



“华为杯”第十五届中国研究生 数学建模竞赛

学 校

参赛队号

队员姓名

“华为杯”第十五届中国研究生 数学建模竞赛

题 目 对恐怖袭击事件记录数据的量化分析

摘 要：

本文主要研究恐怖袭击事件的量化分析问题，通过构造基于等级全息建模的恐怖袭击危害指数评估与分级模型进行恐怖袭击事件危害分级、建立有噪的基于密度聚类（DBSCAN）的恐怖袭击组织识别模型捕获事件相似性以反向追踪恐怖组织、构建风险评估模型于国家层面研判未来反恐态势并搭建基于帕累托图的恐怖袭击聚集性评价模型。具体步骤如下：

针对问题一，引入等级全息建模思想，构建恐怖袭击危害指标体系，建立恐怖袭击危害指数评估与分级模型。以危害性相关特征作为低级指标，通过等级全息建模技术进行恐怖袭击危害指数量化建模。而后，建立基于经验阈值的危害分级模型与基于加权聚类的危害分级模型，根据两个模型的分级结果，进行指标体系权重的动态调整，克服初始权重的主观影响，完成危害指数计算并成功将危害指数分为五个等级。

针对问题二，建立基于 DBSCAN 的恐怖袭击组织识别模型，对犯罪组织及犯罪对象相关指标进行研究，计算密度进行组织的自动聚类，并利用欧式距离度量事件的嫌疑人的嫌疑程度，取得了精确的分类效果。

针对问题三，通过主成分分析提取主要因子——基本风险因子与脆弱性因子，研究近三年来恐怖袭击事件发生的主要原因；构建风险评估模型，剖析不同国家的时间转移特性和空间转移特性，进而分析全球恐怖袭击事件的蔓延特性及级别分布规律，观察到中东南亚等地区是恐怖组织的重点袭击区域。最后通过分析近三年的恐怖袭击特征，对反恐态势提出建议与意见。

针对问题四，引入基尼系数和 Lorenz 曲线，建立基于帕累托图的恐怖袭击聚集性评价模型，对国家、区域、武器类型、攻击类型、攻击目标等因素的空间聚集性展开研究，得到上述考虑的因素大多具有较强聚集性的结论。此外，从时间角度分析死亡人数和受袭击国家数量的基尼系数的变化，进行时间尺度评估，发现近三年随时间变化趋势并不明显。

关键词: 恐怖袭击, 等级全息建模(HHM), 有噪的基于密度聚类算法(DBSCAN), 风险评估, 基尼系数

目录

一、问题重述.....	5
1.1 研究背景.....	5
1.2 研究问题.....	5
二、模型假设.....	6
三、符号说明.....	6
四、问题分析.....	6
4.1 恐怖袭击危害分级问题分析.....	6
4.2 恐怖袭击事件制造者追踪问题分析.....	6
4.3 未来反恐态势问题分析.....	7
4.4 数据集多方面利用问题分析.....	7
五、模型建立与求解.....	7
5.1 模型准备.....	7
5.2 问题一.....	8
5.2.1 恐怖袭击危害等级全息建模框架.....	8
5.2.2 恐怖袭击危害指数评估与分级模型.....	9
5.2.3 恐怖袭击危害量化分级模型求解.....	12
5.3 问题二.....	12
5.3.1 指标选择.....	13
5.3.2 基于 DBSCAN 的恐怖袭击组织识别模型.....	13
5.3.3 恐怖袭击组织识别模型求解.....	15
5.4 问题三.....	16
5.4.1 主成分分析与风险评估模型.....	17
5.4.2 恐怖袭击风险分析.....	18
5.4.3 恐怖袭击主成分分析结果.....	19
5.4.4 恐怖袭击风险时空演变.....	23

5.4.5 下一年反恐态势.....	24
5.5 问题四.....	24
5.5.1 基尼系数与 <i>Lorenz</i> 曲线.....	24
5.5.2 恐怖袭击聚集性分布.....	24
5.5.3 恐怖袭击时间性分析.....	30
5.5.4 小结.....	31
六、模型的评价与推广	31
6.1 模型优点.....	31
6.2 模型缺点.....	31
6.3 模型的改进与推广	31
七、参考文献.....	32

一、问题重述

1.1 研究背景

恐怖袭击是指极端分子人为制造的针对但不仅限于平民及民用设施的不符合国际道义的攻击方式。自 20 世纪九十年代以来，恐怖袭击在全球范围内迅速蔓延，其严峻趋势逐渐引起了世界人民的关注。恐怖袭击事件具有极大的杀伤性与破坏力，不仅明显地直接影响人类生活、人们带来巨大的心理压力，更会持久地影响国际政治、公民自由和经济发展。

为更好的维护世界和平、保障人民生命财产安全，打击恐怖主义行为迫在眉睫，而对恐怖组织或恐怖分子展开深入研究有助于人们建立合理的恐怖袭击预警机制与情报管理机制。如何将在众多恐怖袭击事件相关数据中挖掘出潜藏的恐怖分子信息、恐怖袭击情报信息对反恐防恐有着重要意义。

1.2 研究问题

现要求通过数学建模来完成以下任务：

1) 依据危害性对恐怖袭击事件分级

传统的灾难性事件分级一般采取主观方法，主要按照人员伤亡和经济损失程度划分，然而恐怖袭击的危害性不仅体现在这两个方面，还与发生的时机、地域、针对的对象等等诸多因素有关。如何根据附件 1 及其他相关信息，建立基于数据分析的量化分级模型，将附件 1 给出的事件按危害程度从高到低分为一至五级是本环节研究的主要问题。

此外，建立恐怖袭击危害分级模型后，还需列出近二十年来危害程度最高的十大恐怖袭击事件，并给出要求事件的分级结果。

2) 依据事件特征反向追踪恐怖袭击事件制造者

目前许多在案的恐怖袭击事件尚未确定作案者，为尽早发现新生或隐藏的恐怖分子，要求建立数学模型将 2015、2016 年度发生的、尚未有组织或个人宣称负责的恐怖袭击事件中可能是同一个恐怖组织或个人在不同时间、不同地点多次作案的若干案件归为一类。

建立分类模型后，选出危害性排列前五名的组织或个人作为给定事件的嫌疑人，并按照嫌疑程度对 5 个嫌疑人进行排序。

3) 对未来反恐态势的分析

为提高反恐斗争的针对性和效率，需要根据附件 1 与因特网上的有关信息，建立适当的数学模型，研究近三年来恐怖袭击事件发生的主要原因、时空特性、蔓延特性、级别分布等规律，进而分析研判下一年全球或某些重点地区的反恐态势，并利用图/表表示研究结果，阐述对反恐斗争的见解和建议。

4) 数据的进一步利用

除了恐怖袭击严重程度分析、追踪恐怖袭击作案者以及对未来反恐态势的推测，附件 1 的数据还可以发挥哪些作用，利用模型或方法进行说明。

二、模型假设

1. 假设附件数据真实可靠；
2. 不对由于恐怖袭击而导致的人们心理伤害进行量化作为指标；

三、符号说明

符号	意义
$X = \{x_i i = 1, 2, \dots, n\}$	恐怖袭击事件样本集
D	事件危害指数
ϵ	邻域距离阈值
$Minpts$	距离为 ϵ 的邻域中样本个数的阈值
S	恐怖袭击组织的嫌疑指数
$L(i)$	第 i 个样本序列号

注：这里仅标明了文中常用的符号及其含义，文中另外用到的符号会在相应地方进行说明。

四、问题分析

4.1 恐怖袭击危害分级问题分析

目前灾难性事件多由主观方法分级，划分标准主要按照人员伤亡和经济损失程度，考虑到恐怖袭击事件的组织复杂性、行为复杂性与后果复杂性，本文首先挑选具有代表性危害性相关特征作为低级指标，引入等级全息建模（HHM）技术进行恐怖袭击危害指数建模，通过多方位、多视角的恐怖袭击事件观察，从而进行危害指数量化。而后，分别建立基于经验阈值的危害分级模型与基于加权聚类的危害分级模型，根据两个模型的分级结果，进行恐怖袭击危害指标体系权重的动态调整，以克服初始权重过于主观，提高分级准确度。

4.2 恐怖袭击事件制造者追踪问题分析

考虑到同一恐怖组织或个人在不同时间、不同地点多次作案的若干案件可能具有潜在相似性，或同一时间组的恐怖袭击事件具有一定程度的共同性，本文拟建立基于 DBSCAN(具有噪声的基于密度聚类)的恐怖袭击组织识别模型，该模型主要集中于犯罪组织及犯罪对象相关指标，计算密度进行组织的自动聚类，无需人工干预设置组织个数。针对某事件的嫌疑人（组织）的嫌疑程度可由事件与该

嫌疑人（组织）间的欧式距离进行度量。

4.3 未来反恐态势问题分析

恐怖主义具有潜在性与预谋性，大量的历史数据可以帮助我们分析与预测恐怖袭击的趋势。本文拟从国家层面对恐怖袭击的风险建立模型，通过主成分分析研究近三年来恐怖袭击事件发生的主要原因；通过建立风险评估模型研究不同国家的时间转移特性和空间转移特性，进而分析全球恐怖袭击事件的蔓延特性及级别分布规律。最后通过分析近三年的恐怖袭击特征，对反恐态势提出几点建议与意见。

4.4 数据集多方面利用问题分析

针对该数据集，本文拟建立基于帕累托图的恐怖袭击聚集性评价模型，引入基尼系数和 Lorenz 曲线，对国家、区域、武器类型、攻击类型、攻击目标等因素的空间聚集性展开研究。此外，从时间角度分析死亡人数和受袭击国家数量的基尼系数的变化，进行时间尺度评估。

