Received: 7 April 2020 | Revised: 13 July 2020 | Accepted: 12 August 2020

DOI: 10.4218/etrij.2020-0138

ORIGINAL ARTICLE

ETRI Journal WILEY

Automatic and objective gradation of 114 183 terrorist attacks using a machine learning approach

Wanle Chi[1,2,3] | **Yihong Du**[1]

[1]Department of Information Technology, Wenzhou Polytechnic, Zhejiang, China

[2]Faculty of Information and Communication Technology, Technical University of Malaysia Malacca, Malacca, Malaysia

[3]Wenzhou University, Zhejiang, China

Correspondence
Wanle Chi, Department of Information Technology, Wenzhou Polytechnic, Zhejiang, China.
Email: 358455713@qq.com

Funding information
The Scientific Research Project of Zhejiang Provincial Education Department, Grant/Award Number: Y 201329845); the Scientific Research Project of Wenzhou science and Technology Bureau; Scientific Research Project of Zhejiang Natural Science Foundation, P.R.China, Grant/Award Number: LY20G030018; Scientific Research Project of Wenzhou Science and Technology Bureau, P.R. China, Grant/Award Number: 2013R0017

Catastrophic events cause casualties, damage property, and lead to huge social impacts. To build common standards and facilitate international communications regarding disasters, the relevant authorities in social management rank them in subjectively imposed terms such as direct economic losses and loss of life. Terrorist attacks involving uncertain human factors, which are roughly graded based on the rule of property damage, are even more difficult to interpret and assess. In this paper, we collected 114 183 open-source records of terrorist attacks and used a machine learning method to grade them synthetically in an automatic and objective way. No subjective claims or personal preferences were involved in the grading, and each derived common factor contains the comprehensive and rich information of many variables. Our work presents a new automatic ranking approach and is suitable for a broad range of gradation problems. Furthermore, we can use this model to grade all such attacks globally and visualize them to provide new insights.

KEYWORDS
Automatic rank, gradation, machine learning, terrorist attacks

1 | INTRODUCTION

Natural and man-made disasters are catastrophic events, leading to loss of life, damage to property, and huge social impacts [1,2]. To set common standards and facilitate international communications regarding disasters, the relevant authorities rank and group them on multiple levels in terms of overall impact, which are always important tasks in social management.

Pre-specified criteria imposing uniformity in all cases are formulated subjectively on only two rules, direct economic losses and loss of life, in most accidents and natural disasters. These two rules play leading roles in assessing the overall impact, but they fail to make the best of available information, and many other factors in the accident should also be jointly considered in complex natural and social environments [3]. Subjectively assigned factors when assessing events are incomplete and conform to the opinions of people in some nations or in those periods to some degree [4].

Man-made disasters are more uncontrollable with uncertain human factors such as ethics, policies, economies, and even laws. Terrorism, which is violent conflict with specific purposes carried out in various ways, is endowed with strong intuitions, which make it difficult to assess the overall impact and rank the severity [5]. The extent of terrorist attacks is roughly ranked only by the rule of direct economic losses into four levels: catastrophic, major, minor, and unknown [6].

This is an Open Access article distributed under the term of Korea Open Government License (KOGL) Type 4: Source Indication + Commercial Use Prohibition + Change Prohibition (http://www.kogl.or.kr/info/licenseTypeEn.do).
1225-6463/$ © 2021 ETRI

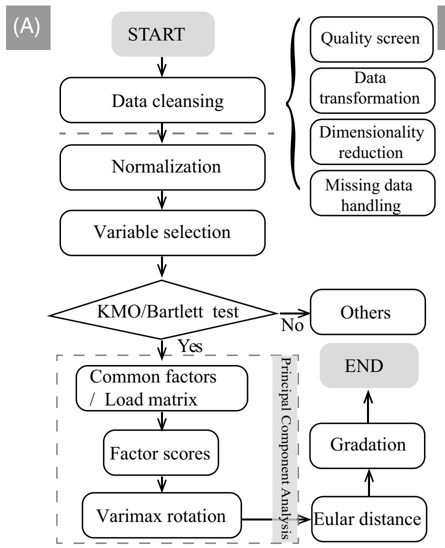
Machine learning requires a set of data that can analyze the information and provide confident solutions. In this research, the Global Terrorism Database (GTD) is used for the dataset. The GTD currently records massive open-source data on terrorist attacks across the world, and the baseline information of each terrorist attack is arranged into 135 items such as the tempo-spatial information, fatalities, terrorist targets, tactics, and weapons employed by the terrorists [7–9]. All 135 items contribute to a comprehensive and detailed description of an attack from all facets [10]. The continuous accumulation and large number of terrorist attack events make it possible to predict the behaviors and even the purpose of terrorists [11,12]. Terrorism research is moving toward internationalizing quantitative criminology [13].

