between 0 and 3 depending on severity |

Then the total raw score for one year can be calculated, and the weight for five-year weighted average is given by

|  |  |
| --- | --- |
| Year | Weight |
| Current year | 16 |
| Previous year | 8 |
| Two years ago | 4 |
| Three years ago | 2 |
| Four years ago | 1 |

In the end, the GTI uses a base 10 logarithmic banding system between 0 and 10 at 0.5 intervals. It maps all values to a band of size 0.5 within the scale of 0-10.

# References

Institute for economics & peace (2018). *Global terrorism index*. <http://www.visionofhumanity.org/reports>.

Wagstaff, K., Cardie, C., & Rogers, S. (2001). Constrained K-means Clustering with Background Knowledge. *Eighteenth International Conference on Machine Learning* (pp.577-584). Morgan Kaufmann Publishers Inc.

Arai, K., & Ali, R. B. (2007). Hierarchical k-means : an algorithm for centroids initialization for k-means. *Reports of the Faculty of Science & Engineering Saga University,* *36*, 25-31.

1. The Global Terrorism Database (GTD) is an open-source database including information on terrorist events around the world, which can be found on the website http://www.start.umd.edu/gtd/. [↑](#footnote-ref-1)