五、模型建立与求解

5.1 模型准备

GTD 全球恐怖主义数据库中记录了各年度世界上发生的恐怖袭击事件记录，包括事件标志、事件信息、事件地点、攻击信息、武器信息与目标/受害者信息等，每种指标分别以数字或文本的形式进行记录，为完成后续的模型建立求解，我们需要对有效指标进行量化赋值。

表 1 指标赋值方法表（示例）

类型	指标名称	元素/意义/依据	赋值方法
有序指标	袭击位置	人口数量、市县级别、经济水平等	依法打分
	袭击能力	武力水平、集团归属等	
	犯罪后果	伤亡人数、财产损失等	
无序指标	攻击类型	暗杀、武装袭击、轰炸、劫持等	根据强烈程度进行排序赋值
	武器类型	化学武器、核武器、炸弹等	
	受害人类型	政府、宗教、教育、海事等	
连续指标	反恐处置能力	该目标的恐怖袭击发生概率与成功概率	计算
	目标威胁	袭击发生的概率	
	目标脆弱	袭击成功率	

观测数据，反应恐怖袭击事件要素的指标（包括附件特征指标与自建指标）

可分为三大类：有序指标、无序指标与连续性指标。其中有序指标指本身可以按序标称；无序指标多为并列式，如恐怖组织结构可分为等级式、网络式、无中心等；连续指标则可通过计算得出。对于不同类型指标，本文将采用不同种赋值方法，如表 1。（考虑到指标数量过多，表中仅列出代表性指标）

5.2 问题一

恐怖袭击通常施行暴力、破坏或流血行动以实现某种政治、经济、宗教或社会目标，其作为一种复杂的政治、社会现象影响的是整个社会秩序与稳定，故而恐怖袭击事件的危害性与人员伤亡、经济损失、发生的时机、地域、针对的对象等等诸多因素相关。问题一中，拟根据附件 1 及其他有关信息，建立基于数据分析的量化分级模型实现恐怖袭击危害分级。

5.2.1 恐怖袭击危害等级全息建模框架

考虑到恐怖袭击事件的组织复杂性、行为复杂性与后果复杂性，本文拟采用等级全息建模（HHM）技术^[1]进行恐怖袭击危害等级建模。HHM 以展现复杂系统于各方面、视角、观点和层级中的不同特征和本质为目的，适用于大规模的、复杂的系统建模。通过多方位、多视角的恐怖袭击事件观察，从而进行危害量化与评级研究。

运用 HHM 进行恐怖袭击危害建模，首先需要通过分析恐怖分子或组织实施的恐怖行为以及可观测的活动进行情景重构。Yacov Y. Haimes 提出恐怖袭击情景具有六个可观测阶段：意图、目标获取、计划、准备、执行、结束期^[1]。结合生命周期理论，本文将六个阶段整合为袭击策划期、具体实施期、行为效应期 3 个风险情景，据此构建恐怖袭击事件危害的 HHM 框架，如图 1 所示。

（1）恐怖袭击的袭击策划期

恐怖袭击策划期主要进行恐怖活动的观察、策划以及准备，此时期可分析资源配置与形势环境两个风险场景，其中资源配置主要考虑恐怖袭击规模、人员数模、是否为独立行动或事件组行动；袭击地区的地理位置在一定程度上影响了事件的危险程度。恐怖袭击的具体实施期

（2）恐怖袭击的具体实施

恐怖袭击的具体实施期主要面向政治、经济、社会、机构与宗教，观测数据可发现，对应不同的袭击对象与目的将会采用不同的攻击类型与武器类型，其危险程度也并不相同。

（3）恐怖袭击的行为效应期

此时期将显现恐怖袭击的效果与影响，将还原人员伤亡、经济损失、国家损失与事件影响等四个危险场景，全息展示恐怖袭击的行为效应。

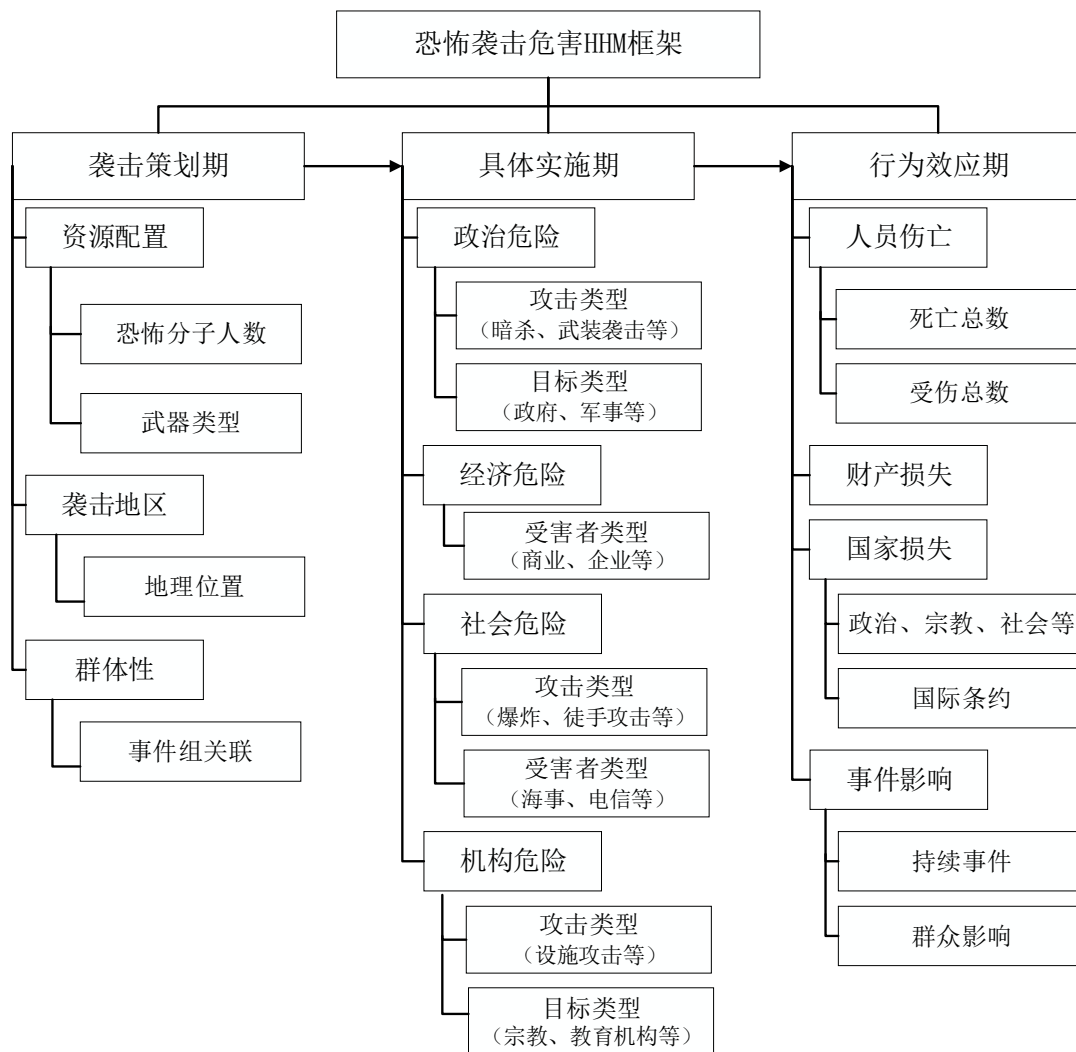


图 1 恐怖袭击危害 HHM 框架

5.2.2 恐怖袭击危害指数评估与分级模型

为进行各个情景分析，在恐怖袭击危害等级全息建模框架构建完成后，还需对各指标因素进行量化。首先本文将 HHM 框架转换为对应的指标体系，如表 2 所示（表 2 中标明一二级指标与对应的三级指标数据来源）。

表 1 恐怖袭击危害指标体系

一级指标	二级指标	三级指标/数据来源
w_1 袭击策划期	$w_{1,1}$ 资源配置	$w_{1,1,1}$ 恐怖分子人数
		$w_{1,1,2}$ 武器类型
	$w_{1,2}$ 袭击地区	$w_{1,2,1}$ 地理位置
	$w_{1,3}$ 群体性	$w_{1,3,1}$ 事件组关联

w_2 具体实施期	$w_{2,1}$ 政治危险	$w_{2,1,1}$ 攻击类型
		$w_{2,1,2}$ 目标类型
	$w_{2,2}$ 经济危险	$w_{2,2,1}$ 受害者类型
	$w_{2,3}$ 社会危险	$w_{2,3,1}$ 攻击类型
		$w_{2,3,2}$ 受害者类型
	$w_{2,4}$ 机构危险	$w_{2,4,1}$ 攻击类型
		$w_{2,4,2}$ 目标类型
w_3 行为效应期	$w_{3,1}$ 人员伤亡	$w_{3,1,1}$ 死亡总数
		$w_{3,1,2}$ 受伤总数
	$w_{3,2}$ 财产损失	$w_{3,2,1}$ 财产损失
	$w_{3,3}$ 国家损失	$w_{3,3,1}$ 目标信息
		$w_{3,4,1}$ 持续事件
	$w_{3,4}$ 事件影响	$w_{3,4,2}$ 群众影响