Lee focused on artificial intelligence techniques to visualize and predict possible terrorist attacks using classification models, decision trees, and random forests [14]. Qiu used six machine learning models to predict global terrorist attacks. The results show that the random forest model, K-nearest neighbors model, and decision tree model, which had the highest R-squared value, perform well [15]. Verma presented three predictive models: attack type, attack region, and weapon type predictive models, which classify attack type, attack region, and weapon type based on millions of attacks using various supervised machine learning algorithms [16].

In this paper, we collected 110 000 records of terrorist attacks and performed a series of preprocessing steps, including data cleansing and integrity steps. To avoid the influence of noise on the experimental results, the data dimensionality was reduced. Subsequently, we used machine learning approaches to select seven variables to describe the overall impact and reduced the number of dimensions into three independent basic structures with principal component analysis and factor analysis [17–20]. In this space, the Euler distance is used to grade attacks in an automatic and objective way.

2 | PREPROCESSING AND VARIABLE SELECTION

We consider all attack events, that is, 114 183 records with 135 items each, globally within the time period from 1998 to 2017 in the GTD. The dataset combines and records data from a variety of collection institutions, leading to considerable data noise and missing values. We apply the data cleansing and information retrieval methods described below to extract



(B)

Item	Variable: occupation	Item	Variable: occupation	Item	Variable: occupation
1	eventid: 100.00% > 10%	46	targsubtype2_txt: 8.86%	91	weaptype3_txt: 1.26%
2	iyear: 100.00% > 10%	47	corp2: 8.44%	92	weapsubtype3: 1.17%
3	imonth: 100.00% > 10%	48	target2: 9.06%	93	weapsubtype3_txt: 1.17%
4	iday: 100.00% > 10%	49	natlty2: 8.88%	94	weaptype4: 0.00%
5	approxdate: 8.06%	50	natlty2_txt: 8.88%	95	weaptype4_txt: 0.00%
6	extended: 100.00% > 10%	51	targtype3: 0.90%	96	weapsubtype4: 0.01%
7	resolution: 0.80%	52	targtype3_txt: 0.90%	97	weapsubtype4_txt: 0.01%
8	country: 100.00% > 10%	53	targsubtype3: 0.86%	98	weapdetail: 47.16% > 10%
9	country_txt: 100.00% > 10%	54	targsubtype3_txt: 0.86%	99	nkill: 96.25% > 10%
10	region: 100.00% > 10%	55	corp3: 0.81%	100	nkillus: 99.68% > 10%
11	region_txt: 100.00% > 10%	56	target3: 0.90%	101	nkillter: 97.66% > 10%
12	provstate: 99.99% > 10%	57	natlty3: 0.88%	102	nwound: 92.79% > 10%
13	city: 99.62% > 10%	58	natlty3_txt: 0.88%	103	nwoundus: 99.49% > 10%
14	latitude: 99.18% > 10%	59	gname: 100.00% > 10%	104	nwoundte: 96.05% > 10%
15	longitude: 99.18% > 10%	60	gsubname: 2.78%	105	property: 100.00% > 10%
16	specificity: 99.99% > 10%	61	gname2: 1.63%	106	propextent: 41.02% > 10%
17	vicinity: 100.00% > 10%	62	gsubname2: 0.13%	107	propextent_txt: 41.02% > 10%
18	location: 47.86% > 10%	63	gname3: 0.27%	108	propvalue: 25.30% > 10%
19	summary: 100.00% > 10%	64	gsubname3: 0.02%	109	propcomment: 48.05% > 10%
20	crit1: 100.00% > 10%	65	motive: 43.33% > 10%	110	ishostkid: 100.00% > 10%
21	crit2: 100.00% > 10%	66	guncertain1: 99.67% > 10%	111	nhostkid: 8.19%
22	crit3: 100.00% > 10%	67	guncertain2: 1.58%	112	nhostkidus: 8.14%
23	doubtterr: 100.00% > 10%	68	guncertain3: 0.27%	113	nhours: 2.28%
24	alternative: 16.18% > 10%	69	individual: 100.00% > 10%	114	ndays: 6.10%
25	alternative_txt: 16.18% > 10%	70	nperps: 88.47% > 10%	115	divert: 0.14%
26	multiple: 100.00% > 10%	71	nperpcap: 97.05% > 10%	116	kidhijcountry: 0.19%
27	success: 100.00% > 10%	72	claimed: 100.00% > 10%	117	ransom: 8.99%
28	suicide: 100.00% > 10%	73	claimmode: 16.40% > 10%	118	ransomamt: 0.59%
29	attacktype1: 100.00% > 10%	74	claimmode_txt: 16.40% > 10%	119	ransomamtus: 0.49%
30	attacktype1_txt: 100.00% > 10%	75	claim2: 1.63%	120	ransompaid: 0.53%
31	attacktype2: 5.48%	76	claimmode2: 0.52%	121	ransompaidus: 0.48%
32	attacktype2_txt: 5.48%	77	claimmode2_txt: 0.52%	122	ransomnote: 0.44%
33	attacktype3: 0.37%	78	claim3: 0.27%	123	hostkidoutcome: 8.18%
34	attacktype3_txt: 0.37%	79	claimmode3: 0.11%	124	hostkidoutcome_txt: 8.18%
35	targtype1: 100.00% > 10%	80	claimmode3_txt: 0.11%	125	nreleased: 7.97%
36	targtype1_txt: 100.00% > 10%	81	compclaim: 4.18%	126	addnotes: 23.85% > 10%
37	targsubtype1: 93.84% > 10%	82	weaptype1: 100.00% > 10%	127	scite1: 99.95% > 10%
38	targsubtype1_txt: 93.84% > 10%	83	weaptype1_txt: 100.00% > 10%	128	scite2: 66.30% > 10%
39	corp1: 90.93% > 10%	84	weapsubtype1: 92.04% > 10%	129	scite3: 37.33% > 10%
40	target1: 99.85% > 10%	85	weapsubtype1_txt: 92.04% > 10%	130	dbsource: 100.00% > 10%
41	natlty1: 99.01% > 10%	86	weaptype2: 8.49%	131	INT_LOG: 100.00% > 10%
42	natlty1_txt: 99.01% > 10%	87	weaptype2_txt: 8.49%	132	INT_IDEO: 100.00% > 10%
43	targtype2: 9.10%	88	weapsubtype2: 7.60%	133	INT_MISC: 100.00% > 10%
44	targtype2_txt: 9.10%	89	weapsubtype2_txt: 7.60%	134	INT_ANY: 100.00% > 10%
45	targsubtype2: 8.86%	90	weaptype3: 1.26%	135	related: 15.80% > 10%

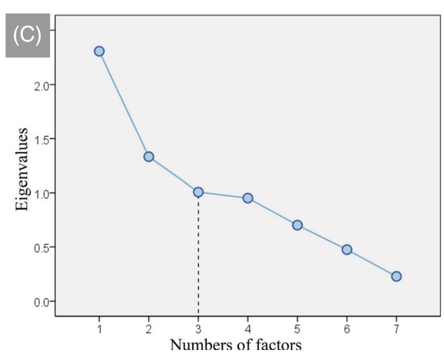


FIGURE 1 (A) Flow chart of data processing. (B) Visualization of the proportion of occupied values for all 135 items in the GTD. (C) Scree plot of eigenvalues for the correlation coefficient matrix

TABLE 1 Primary variables describing an attack

Variable	Categories
Multiple	Incident information
Success	Attack information
Nperps	Perpetrator information
Nhostkid	Casualties and consequences
Nkill	Casualties and consequences
Nwound	Casualties and consequences
Propextent	Casualties and consequences

structured data from raw data that are not readily available (the flow chart is shown in Figure 1A). We first screened for quality and tightened the scope to 95 710 incidents with the DoubtTerrorismProper restriction doubtterr = 0, which means there is essentially no doubt that the incident is an act of terrorism. Subsequently, we removed items that are less than 10% occupied, as shown in Figure 1B, since they are mostly less informative. For those remaining items, missing values are also very common. We then used statistical attributes (such as mean and median) to fill in the missing values for "NaN" and blanks, since terrorist attacks are a series of unethical events with related and similar characteristics, and the data may obey a certain distribution in the broad sense. To avoid the impact of special extreme event error processing, we use the correlation and class mean interpolation method to eliminate singularities. Finally, to consider measurement metrics and the dimensions of different variables, we normalized each variable with a standard deviation equal to 1 and a mean value equal to 0 to ensure comparability among them.

Considering the consistency of the range of dataset changes and effective metrics, to make the variables more representative and distinguishable, we investigated the frequency statistics of variables with high occupation proportions and visualize their underlying distribution in the supplementary materials. We pursued maximum identification and picked the following seven primary variables to describe an attack in a comprehensive and diverse manner, as also shown in Table 1 (the right column indicates the categories of the GTD sets [10]):

- "Multiple" indicates that a particular attack was part of multiple incidents and implies that organized crimes committed in mature operations bring huge potential dangers.
- "Success" shows the incident was successful and caused the damage the criminals expected.
- "Nperps" stores the total number of terrorists participating in the incident.
- "Nhostkid" counts the number of hostages or kidnapping victims.
- "Nkill" records the number of total confirmed fatalities for the incident.
- "Nwound" counts confirmed non-fatal injuries in the attack. These four variables reflect the severity objectively from different aspects.
- "Propextent" describes the extent of the property damage.