令 i, j, z 分别代表一级、二级、三级指标， $A_{i,j,z}$ 为对应指标值， $w_{i,j,z}$ 为各指标对应权重，则对应样本的危害指数为

$$D = \sum_{i=1}^3 w_i A_i \quad (1)$$

初始权重主要受主观因素影响，为得到比较客观的分级结果，在评级过程中需对各指标权重进行动态调整，具体调整思路如下：

(1) 基于经验阈值的危害分级模型

恐怖袭击事件的危害程度往往是相对的，通常使用基于阈值的危害分级方法，为完成危害程度由高到低的五级划分设置 4 个划分阈值 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ ，并根据危险指数进行等级的确定。

将样本根据危险指数从高到低进行重新排列，则有式 (2)，其中危害等级按照 1,2,3,4,5 顺序逐渐降低。

$$y = \begin{cases} 1, & l_x \geq \alpha_1 * N \\ 2, & \alpha_1 * N > l_x \geq \alpha_2 * N \\ 3, & \alpha_2 * N > l_x \geq \alpha_3 * N \\ 4, & \alpha_3 * N > l_x \geq \alpha_4 * N \\ 5, & other \end{cases} \quad (2)$$

(2) 基于加权聚类的危害分级模型

相同危害程度的恐怖袭击事件通常具有一定程度的相似性，如伤亡人数巨大，政治影响深远、经济损失巨大等，在不使用标签的情况下，使用无监督学习对加权指标数据进行聚类分析，并由此进行危害分级。

设样本数据集 $X = \{x_i | i = 1, 2, \dots, n\}$ ， $C_j = (j = 1, 2, \dots, k)$ 为初始聚类中心，本模型中将所有样本聚为五类，基本思想为如下：

- 从输入的训练样本集 $\{x_i\}$ 中随机选择 k 个数据点 (c_1, c_2, \dots, c_k) 作为初始聚类中心点；
- 计算数据集 $\{x_i\}$ 中的所有数据点到所有初始聚类中心点的距离，并将该点分至与其距离最近的中心点 c_i 所在的聚类中；
- 更新各聚类的中心点 $c'_i = (i = 1, 2, \dots, k)$ ， $c'_i = \frac{1}{n} \sum_{x \in S_i} x$ ， n 为 S_i 领域内包含的样本个数；
- 若聚类中心的变化小于阈值或达到预先设置的迭代次数，说明算法收敛，结束。否则，重新回到 (2) 开始迭代。

(3) 基于混合模型的权重动态调整

通过基于经验阈值的危害分级模型与基于加权聚类的危害分级模型，针对同一样本可获得两种分类结果，如图 2 所示。

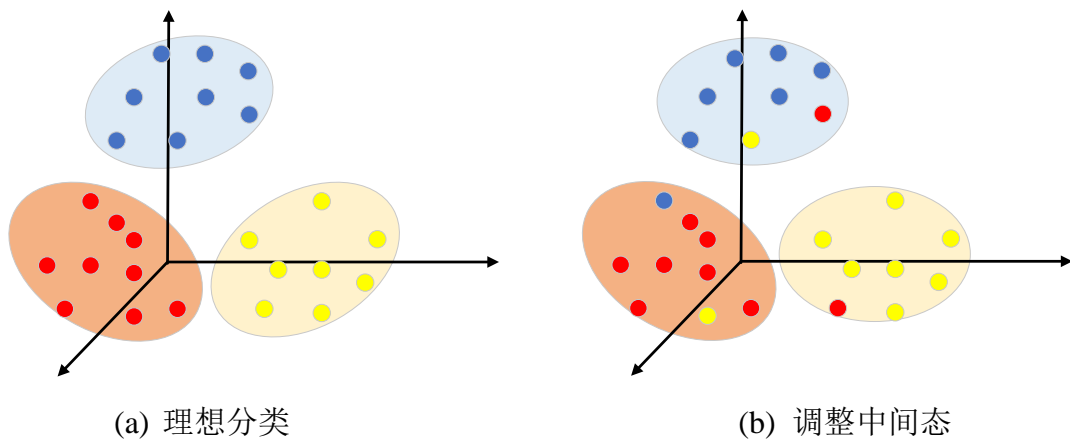


图 2 分类效果图

图 2 中根据阴影部分观测等级（蓝色>红色>黄色），根据数据点颜色观测样

本危害程度（蓝色>红色>黄色），图 2(a)为理想的分类效果，即危害等级高的样本危害指数高于危害等级低的样本，而图 2(b)为存有分歧的分类效果，是求解模型的常见状态，此时需进行权重动态调整，直至错分样本达到设定阈值，达到期望状态。

5.2.3 恐怖袭击危害量化分级模型求解

到本问题主要针对危害级别进行分级建模，故从中挑选 12 个基础指标构建恐怖袭击危害性指标体系，如表 2 所示，其中每个指标依据赋值规则进行赋值。

对于初始权重，首先采用均分思想，即同一级指标下的次级指标权重相同，其次考虑各指标实际意义，针对具体实施期，恐怖袭击造成的危险是联动的，即政治危险中也同样具有经济危险、文化危险、社会危险等，故而不作区分，即初始权重设为相同，针对行为效应期的 4 类二级指标，本文拟根据《中华人民共和国刑法》对不同行为后果的量刑程度进行初始权重的计算，由此计算危险指数。

而后，根据危险指数，通过经验阈值与加权聚类进行危害程度分类，并根据分类结果反向动态调整权重，直至达到期望目标。

最终根据恐怖袭击危害量化分级模型中危险指数计算得近二十年来危害程度最高的十大恐怖袭击事件如表 3 典型事件危害级别如表 4 示。

表 3 二十年来危害程度最高的十大恐怖袭击事件

排名	事件编号	排名	事件编号
1	200109110004	6	201710140002
2	201406150063	7	201408200027
3	199808070002	8	200403210001
4	201408090071	9	200409010002
5	201406100042	10	201607020002

表 4 典型事件危害级别

事件编号	危害级别	事件编号	危害级别
200108110012	1	201411070002	5
200511180002	2	201412160041	2
200901170021	1	201508010015	5
201402110015	3	201705080012	2
201405010071	4		

5.3 问题二

目前发生的恐怖袭击事件，并非所有的活动都有人表示对其负责或查到作案组织（个人），本问拟将 2015、2016 年度发生的、尚未有组织或个人宣称负责的恐怖袭击事件通过恐怖袭击组织识别模型，进行事件归类，即将可能是同一个恐怖组织或个人在不同时间、不同地点多次作案的若干案件归为一类。

5.3.1 指标选择

考虑到本问题主要针对犯罪组织及犯罪对象,与上节主要针对内容为事件危害不同,本模型将重新选择基础指标建立恐怖袭击组织识别模型,指标包括国家,地区,维度,经度,攻击类型,目标/受害者类型,受害者国籍,凶手数量,声称负责相关信息,武器类型,持续性等特征。

5.3.2 基于 DBSCAN 的恐怖袭击组织识别模型

DBSCAN(Density-Based Spatial Clustering of Applications with Noise, 具有噪声的基于密度聚类)是一个具有代表性的基于密度的聚类算法^[3,4], 它将簇定义为密度相连的点的最大集合, 能够把具有足够高密度的区域划分为簇, 可在噪声的空间数据库中发现任意形状的聚类, 且不需要预先设定聚类个数。适用于未知分类的情况, 故此本文拟建立基于 DBSCAN 的恐怖袭击组织识别模型实现恐怖袭击事件的组织识别。

DBSCAN 基于一组邻域描述样本集的紧密程度, 参数 $(\epsilon, Minpts)$ 用来描述邻域的样本分布紧密程度。其中, ϵ 为某一样本的邻域距离阈值, $Minpts$ 为某一样本的距离为 ϵ 的邻域中样本个数的阈值。对于样本集 $X = \{x_i | i = 1, 2, \dots, n\}$, DBSCAN 具体的密度描述定义如下:

- 1) 给定任意一点 p , 其半径 ϵ 内的邻域即点 p 的 ϵ 邻域;
- 2) 若某点 p 的 ϵ 邻域至少包含最小数目 $Minpts$ 个点, 则称点 p 为核心点;
- 3) 若点 q 在点 p 的 ϵ 邻域内, 而 p 是核心点, 则称点 p 到点 q 关于 ϵ 和 $Minpts$ 直接密度可达;
- 4) 对于 x_i 和 x_j , 如果存在样本序列 p_1, p_2, \dots, p_T , 满足 $p_1 = x_i, p_T = x_j$, 且 p_{t+1} 由 p_t 密度直达, 则称 x_j 由 x_i 密度可达。
- 5) 对于 x_i 和 x_j , 如果存在核心对象样本 x_k , 使 x_i 和 x_j 均由 x_k 密度可达, 则称 x_i 和 x_j 密度相连。

模型聚类流程为: 从事件样本集 X 中某一个点(随机选择事件样本) p 出发, 若点 p 的 ϵ 邻域包含点多于 $Minpts$ 个, 则表明点 p 是核心对象, 创建以 p 为核心的簇, 将其 ϵ 邻域中的直接密度可达的点加入该簇中。将与簇中所有核心对象直接密度可达的点加入簇中, 迭代本过程直至所有与点 p 密度相连的点都加入簇中时, 选定尚未被加入任意簇的另一个点出发, 重复上述过程, 直至没有新的点可加入任意簇中时, 聚类算法结束, 未被加入任何簇的点即为噪声点, 如图 3 所示。