Correlation coefficients between these seven variables are derived in Table 2 and they are closely interrelated, because they pass the KMO and Bartlett's tests, which indicate that factor analysis is suitable for reducing dimensions and

TABLE 2 Correlation coefficients among the seven variables

	Nperps	Nkill	Nwound	Nhostkid	Multiple	Success	Proextent
Correlation coefficients							
Nperps	1	0.046	0.058	−0.016	0.014	0.026	0.328
Nkill	0.046	1	0.687	0.642	0.141	0.025	0.244
Nwound	0.058	0.687	1	0.521	0.11	0.025	−0.004
Nhostkid	−0.016	0.642	0.521	1	0.13	0.025	0.015
Multiple	0.014	0.141	0.11	0.13	1	0.035	0.088
Success	0.026	0.025	0.025	0.025	0.035	1	0.02
Propextent	0.328	0.244	−0.004	0.015	0.088	0.02	1
Significance							
Nperps		0.255	0.204	0.41	0.419	0.354	0
Nkill	0.255		0	0	0.022	0.362	0
Nwound	0.204	0		0	0.058	0.361	0.475
Nhostkid	0.41	0	0		0.031	0.363	0.416
Multiple	0.419	0.022	0.058	0.031		0.308	0.103
Success	0.354	0.362	0.361	0.363	0.308		0.388
Propextent	0	0	0.475	0.416	0.103	0.388	

TABLE 3 Factor load matrix after varimax rotation

Descriptive variable	f_1	f_2	f_3
Nperps	−0.020	−0.796	0.009
Nkill	0.894	0.168	0.077
Nwound	0.846	−0.007	0.052
Nhostkid	0.829	−0.061	0.077
Multiple	0.179	0.091	0.473
Success	−0.179	−0.042	0.897
Propextent	0.073	0.824	0.070

extracting the principal components. We adopted the principal component method. Extracting three common factors has an optimal margin of the total variance information of the original variables, and the initial solutions are shown in Figure 1C. Our model is suitable for predicting each of the seven observable variables from the values of three unobservable common factors.

3 | MODELING

For a deeper understanding and a simplified expression of factors, we rotate the orthogonal basis to align with the actual coordinate system. Assuming that $\gamma_1, \gamma_2, \ldots, \gamma_p$ are p (p equals 7 in this paper), the eigenvalues for the correlation coefficient matrix, $\eta_1, \eta_2, \ldots, \eta_p$ are the corresponding standard orthogonal eigenvectors. The load matrix **A** of the principal component factor analysis of the sample correlation coefficient matrix **R** is

$$\mathbf{A} = \left(\sqrt{\gamma_1}\eta_1, \sqrt{\gamma_2}\eta_2, \ldots, \sqrt{\gamma_m}\eta_m\right). \tag{1}$$

The variances of specific factors δ_i^2 are estimated using the expression $\mathbf{R} - \mathbf{A}\mathbf{A}^{\mathrm{T}}$ as follows:

$$\delta_i^2 = 1 - \sum_{j=1}^{m} a_{ij}^2. \tag{2}$$

Here, we adopted the Kaiser varimax rotation method and obtained convergence after four iterations [21]. Load matrix **A** of the analytic solutions is shown in Table 3 and listed as the following system of equations.

$$\text{Nperps} = -0.020f_1 + 0.796f_2 + 0.009f_3. \tag{3}$$

$$\text{Nkill} = 0.894f_1 + 0.168f_2 + 0.077f_3. \tag{4}$$

$$\text{Nwound} = 0.846f_1 - 0.007f_2 + 0.052f_3. \tag{5}$$

$$\text{Nhostkid} = 0.829f_1 - 0.061f_2 + 0.077f_3. \tag{6}$$

$$\text{Multiple} = 0.179f_1 + 0.091f_2 + 0.473f_3. \tag{7}$$

$$\text{Success} = -0.079f_1 - 0.042f_2 + 0.897f_3. \tag{8}$$

$$\text{Propextent} = -0.073f_1 + 0.824f_2 + 0.070f_3. \tag{9}$$

According to (9) and Table 3, Nkill (0.894) and Nwound (0.846) have the greatest load values at f_1, indicating that f_1 is mainly related to fatal and non-fatal injuries in the attack and can be referred to as the factor of fatalities; f_2 shows a strong relevance with Nperps and Propextent, whose load values are 0.796 and 0.824, respectively. Thus, it can be referred to as the factor of property damage and scale of the attack. In addition, Multiple (0.473) and Success (0.897) have the greatest load values at f_3, reflecting that f_3 carries the social influence information of a terrorist attack. For example, if an attack is one of a series of successful incidents, it will cause social panic and other hazards.