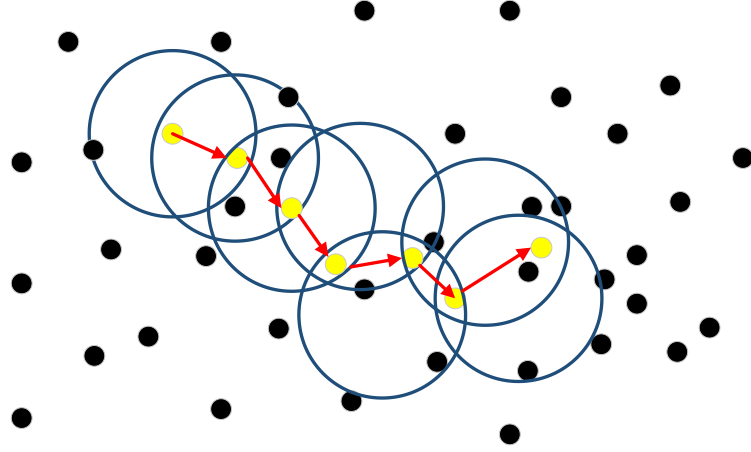


图 3 DBSCAN 流程说明图

其中样本与样本间距离的度量方式包括欧式距离、曼哈顿距离、且必须额度距离、民科夫斯基距离与马氏距离等，以下列出几种代表性距离计算公式：

欧氏距离：

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

曼哈顿距离：

$$d = \sum_{i=1}^n |x_i - y_i| \quad (4)$$

切比雪夫距离：

$$d = \max |x_i - y_i|, i = 1, 2, 3, \dots, n \quad (5)$$

算法流程如表 5 所示。

表 5 基于 DBSCAN 的恐怖袭击组织识别模型

输入：恐怖事件集 $X = \{x_i | i = 1, 2, \dots, n\}$ ，邻域参数 $(\epsilon, Minpts)$ ，样本距离度量方式

1) 初始化核心对象集合 $\Omega = \emptyset$ ，初始化聚类簇数 $k = 0$ ，初始化未访问样本集合 $\Gamma = X$ ，簇划分 $C = \emptyset$

2) 对于 $j = 1, 2, \dots, n$ ，按下面的步骤找出所有的核心对象：

a) 通过距离度量方式，找到样本 x_j 的 ϵ 邻域子样本集 $N_\epsilon(x_j)$

b) 如果子样本集样本个数满足 $|N_\epsilon(x_j)| \geq MinPts$ ，将样本 x_j 加入核心对象

样本集合： $\Omega = \Omega \cup \{x_j\}$

3) 如果核心对象集合 $\Omega = \emptyset$ ，则算法结束，否则转入步骤 4.

4) 在核心对象集合 Ω 中，随机选择一个核心对象 o ，初始化当前簇核心对象队

列 $\Omega_{cur} = \{o\}$ ，初始化类别序号 $k = k + 1$ ，初始化当前簇样本集合 $C_k = \{o\}$ ，

更新未访问样本集合 $\Gamma = \Gamma - \{o\}$

5) 如果当前簇核心对象队列 $\Omega_{cur} = \{\emptyset\}$ ，则当前聚类簇 C_k 生成完毕，更新簇划分

分 $C = \{C_1, C_2, \dots, C_k\}$ ，更新核心对象集合 $\Omega = \Omega - C_k$ ，转入步骤 3

6) 在当前簇核心对象队列 Ω_{cur} 中取出一个核心对象 o' ，通过邻域距离阈值 ϵ 找出

出所有的 ϵ 邻域子样本集 $N_\epsilon(o')$ ，令 $\Delta = N_\epsilon(o') \cap \Gamma$ ，更新当前簇集合

$C_k = C_k \cup \Delta$ ，更新未访问样本集合 $\Gamma = \Gamma - \Delta$ ，

更新 $\Omega_{cur} = \Omega_{cur} \cup (\Delta \cap \Omega) - o'$ ，转入步骤 5.

输出：恐怖袭击事件组织划分 $C = \{C_1, C_2, \dots, C_k\}$

5.3.3 恐怖袭击组织识别模型求解

以 2015-2016 年度发生的所有恐怖袭击事件作为输入，建立恐怖袭击组织识别模型，通过基于 DBSCAN 的恐怖袭击组织识别模型进行自主分类，将此两年的所发生恐怖袭击事件共分为 703 类，即所有事件由约 703 个组织分别策划。

1) 恐怖袭击组织危害程度标定

每个组织的危害性由疑似该组织策划的所以事件的危险指数均值标定，如下式

$$D_{C_i} = \bar{D}_{x_j}, x_j \in C_i \quad (6)$$

通过计算得恐怖袭击组织危害指数前十名如表 6.

表 6 恐怖袭击组织危害指数排名

排名	危害程度	排名	危害程度
嫌疑人 1	9975.2	嫌疑人 6	8059.8
嫌疑人 2	9488.6	嫌疑人 7	8046.8
嫌疑人 3	8505.0	嫌疑人 8	8043.7
嫌疑人 4	8239.2	嫌疑人 9	8037.5
嫌疑人 5	8127.7	嫌疑人 10	8035.3

2) 恐怖袭击组织嫌疑程度标定

对于每起恐怖袭击事件，某一恐怖袭击组织的嫌疑程度由该点至该组织簇内

各事件点的平均欧式距离标定，如下式

$$S_{C_i} = \bar{d}_{p,x_j}, \quad x_j \in C_i$$

$$d_{p,x} = \sqrt{\sum_{m=1}^n (p_m - x_m)^2} \tag{7}$$

其中 m 为特征维数。经此计算，得恐怖分子关于典型事件的嫌疑程度如表 7,事件与嫌疑人分布情况如图 4。

表 7 恐怖分子关于典型事件的嫌疑度

	1 号嫌疑人	2 号嫌疑人	3 号嫌疑人	4 号嫌疑人	5 号嫌疑人
201701090031	5	4	3	1	2
201702210037	5	4	3	1	2
201703120023	5	4	3	1	2
201705050009	5	4	3	1	2
201705050010	5	4	3	1	2
201707010028	5	4	3	1	2
201707020006	5	1	4	2	3
201708110018	5	4	3	1	2
201711010006	5	4	3	2	1
201712010003	5	4	3	1	2



图 4 典型事件与嫌疑人分布情况

5.4 问题三

对未来反恐态势的分析评估有助于提高反恐斗争的针对性和效率。由于恐怖

主义的潜在性与预谋性，很难预测恐怖组织的行动，但大量的历史数据可以告诉我们恐怖组织袭击的规律与风险。本章从国家层面分析恐怖袭击的风险，通过建立风险评估模型，研究近三年来恐怖袭击事件发生的主要原因、时空特性、蔓延特性、级别分布等规律，进而分析研判下一年全球或某些重点地区的反恐态势。

5.4.1 主成分分析与风险评估模型

主成分分析是一种多元统计方法，通过降维变换，提取出多维特征中的主要因子，去除特征中的冗余信息，更有效的对问题进行预测与综合评价。主成分分析的原理是相关矩阵的变化，本章通过把影响恐怖袭击的特征进行数学变换，通过主成分分析提取出主要因子，得到国家恐怖袭击风险综合得分。主成分分析不需要给每个特征指定权重，他可以通过观测样本自动为每个特征赋予权重，因此降低了分析结果中的主观因素。

a. 数学模型

假设有 p 个特征： x_1, x_2, \dots, x_p ，将其标准化为正态分布。每个变量可由 k ($k < p$)

个因子线性组合表示，分别为 f_1, f_2, \dots, f_k ，建立数学模型为：

$$\begin{cases} x_1 = a_{11}f_1 + a_{12}f_2 + \dots + a_{1k}f_k + \zeta_1 \\ x_2 = a_{21}f_1 + a_{22}f_2 + \dots + a_{2k}f_k + \zeta_2 \\ \dots \\ x_p = a_{p1}f_1 + a_{p2}f_2 + \dots + a_{pk}f_k + \zeta_p \end{cases} \quad (8)$$

用矩阵形式表示为：

$$X = AF + \zeta \quad (9)$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pk} \end{bmatrix} \quad (10)$$

其中， F 为公共因子； X 为标准化后的特征； A 为因子载荷矩阵， a_{ij} 是 X_i 和 f_i 的

协方差， a_{ij} 的值越大，则 X_i 和 f_i 的依赖程度就越大。

因子载荷可由最小二乘法、极大似然法、主轴因子法、主成分分析法等方法求得。本文文采用主成分分析法，计算过程如下：

- (1) 将原始数据标准化处理，使其服从正太分布；
- (2) 计算标准化后的数据的协方差矩阵；
- (3) 求协方差矩阵的非零特征根并由大到小排序 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ ，及对应的

正交特征向量 $e_i (i=1, 2, \dots, p)$ ；

- (4) 计算因子载荷矩阵 $A = (\sqrt{\lambda_1}e_1, \sqrt{\lambda_2}e_2, \dots, \sqrt{\lambda_k}e_k)$ 。

b. 风险评估流程

本文通过下列流程得到国家的风险评估模型：

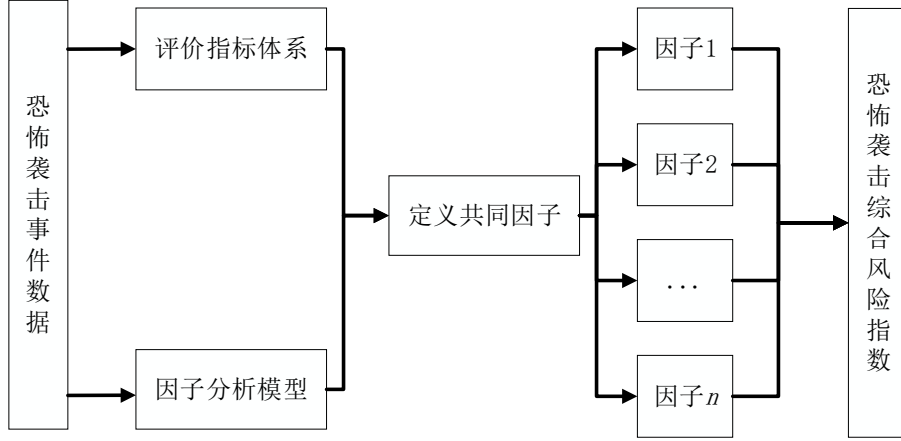


图 5 恐怖袭击风险分析模型流程图

c. 主成分分析显著性检验方法

为了检验原始数据是否适用主成分分析方法，我们用 **KMO** 检验方法进行检验。**KMO** (Kaiser-Meyer-Olkin) 检验统计量是用于比较变量间简单相关系数和偏相关系数的指标。主要应用于多元统计的因子分析。统计量为：

$$KMO = \frac{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2}{\sum_{i \neq j} \sum_{j \neq i} r_{ij}^2 + \sum_{i \neq j} \sum_{j \neq i} p_{ij}^2} \quad (11)$$