In addition, we can see that the three common factors are mutually independent in Table 4 when orthogonal rotation is adopted. That is, we extracted three basic structures from the seven observable variables representing terrorist attacks (the factor of fatalities f_1, factor of property damage and attack scale f_2, and factor of social influences f_3). In this framework, we can explicitly address and discuss the issue of the classification and gradation of terrorist attacks.

We estimate the factor scores for each observation (f_1, f_2, and f_3) using the retrogressive method, and we can obtain the following factor score coefficient solution:

$$\begin{aligned} f1 = &-0.048\,\text{Nperps} + 0.393\,\text{Nkill} + 0.382\,\text{Nwound} \\ &+ 0.375\,\text{Nhostkid} + 0.038\,\text{Multiple} \\ &- 0.107\,\text{Success} - 0.012\,\text{Propextent}, \\ f2 = &+0.595\,\text{Nperps} + 0.079\,\text{Nkill} - 0.048\,\text{Nwound} \\ &- 0.089\,\text{Nhostkid} + 0.037\,\text{Nultiple} \\ &- 0.070\,\text{Success} + 0.609\,\text{Propextent}, \\ f3 = &-0.027\,\text{Nperps} - 0.006\,\text{Nkill} - 0.018\,\text{Nwound} \\ &+ 0.010\,\text{Nhostkid} + 0.441\,\text{Multiple} \\ &+ 0.882\,\text{Success} + 0.024\,\text{Propextent}. \end{aligned} \tag{10}$$

TABLE 4 Matrix of correlation coefficients among the three common factors

Common factor	f_1	f_2	f_3
f_1	1.000	0.000	0.000
f_2	0.000	1.000	0.000
f_3	0.000	0.000	1.000

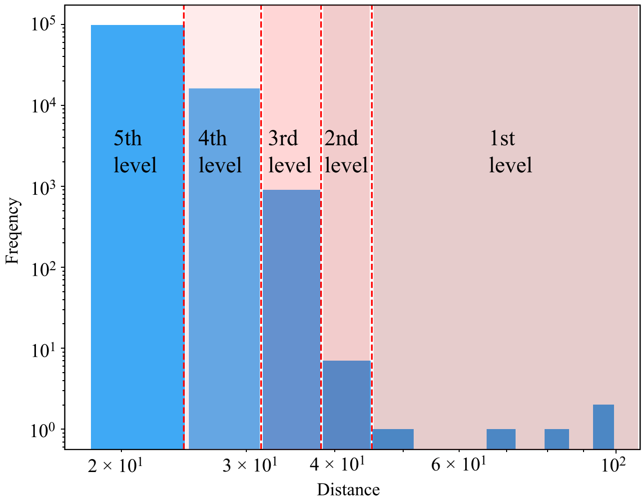


FIGURE 2 Ranking and frequency distribution of the events according to the calculated Euler distance

To make the weights of the factors equivalent and remove dimensions, let

$$f_1 = \frac{f_1 - \min f_1}{\max f_1 - \min f_1}, \tag{11}$$

$$f_2 = \frac{f_2 - \min f_2}{\max f_2 - \min f_2}, \tag{12}$$

$$f_3 = \frac{f_3 - \min f_3}{\max f_3 - \min f_3}. \tag{13}$$

In this manner, the factors are normalized to unity in the interval [0.1]. Remarkably, they describe fatalities, property damage, and attack scale and social influences from three independent aspects, which are all positively related to the severity of a terrorist attack. Here, we construct a three-dimensional factor space with each axis representing one factor, and a terrorist attack can be regarded as a point in the space. The point at the origin indicates no fatalities, damage, or social influence and no hazard, whereas points further from the origin indicate more severity. To quantify and grade the overall impacts of terrorist attacks, we calculate the Euclidean distance **R** from the origin in the factor space as follows:

$$\mathbf{R} = \sqrt{f_1^2 + f_2^2 + f_3^2}. \tag{14}$$

We divide terrorist attacks into five levels according to the overall impacts, as shown in Figure 2. The horizontal axis is the distance from the origin, and the vertical axis shows the number of attacks. This gradation method, which integrates a large amount of data cleansing and factor analysis in the early stage, picks the primary factors automatically and create a Euclidean distance to measure the overall impacts objectively, where no subjective claims or personal preferences are involved in the attack grading. In addition, each factor contains comprehensive and rich information using many variables, and thus, it can be used as a criterion for grading terrorist attacks.

4 | EVALUATION AND VISUALIZATION

We ranked the top 10 worst terrorist attacks in the past two decades to evaluate our classification method in Table 5.