其中， $r_{ij} = \frac{\sum_{k=1}^n (x_k^i - \bar{x}^i)(x_k^j - \bar{x}^j)}{\sqrt{\sum_{k=1}^n (x_k^i - \bar{x}^i)^2} \sqrt{\sum_{k=1}^n (x_k^j - \bar{x}^j)^2}}$ ，表示变量 x^i 与变量 x^j 之间的简单相关系

数； p_{ij} 表示变量 x^i 与变量 x^j 之间的偏相关系数。**KMO** 越接近于 1，表示两个变量的相似性越高，越适合做因子分析，相反越低表示两个变量的相关性很弱，不能提取公共因子。**Kaiser** 提出检验标准，当 **KMO**>0.9 时非常适合，0.5~0.7 适合，0.5~0.7 不太适合，小于 0.5 不适合。

5.4.2 恐怖袭击风险分析

在 2015 年到 2017 年之间，总共发生了 39452 起恐怖袭击事件，死亡人数达 100169 人，受伤人数 108971，涉及国家 132 个。在 **IEP** 报告基础上，将国家分类，选取七个风险指标，分别为：近三年内总共发生恐怖袭击的件数，近三年总死亡人数，近三年总受伤人数，近三年财产损失总数，近三年死亡人数大于 10 人的件数，死亡人数大于 10 人的件数占总件数的百分比，平均每起恐怖袭击的死亡人数^[5]。

由表 8 可知，在 2015 到 2017 这三年间，伊拉克和阿富汗等国家受恐怖袭击最严重，且恐怖袭击主要发生在中东地区。

表 8 2015 年-2017 年恐怖袭击起数前 10 个国家

国家	起数	排名
Iraq	8577	1
Afghanistan	4959	2
India	2875	3
Pakistan	2826	4
Philippines	2045	5
Nigeria	1655	6
Somalia	1634	7
Yemen	1415	8
Egypt	1247	9
Syria	1207	10

KMO 检验的结果为 0.781，结果显示，提取的这七个特征，适合做主成分分析。

5.4.3 恐怖袭击主成分分析结果

首先我们采用主成分分析法分析 2015 年到 2016 年的数据，主成分的因子个数确定原则为特征根大于 1。主成分分析结果如表 9 所示，可以看出，前两个因子的累计百分比达到 94.9%，因此，选择两个因子是非常合理的，并且达到了降维的目的。第一个主因子 F_1 与五个指标相关：近三年内总共发生恐怖袭击的件数，近三年总死亡人数，近三年总受伤人数，近三年财产损失总数，近三年死亡人数大于 10 人的件数；主要反映了国家总体遭受恐怖袭击的风险，因此第一主因子可命名为基本风险因子；第二个主因子 F_2 与另外两个特征指标相关：死亡人数大于 10 人的件数占总件数的百分比，平均每起恐怖袭击的死亡人数；主要反映了国家应对恐怖袭击的能力，可理解为脆弱性因子。

表 9 因子变量方差结果表

因子	方差百分比	累计百分比
1	68.5	68.5
2	26.4	94.9
3	2.8	97.7
4	1.4	99.1
5	0.8	99.9

6	0.1	100
7	0	0

求得的因子载荷矩阵为:

$$A = \begin{bmatrix} 0.444 & 0.440 & 0.450 & 0.445 & 0.432 & 0.116 & 0.095 \\ -0.008 & -0.114 & -0.031 & -0.074 & -0.104 & 0.692 & 0.700 \end{bmatrix} \quad (12)$$

因此, 第一因子 F_1 和第二因子 F_2 的得分函数计算如下:

$$F_1 = 0.444x_1 + 0.440x_2 + 0.450x_3 + 0.445x_4 + 0.432x_5 + 0.116x_6 + 0.095x_7 \quad (13)$$

$$F_2 = -0.008x_1 - 0.114x_2 - 0.031x_3 - 0.074x_4 - 0.104x_5 + 0.692x_6 + 0.700x_7 \quad (14)$$

根据每个因子的方差比重可由下式计算得到综合得分:

$$F = 0.685F_1 + 0.264F_2 \quad (15)$$

1) 第一因子与第二因子分析

表 10 给出了总分排名前 50 的国家的得分因子。第一因子表示遭受恐怖袭击的基本风险指数, 从表里可以看出, 2015 到 2016 年间, 风险最高的前十个国家为: 伊拉克, 阿富汗, 尼日利亚, 叙利亚, 也门, 巴基斯坦, 索马里, 印度, 土耳其和菲律宾。从世界地图上看, 第一因子得分较高的国家主要分布在中东、南亚等地区, 且多为发展中国家, 说明他们很容易遭到侵略。

脆弱因子主要反映国家遭受恐怖袭击发生伤亡的风险高低, 从表中可以看出第二因子排名前十的国家为: 埃塞俄比亚, 乍得, 尼日尔, 科威特, 比利时, 南苏丹共和国, 布基纳法索, 象牙海岸, 突尼斯和中国。第二因子得分较高的国家除了比利时位于欧洲西北部, 其他主要分布在中东, 非洲等地方, 这些国家容易在恐怖袭击中发生伤亡, 排名第一的埃塞俄比亚在 2015 年和 2016 年间, 发生了 20 起恐怖袭击, 死亡人数达 327 人, 15 起有人死亡, 死亡事件比例达 75%。欧美等国家普遍得分低, 以为发达国家有比较完整的社会安全应急机制和法律, 公民整体素质高, 生活水平好。

2) 综合因子得分分析

综合评分综合第一因子和第二因子, 得分越高表明恐怖袭击风险水平越高, 得分为零表示处在平均风险水平, 负分表示处于平均水平以下。比较 2015-2016 年和 2017 年的综合得分可以看到, 伊拉克和阿富汗排名没有变化, 仍位居首位。图和图分别代表 2015-2016 年间前五十个国家的综合得分分布图和 2017 年前五十个国家的综合得分分布图。从图 6、图 7 中可以看到, 中东地区没有什么明显的变化, 且恐怖袭击比较集中, 均处于第四到第一风险区, 而南美颜色加深, 由第四风险区变为第二风险区, 说明综合得分升高, 其他地区都有减小的趋势。2017 年综合指数排名前十的国家为: 伊拉克, 阿富汗, 叙利亚, 中非共和国, 南苏丹共和国, 索马里, 巴基斯坦, 尼日利亚, 尼日尔和乍得。

表 10 恐怖袭击风险指数及时空特性

国家	2015-2016 年因子得分		2015-2016 年 综 合 得 分	2015-2016 年 综 合 排 名	2017 年 综合得 分	2017 年 综 合 排 名
	因子 1	因子 2				
Iraq	18.1	-1.97	11.9	1	8.8	1
Afghanistan	10.8	-0.831	7.19	2	8.38	2
Nigeria	5.42	1.61	4.14	3	2.24	8
Syria	4.45	1.87	3.54	4	3.28	3
Ethiopia	0.765	8.01	2.64	5	0.155	18
Yemen	3.05	0.366	2.19	6	1.16	12
Pakistan	3.56	-1.01	2.17	7	2.24	7
Chad	0.396	5.21	1.65	8	1.86	10
Somalia	2.23	0.252	1.59	9	2.33	6
Niger	0.508	4.57	1.56	10	1.97	9
Kuwait	0.0184	4.12	1.1	11	N/A	N/A
Cameroon	0.523	2.79	1.1	12	0.149	19
India	2.04	-1.13	1.1	13	1.69	11
Turkey	1.77	-0.553	1.07	14	0.021	22
Belgium	-0.115	3.21	0.768	15	-0.616	68
South Sudan	0.202	2.3	0.745	16	2.53	5
Libya	1.26	-0.453	0.743	17	-0.002	23
Egypt	1.28	-0.527	0.735	18	1.05	14
Philippines	1.3	-0.926	0.643	19	1.07	13
Burkina Faso	-0.153	2.78	0.629	20	-0.274	30
Ivory Coast	-0.239	2.56	0.513	21	-0.635	73
Tunisia	-0.179	2.3	0.485	22	-0.545	48
Ukraine	0.742	-0.386	0.406	23	-0.355	34
Democratic Republic of the Congo	0.344	0.551	0.381	24	0.952	15
China	-0.279	1.85	0.297	25	-0.335	33

Central African Republic	-0.147	1.35	0.255	26	2.61	4
France	-0.166	0.878	0.118	27	-0.562	52
Thailand	0.365	-0.785	0.0425	28	-0.123	25
Mali	-0.0655	0.124	-0.0123	29	0.314	17
Haiti	-0.462	1.09	-0.0271	30	N/A	N/A
Bangladesh	0.238	-0.781	-0.043	31	-0.546	49
Sudan	0.0684	-0.351	-0.0459	32	-0.253	29
Saudi Arabia	-0.16	-0.0367	-0.12	33	-0.503	43
Kenya	-0.266	0.202	-0.129	34	-0.202	28
Iran	-0.499	0.754	-0.143	35	0.132	20
Jordan	-0.501	0.7	-0.158	36	-0.614	67
Uganda	-0.532	0.418	-0.254	37	-0.604	65
Lebanon	-0.4	0.0468	-0.261	38	-0.551	50
Burundi	-0.315	-0.196	-0.268	39	-0.494	42
Japan	-0.556	0.426	-0.268	40	N/A	N/A
Myanmar	-0.395	-0.0737	-0.29	41	-0.137	26
Republic of the Congo	-0.554	0.336	-0.291	42	N/A	N/A
West Bank and Gaza Strip	-0.192	-0.637	-0.3	43	-0.465	41
United States	-0.365	-0.2	-0.303	44	0.0718	21
Peru	-0.6	0.226	-0.351	45	-0.567	54
Colombia	-0.33	-0.511	-0.361	46	-0.306	32
Kazakhstan	-0.61	0.185	-0.369	47	N/A	N/A
Madagascar	-0.617	0.101	-0.396	48	N/A	N/A
Germany	-0.407	-0.518	-0.416	49	-0.601	64
Tajikistan	-0.631	0.0525	-0.418	50	-0.633	71

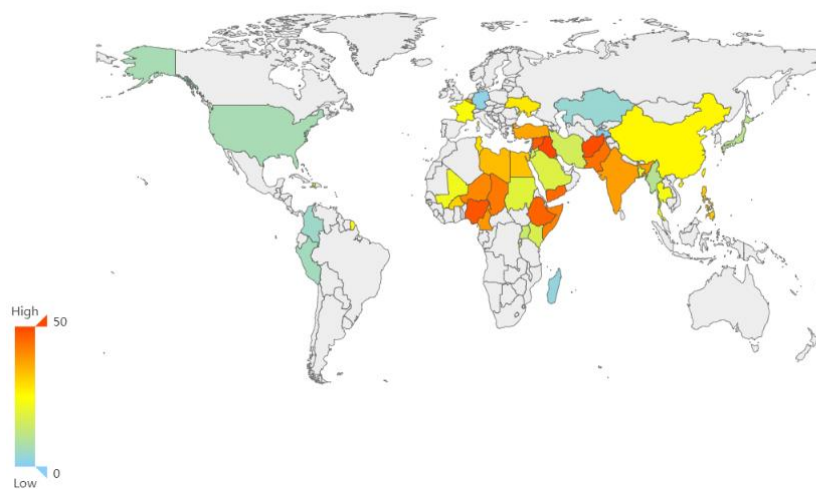


图 6 2015 年-2016 年全球恐怖袭击风险综合得分图

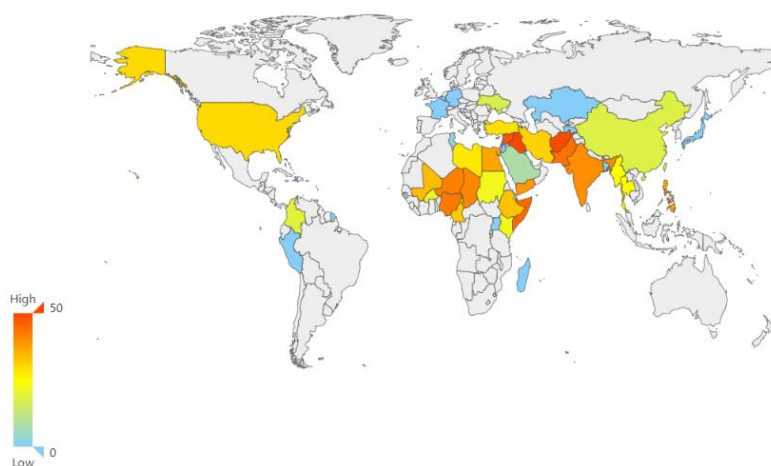


图 7 2017 年全球恐怖袭击风险综合得分图

5.4.4 恐怖袭击风险时空演变

从表 10 中可以看出，在 2015-2016 年时间段内，综合得分在平均线以上的国家有 28 个，而在 2017 年只有 20 个，表示恐怖袭击在收缩，且发生恐怖袭击的地区在减少。另一方面，恐怖袭击表现为聚集性，主要聚集在中非等地方，从空间角度看，南亚和中东为恐怖袭击高风险地区；北美、非洲和中国为第二风险地区；西欧和南美为第三风险地区；大洋洲恐怖袭击风险最低。据统计，根据 1998-2017 年期间世界范围内恐怖事件发生的次数，前十大恐怖组织分别是塔利班、伊拉克和黎凡特伊斯兰国（ISIL）、青年党、博科圣地等。这十大恐怖组织与宗教极端主义有关的组织有 6 个，其组织发动的恐怖袭击次数占 70% 以上。过去二十年，宗教极端主义是恐怖主义的主要原因。

5.4.5 未来反恐态势

从以上分析表明,中东、北非、南亚是恐怖组织重点袭击区域,处于这些地区的国家应采取武装防范措施。北非和撒哈拉以南非洲地区的恐怖组织非常复杂,加之信息闭塞,媒体关注度低,使得这里的安全形势和恐怖威胁被低估。2017年1月至10月,非洲发生了343起恐怖袭击,造成至少2600人死亡,比欧洲高22倍。而同年,“伊斯兰国”在叙利亚和伊拉克全线崩溃,标志着国际反恐取得巨大胜利。《2017年全球恐怖主义指数报告》^[6]指出:

- (1) 第一,伊斯兰极端宗教恐怖活动仍是国际恐怖活动的主要表现形态。
- (2) 第二,民族分离型恐怖活动急剧上升,成为新的全球恐怖震荡源。
- (3) 极右翼恐怖活动沉渣泛起,欧美成为重灾区。
- (4) 第四,网络恐怖活动进入新阶段,对全球安全形成新挑战。

5.5 问题四

为了进一步研究恐怖袭击的时空聚集模式,为国际反恐提供参考。在本文中我们针对近三年恐怖袭击在时间和空间上的聚集特性展开研究。利用基尼系数^[5]和 *Lorenz* 曲线,并基于帕累托图建立恐怖袭击聚集性评价模型,得到不同因素在空间上的聚集和蔓延特性;并从时间角度分析了死亡人数和受袭击国家数量的变化趋势。通过分析研究恐怖袭击事的时间和空间特性,分析得到下一年全球的反恐态势。

5.5.1 基尼系数与 *Lorenz* 曲线

基尼系数原本是一个用于衡量一个国家或地区居民收入差距的指标,引入恐怖袭击研究领域后用于表示极端事件的比例。首先,把不同因素根据其相关的恐怖袭击频率从小到大进行排序,并计算频率累积百分比作为 *y* 轴,将因素数量的累积百分比作为 *x* 轴,构建 *Lorenz* 曲线。

如果恐怖袭击在研究的因素上均匀分布,则 *Lorenz* 曲线与 $y = x$ 直线重合,说明不存在聚集性。如果恐怖袭击在研究因素上表现出偏好,则 *Lorenz* 曲线在前期增长缓慢,后期迅速增长。基尼系数则被定义为在[0,1]范围内 *Lorenz* 曲线的积分与直线 $y = x$ 的积分,基尼系数越大,说明分布越不均匀,计算公式如下所示。

$$G = 1 - 2 \int_0^1 L dx \quad (16)$$

为了更加清晰地分析恐怖袭击事件的主要因素和次要因素,引入帕累托图的思想将 *Lorenz* 累积频率曲线分成三个区间。以较少因素占据 0%~80%的累积频率的区间为 A 类区间,80%~90%为 B 类区间,剩下的是 C 类区间。A、B、C 三类区间的对应因素的重要程度依次降低。

5.5.2 恐怖袭击聚集性分布

恐怖袭击的目标类型可以体现恐怖主义者发动袭击事件的目标和意向。按照基尼系数,对全球范围内目标类型聚集性进行研究。首先,计算共 22 种目标类型的发生频率,并按升序排序,计算各目标类型的频率和累积频率百分比,以及

目标数量的累积百分比。最后，得到基尼系数为 0.705。具体计算过程如表 11 所示。

表 11 目标类型基尼系数计算过程

袭击目标	编号	袭击目标累积百分比	频率	频率百分比	累积频率百分比	基尼系数
流产有关	1	0.045	7	0.0002	0.0002	0.705
游客	2	0.091	24	0.0006	0.0007	
食物或水供应	3	0.136	39	0.0010	0.0018	
其他	4	0.182	40	0.0010	0.0028	
海事（包括港口和海上设施）	5	0.227	52	0.0013	0.0041	
机场和飞机	6	0.273	73	0.0018	0.0059	
电信	7	0.318	122	0.0031	0.0090	
非政府组织（NGO）	8	0.364	159	0.0040	0.0131	
政府（外交）	9	0.409	340	0.0086	0.0217	
新闻记者	10	0.455	446	0.0113	0.0330	
暴力政党	11	0.500	484	0.0123	0.0453	
教育机构	12	0.545	656	0.0166	0.0619	
运输（航空除外）	13	0.591	755	0.0191	0.0810	
公用事业	14	0.636	762	0.0193	0.1003	
恐怖分子/非州立民兵组织	15	0.682	812	0.0206	0.1209	
宗教人物/机构	16	0.727	935	0.0237	0.1446	
商业	17	0.773	2832	0.0718	0.2164	
未知	18	0.818	3019	0.0765	0.2929	
政府（一般意义）	19	0.864	3114	0.0789	0.3719	
警察	20	0.909	5360	0.1358	0.5077	
军事	21	0.955	7920	0.2007	0.7085	
公民自身和私有财产	22	1.000	11501	0.2915	1.0000	