TABLE 5 Ten worst terrorist attacks over two decades

Rank	Event ID	Score	Nperps	Nkill	Nwound	Nhostkid	Propextent
1[a]	200109110004	99.451	5.0	1384.0	8190.0	88.0	1
2[b]	201408090071	83.772	NaN	953.0	NaN	5350.0	NaN
3	201406150063	65.672	NaN	1570.0	NaN	1686.0	NaN
4[c]	199808070002	49.166	NaN	224.0	4000.0	NaN	2
5[d]	201406100042	41.308	NaN	670.0	0.0	NaN	3
6[e]	199811010001	41.074	700	80.0	NaN	45.0	2
7[f]	201710140002	40.831	1.0	588.0	316.0	NaN	NaN
8[g]	200708160008	39.282	2.0	250.0	750.0	NaN	2
9	201608050023	37.377	NaN	97.0	NaN	3000.0	NaN
10[h]	200403110003	36.297	NaN	73.0	450.0	NaN	2

Note: The marked attacks were part of Multiple incidents: [a]200109110004 200109110006 200109110007; [b]201408030057 201408030059 201408090071; [c]199808070002 199808070003; [d]201406100041 201406100042 201406100043 201406100044 201406100045 201406100049; [e]199811010001 199811020001; [f]201710140002 201710140003; [g]200708150005 200708160008; [h]200403110001 200403110003 200403110004 200403110005 200403110006 200403110007.

The September 11 attacks in 2001, which consist of multiple incidents caused by four coordinated terrorist attacks, score 99.451 and are ranked first beyond doubt. The Sinjar massacre by ISIS in early August 2014 (score: 83.772) is ranked second. Remarkably, the extent of property damage is unknown in the Sinjar massacre, and it cannot be ranked by a traditional ranking system judged by the rule of property damage alone, which leads to inaccurate and biased ranking. On the one hand, our grading method overcomes the obstacles that occur when many variables in an actual incident are missing or *NaN*; on the other hand, we can provide a comprehensive and objective evaluation of an incident from Multiple aspects.

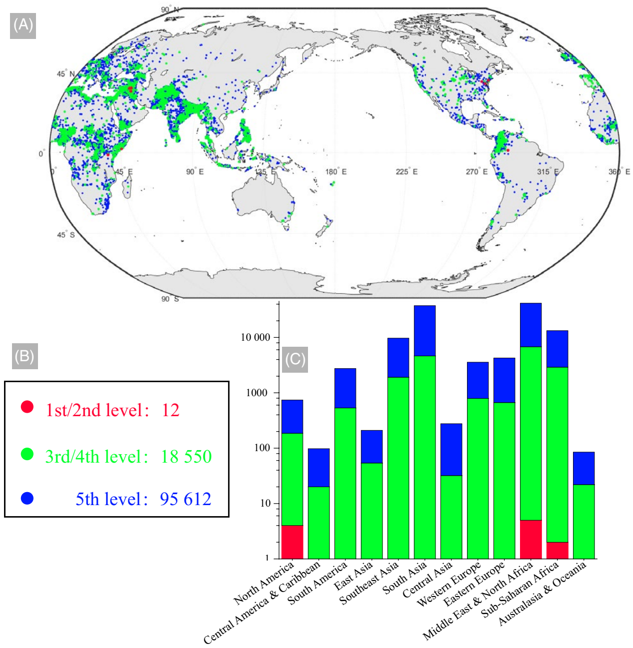


FIGURE 3 Overall impact ranking distribution of terrorist attacks in different regions

Subsequently, we used this model to grade all 114 183 attacks across the world recorded between 1998 and 2017 in the GTD, as shown in Figure 3A. To improve the visualization and avoid visual confusion, we chose three high-contrast colors, red, green, and blue, to mark the first and second levels (12 events), the third and fourth levels (18 550 events), and the fifth level (95 612 events) in Figure 3B, respectively. In addition, we counted the number of attacks that took place in each of 12 regions logarithmically, as shown in Figure 3C. By mapping complex and abstract data in the GTD onto a world map, we can learn that North America, the Middle East, and Africa are relatively prone to more severe attacks according to the grading results. We can also find a strong contrast in that only about 100 attacks took place in Central America, East Asia, Central Asia, and Oceania/Australia, whereas thousands of attacks occurred in South America,