同样地，可以得到国家、武器、区域、袭击方式的 Lorenz 曲线，如图 8 所示。可以看出，包括袭击目标在内的不同因素都具有较强的聚集性。特别是国家因素的累积频率在国家数量累计百分比小于 0.85 时缓慢增长，到 0.85 之后变化增长十分迅速。表明 15% 的国家所发生的恐怖袭击次数占全世界的 90%，即恐怖袭击存在热点国家。袭击方式尽管相较于其他四个因素而言曲线的变化趋势稍显平缓，但同样表现出较强的聚集性。

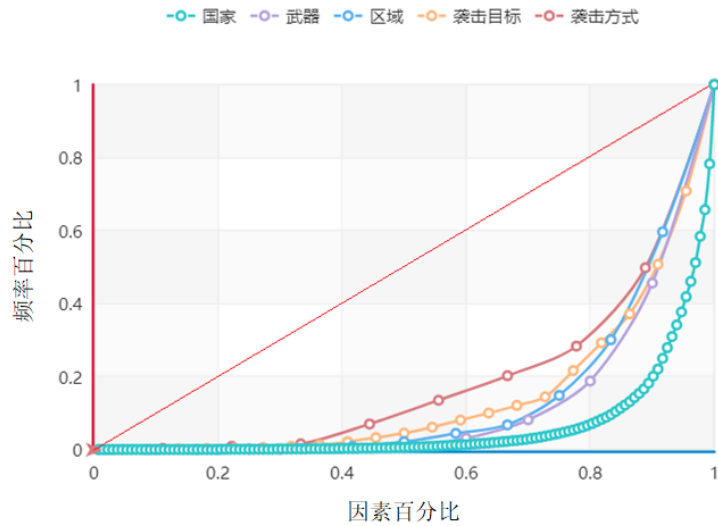


图 8 不同因素恐怖袭击频率 Lorenz 曲线

将 Lorenz 曲线根据帕累托图划分为 A、B、C 三个区间，可以清楚地从 A 区间得到不同因素的聚集度和热点。如表 12 所示。

表 12 不同因素恐怖袭击频率热点

因素	Gini	热点
国家	0.870	伊拉克、阿富汗、哥伦比亚、叙利亚、巴基斯坦、萨尔瓦多、菲律宾、土耳其、印度、黎巴嫩、秘鲁、阿尔及利亚、美国、南非、意大利、俄罗斯、尼加拉瓜
武器	0.746	爆炸物/炸弹/炸药、轻武器
区域	0.717	中东和北非、南亚、南美
袭击目标	0.705	军事、公民自身和私有财产、商业、警察、机场及飞机
袭击方式	0.618	轰炸/爆炸、武装袭击、设施/基础设施攻击

通过基尼系数和 Lorenz 曲线可知，在国家、武器、区域、袭击目标、袭击方式这五个因素中，国家因素表现出了最强的聚集性。图 9 展现了所有国家和地区的恐怖袭击频率。显然，大部分的恐怖袭击事件发生在 20 个热点国家。接下来分析武器类型、攻击类型和目标类型三个因素在不同国家的聚集情况。

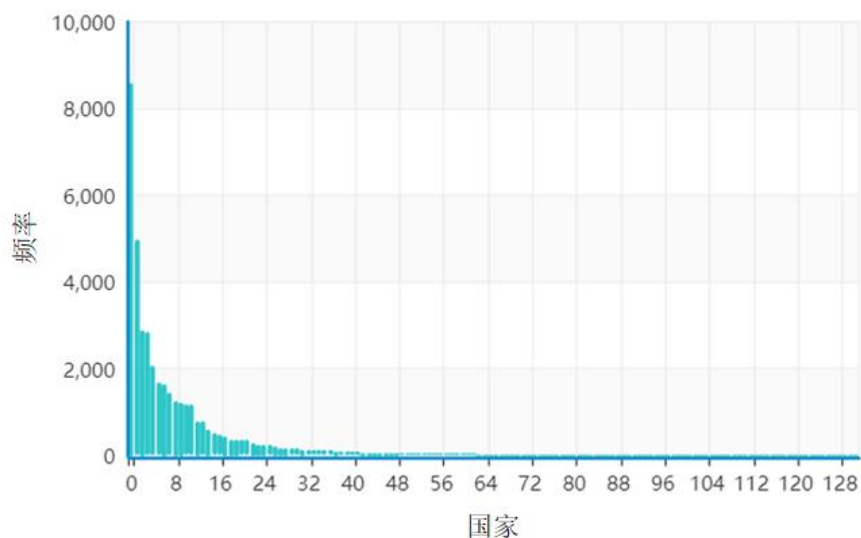


图 9 所有国家和地区恐怖袭击发生频率

表 13 和图 10 分别表示武器类型的基尼系数和 *Lorenz* 曲线。除了破坏设备和其他这两种武器类型的基尼系数偏低以外，其他武器表现出了强烈的分布不均衡的特点。其中，基尼系数最高的是假武器（1）和爆炸物（0.870）和轻武器（0.820）。爆炸物武器恐怖袭击频率最高，并且有 55.5% 的恐怖袭击事件发生在前 5% 的国家，70% 以上的恐怖袭击发生在 10% 的国家，其热点国家为：伊拉克、秘鲁、哥伦比亚、巴基斯坦和印度。

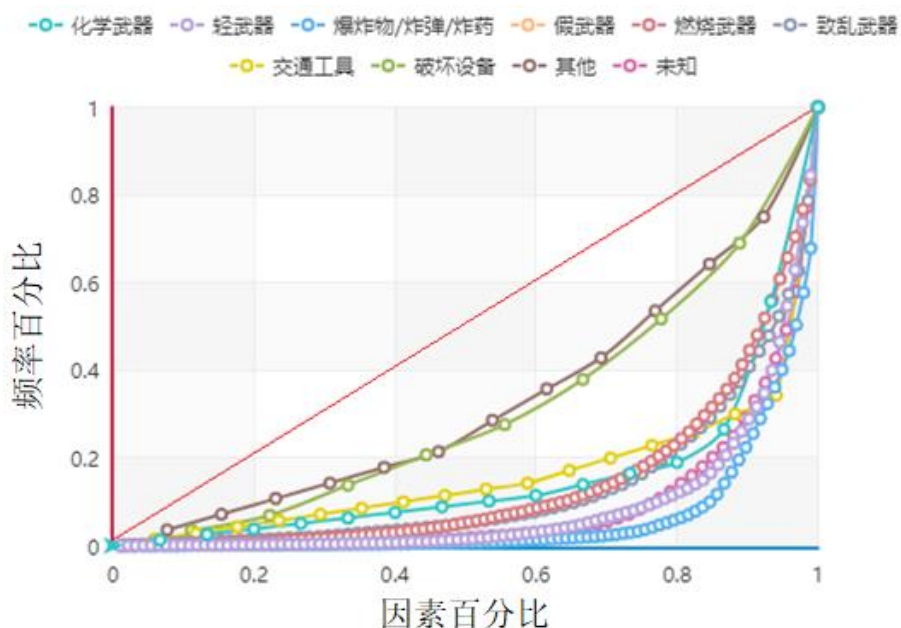


图 10 不同武器类型的恐怖袭击频率 *Lorenz* 曲线

表 13 武器类型热点分析情况

编号	武器	Gini	频率	5%国家	10%国家	20%国家
1	化学武器	0.683	79	0.013	0.443	0.810
2	轻武器	0.820	10580	0.453	0.711	0.877
3	爆炸物/炸弹/炸药	0.870	21448	0.555	0.744	0.936
4	假武器	1	1	1	1	1
5	燃烧武器	0.718	1989	0.343	0.554	0.756
6	致乱武器	0.736	1028	0.426	0.591	0.768
7	交通工具	0.673	70	0.014	0.657	0.743
8	破坏设备	0.379	29	0.034	0.098	0.310
9	其他	0.349	28	0.036	0.25	0.357
10	未知	0.817	4200	0.507	0.670	0.859

表 14 和图 11 分别表示攻击类型的基尼系数和 *Lorenz* 曲线。分析各攻击类型的基尼系数，劫持的基尼系数和发生频率都是最小，相比于暗杀和爆炸等攻击类型，其空间聚集度最小，说明劫持的空间密集度较小，是较为普遍的攻击方式。基尼系数最大的为爆炸，其大部分热点为中东与北非地区。且爆炸事件发生频率远远超出其他攻击类型，这也说明了在众多攻击类型中，应把爆炸作为重点预防的恐怖袭击方式，在重要公共场所加强对爆炸物的排查。