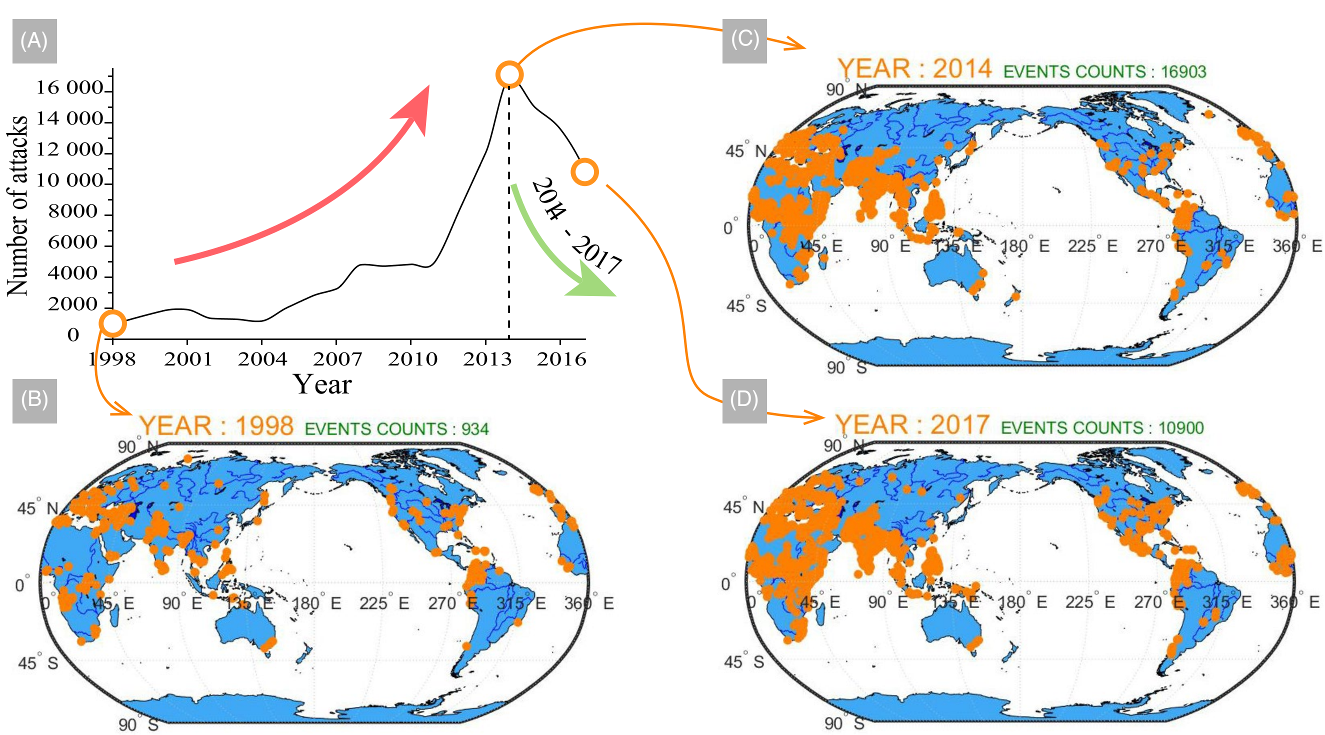


FIGURE 4 (A)Visualization of the trend of terrorist attacks since 1998 and the spatial distributions of attack events in (B) 1998, (C) 2014, and (D) 2017

Southeast Asia, South Asia, Europe, the Middle East, and Africa.

The trend of terrorist attacks since 1998 is visualized in Figure 4A, and Figure 4B shows the spatial distribution of all 934 attacks in 1998. It can be seen that few terrorist attacks occurred and most of them were local clusters, but Figure 4C shows that 16 903 attacks were carried out in the year 2014 and reached a peak. It is clear that the Middle East, South Asia, Southeast Asia, Europe, and Africa have been covered by a wide range of attacks, where one succeeded every 30 minutes on average. Terrorist attacks have gradually decreased to 10 900 in 2017, as shown in Figure 4D. Although the overall number of attacks has been controlled, it can be seen from the figure that the spreading trend still exists.

5 | CONCLUSION

In this paper, we collected 114 183 records of terrorist attacks. After a series of preprocessing steps, we chose seven variables to describe the overall impact and reduced the number of dimensions into three independent compact basic structures. Mapping events in this space, we computed the Euler distance to grade attacks in an automatic and objective manner. No subjective claims or personal preferences are involved in the gradation, and each common factor contains comprehensive and rich information on many variables. We also use this machine learning model to grade all such attacks across the world and visualize them to provide new insights.

Our work offers an elegant and transferable way of dealing with the analysis of massive data and ranking automatically. It is suitable for a broad range of gradation problems ranging from traffic accidents, meteorological and earthquake disasters to social behavior, and even urban planning. The limitation of this method is that its scalability is not as good as that of the hybrid artificial intelligence algorithm.

ACKNOWLEDGEMENTS

This work was supported by the Scientific Research Project of Zhejiang Natural Science Foundation, P.R. China (No. LY20G030018) and the Scientific Research Project of Wenzhou Science and Technology Bureau, P.R. China (No. 2013R0017).

CONFLICT OF INTERESTS

The authors declare no competing interests.