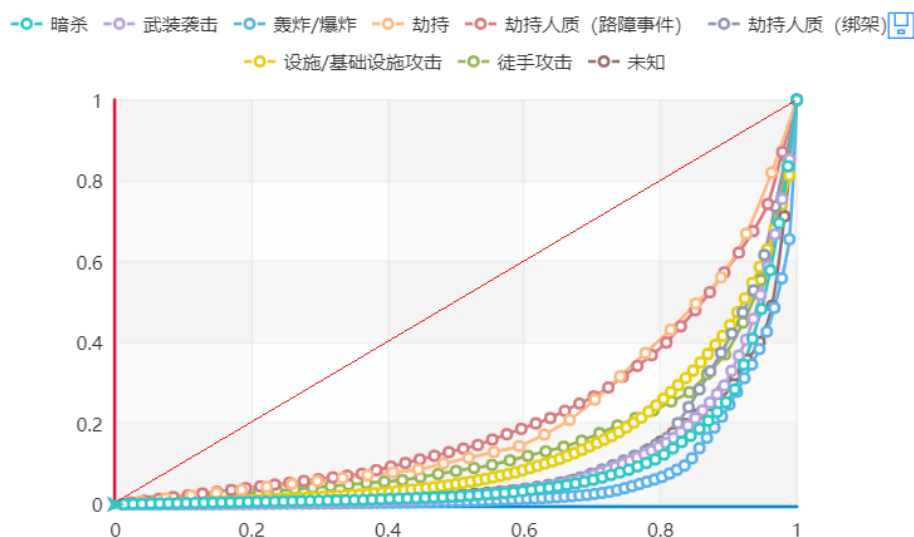


图 11 不同攻击类型的恐怖袭击频率 Lorenz 曲线

表 14 攻击类型热点分析情况

编号	攻击类型	Gini	频率	5% 国家	10% 国家	20% 国家
1	暗杀	0.816	2564	0.421	0.716	0.877
2	武装袭击	0.801	8466	0.413	0.670	0.848
3	轰炸/爆炸	0.866	19794	0.573	0.753	0.932
4	劫持	0.566	139	0.180	0.331	0.568
5	劫持人质（路障事件）	0.556	225	0.258	0.378	0.600
6	劫持人质（绑架）	0.772	3207	0.383	0.578	0.836
7	设施/基础设施攻击	0.714	2152	0.370	0.578	0.836
8	徒手攻击	0.691	267	0.262	0.551	0.745
9	未知	0.809	2638	0.509	0.681	0.845

表 15 和图 12 分别表示攻击类型的基尼系数和 *Lorenz* 曲线。游客的基尼系数最小（0.247），“流产相关”基尼系数最大（1）。“流产相关”聚集性最强，在近三年只发生在美国，只有 7 起，可见该类袭击主要发生在发达国家，大部分针对“流产”发动恐怖袭击的动机通常是反对人类堕胎。

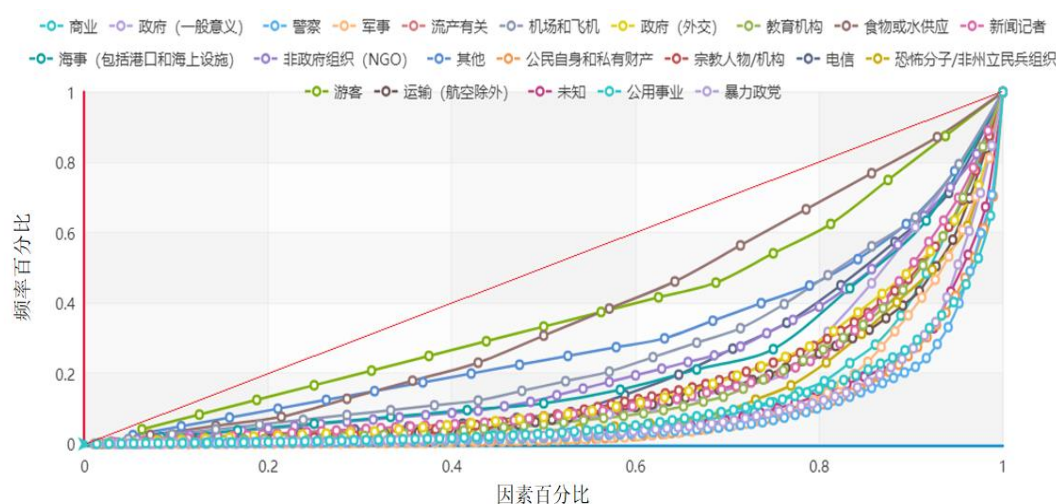


图 12 目标类型的恐怖袭击频率 Lorenz 曲线

表 15 目标类型热点分析情况

编号	目标类型	Gini	频率	5% 国家	10% 国家	20% 国家
1	商业	0.808	2832	0.546	0.704	0.843
2	政府（一般意义）	0.814	3114	0.490	0.728	0.876
3	警察	0.847	5360	0.600	0.781	0.889

4	军事	0.801	7920	0.392	0.584	0.854
5	流产有关	1	7	1	1	1
6	机场和飞机	0.487	73	0.205	0.356	0.521
7	政府（外交）	0.660	340	0.265	0.453	0.679
8	教育机构	0.695	656	0.300	0.492	0.732
9	食物或水供应	0.256	39	0.016	0.128	0.231
10	新闻记者	0.666	446	0.300	0.484	0.742
11	海事（包括港口和海上设施）	0.566	52	0.059	0.365	0.558
12	非政府组织（NGO）	0.549	159	0.176	0.352	0.610
13	其他	0.414	40	0.018	0.225	0.475
14	公民自身和私有财产	0.843	11501	0.528	0.724	0.887
15	宗教人物/机构	0.665	935	0.307	0.488	0.702
16	电信	0.579	122	0.042	0.287	0.549
17	恐怖分子/非州立民兵组织	0.742	812	0.381	0.505	0.768
18	游客	0.247	24	0.048	0.125	0.375
19	运输（航空除外）	0.699	755	0.302	0.565	0.743
20	未知	0.819	3019	0.462	0.724	0.850
21	公用事业	0.756	762	0.216	0.516	0.822
22	暴力政党	0.656	484	0.205	0.385	0.682

5.5.3 恐怖袭击时间性分析

上一节从空间特性分析了不同因素的密集度，根据密集性采取有效的措施来防控恐怖主义。接下来从时间特性来分析恐怖袭击的蔓延特性。以 2015 至 2017 年全部国家和地区为研究对象，分别计算每年内每个月发生恐怖袭击事件频率、死亡人数的基尼系数，并给出每年受袭击的国家数目，如表 16 所示。

表 16 恐怖袭击 2015-2017 年基尼系数

年份	频率 Gini	死亡人数 Gini	受袭击国家数目
2015	0.054	0.095	99
2016	0.045	0.077	108
2017	0.070	0.101	102

可以看出，每年的基尼系数都相差无几，表明恐怖袭击聚集性与时间的相关度较小。从 2015 到 2017 年，恐怖袭击风险最大的五个国家从伊拉克、阿富汗、尼日利亚、叙利亚、埃塞尔比亚转移到了伊拉克、阿富汗、叙利亚、中非共和国、

南苏丹^[7]。

5.5.4 小结

本文利用基尼系数和 Lorenz 曲线,并基于帕累托图建立恐怖袭击聚集性评价模型,分别考虑了国家、区域、武器类型、攻击类型、攻击目标等因素的空间聚集性,并分析国家层面的武器类型、攻击类型、攻击目标的空间聚集性,得到上述考虑的因素大多具有较强聚集性的结论。最后,从时间角度分析了并从时间角度分析了死亡人数和受袭击国家数量的基尼系数的变化,发现近三年随时间变化趋势并不明显。

六、模型的评价与推广

6.1 模型优点

1. 利用等级全息建模 (HHM) 技术全面考量恐怖袭击事件情景,构建恐怖袭击危害性指标体系进行危害指数量化。针对初始权重含有主观因素一问题,分别建立基于经验阈值的危害分级模型与基于加权聚类的危害分级模型,对恐怖袭击危害指标体系权重进行动态调整,提高分级准确度。

2. 采用 DBSCAN 算法进行恐怖组织的自分类,无须人工干预,设定聚类总数。

3. 引入 Lorenz 曲线和帕累托图分类管理法分析,使用基尼系数量化恐怖袭击的聚集性分布,从而有效避免了时空特性的聚集特性难以量化表示的缺点。

6.2 模型缺点

1. 问题一的初始权重部分根据法律判处严重程度进行设定,若通过专家评分也许可提高准确度。

2. 数据缺失值过多,信息不完整,经数据清洗后损失了部分信息。

6.3 模型的改进与推广

本文在对恐怖袭击事件进行评估并分级时,无法验证其性能好坏。为了提高模型的准确度,可以在一定范围内引入专家知识,对模型进行小范围的验证。例如,在任务一中根据专家知识为小部分事件进行评级,并用这些结果验证与模型结论相比较,若几乎一致,则模型建立相对来说更准确。

文章提出的诸多模型面向现实生活中的恐怖袭击事件,然而随着互联网的发展,恐怖袭击方式已经不仅仅局限于线下攻击,如恐怖音频、网络病毒、木马程序等随时可借助网络系统、通信系统以及各种平台发动一场线上袭击,而本文建立的恐怖袭击危害量化与分级模型于此场景同样适用。

七、参考文献

- [1] 夏一雪, 兰月新, 王沙骋. 大数据环境下网络恐怖主义风险分析与防范对策研究[J]. 情报杂志, 36(11):16-22,2017.
- [2] Yacov Y. Haimes. 风险建模、评估和管理 [M]. 西安: 西安交通大学出版社, 20-130, 2007: 677.
- [3] 刘建平, DBSCAN 密度聚类算法, <https://www.cnblogs.com/pinard/p/6208966.html>, 20180918
- [4] 杨帆, 徐建刚, 周亮. 基于 DBSCAN 空间聚类的广州市区餐饮集群识别及空间特征分析[J]. 经济地理, 36(10):110-116, 2016.
- [5] 李国辉. 全球恐怖袭击时空演变及风险分析研究[D]. 中国科学技术大学, 2014.
- [6] 马愿. 《2017 年全球恐怖主义指数报告》解读[J]. 国际研究参考, 2018(2).
- [7] 位珍珍. 后 911 时代恐怖主义的 GTD 数据分析[J]. 情报杂志, (7):10-15, 2017.