ORCID

Wanle Chi https://orcid.org/0000-0002-3911-3200

REFERENCES

1. W. H. Bakun et al., *Implications for prediction and hazard assessment from the 2004 parkfield earthquake*, Nature **437** (2005), 969–974.
2. J. A. Tenreiro Machado, C. M. A. Pinto, and A. Mendes Lopes, *Power law and entropy analysis of catastrophic phenomena*, Math. Prob. Eng. **2013** (2013), 1–10.
3. H. Mata-Lima et al., *Impactos dos desastres naturais nos sistemas ambiental e socioeconômico: O que faz a diferença?*, Amb. Soc. **16** (2013), 45–64.
4. M. Robin, *Hogarth, Cognitive processes and the assessment of subjective probability distributions*, J. Am. Stat. Assoc. **70** (1975), 271–289.
5. W. Enders, T. Sandler, and K. Hartley. *Chapter 26 terrorism: An empirical analysis*, in Handbook of Defense Economics, vol. **2**, Elsevier, North Holland, Netherlands, 2007, pp. 815–866.
6. R. Caruso and A. Locatelli, Understanding Terrorism: A Socio-Economic Perspective, vol. **22**, Emerald Publishing Group, Bingley, UK, 2014.
7. X. Wang et al., *Investigative visual analysis of global terrorism*, Comp. Graphics Forum **27** (2008), 919–926.
8. G. Lafree, *Using open source data to counter common myths about terrorism*, in Criminologists on Terrorism and Homeland Security, Cambridge University Press, Cambridge, UK, 2011, pp. 411–442.
9. K. Kis-Katos, H. Liebert, and G. G. Schulze, *On the origin of domestic and international terrorism*, Eur. J. Polit. Econ. **27** (2011), S17–S36.
10. START, *Codebook: Inclusion criteria and variables*, J. Educ. Stat. (2018), 453.
11. G. LaFree and L. Dugan, *Introducing the global terrorism database*, Terrorism Polit. Violence **19** (2007), 181–204.
12. Z. Li et al. *Terrorist group behavior prediction by wavelet transform-based pattern recognition*, Dis. Dynam. Nat. Soc. **2018** (2018), 1–16.
13. G. LaFree and J. D. Freilich, *Editor's introduction: Quantitative approaches to the study of terrorism*, J. Quant. Criminol. **28** (2012), 1–5.
14. E. L. Huamaní, A. Mantari, and A. Roman-Gonzalez, *Machine learning techniques to visualize and predict terrorist attacks worldwide using the global terrorism database*, Int. J. Advan. Comp. Sci. Appl. **11** (2020), 563–570.
15. Q. I. U. Lingfeng et al., *Study on prediction of global terrorist attacks based on machine learning and fragile states index*, J. Catastrophol. **34** (2019), 211–214.
16. C. Verma, S. Malhotra, and V. Verma, *Predictive modeling of terrorist attacks using machine learning*, Int. J. Pure Appl. Math. **119** (2018), no. 15, 49–60.
17. K. F. Widaman and R. P. McDonald, *Factor analysis and related methods*, J. Educ. Stat. **12** (1987), 308–313.
18. B. Price, *A first course in factor analysis*, Technometrics **35** (1993), 453.
19. J. G. Liu and P. J. Mason, *Principal component analysis*, in Essential Image Processing and GIS for Remote Sensing, Wiley, London, UK, 2009, pp. 77–90.
20. L. A. Pasa et al., *A contribution to the study of ensemble of self-organizing maps*, Math. Prob. Eng. **2015** (2015), 1–10.
21. H. F. Kaiser, *The varimax criterion for analytic rotation in factor analysis*, Psychometrika **23** (1958), 187–200.

AUTHOR BIOGRAPHIES

Wanle Chi received his BS degree in computer science from Dalian University of Technology, Dalian, Liaoning, P.R. China, in 2000, and his MS degree in communication engineering from Zhejiang University, Hangzhou, Zhejiang, P.R. China, in 2007. Currently, he is studying for his PhD degree at the Faculty of Information and Communication Technology, Technical University of Malaysia Malacca, Malacca, Malaysia and works at the Department of Information Technology, Wenzhou Polytechnic, Wenzhou, Zhejiang, P.R. China, where he is now a lecturer. His research interests include pattern recognition, machine learning, and big data analysis.

Yihong Du received her BS degree in electronics engineering from Hangzhou Business University, Hangzhou, Zhejiang, P.R. China, in 1995, and her MS degree in communication engineering from Zhejiang University, Hangzhou, Zhejiang, P.R. China, in 2007. Currently, she works in the Department of Information Technology, Wenzhou Polytechnic, Wenzhou, Zhejiang, P.R. China, where she is now a professor. Her research interests include databases and big data analysis.