

# “华为杯”第十五届中国研究生 数学建模竞赛

题 目            基于模糊聚类 and 距离测度的恐袭  
数据的量化分析

---

## 摘            要：

自美国“9.11”事件发生以来，全球恐怖主义日益嚣张，给人类社会的和谐安定带来了巨大的威胁。恐怖事件具有极大的杀伤力与破坏力，恐怖分子是人类共同的敌人对恐袭历史数据进行深度剖析能够积极融入我国反恐斗争的行列，对于在新形势下维护国家安全与社会稳定具有十分迫切的意义。

本文以全球恐怖主义数据库（GTD）中 1998-2017 年发生的恐袭事件为主要数据来源，首先我们构建了基于数据分析的量化分级模型、运用了模糊 C 均值（FCM）聚类算法研究了近 20 年来危害程度最高的十大恐怖袭击事件，并将恐怖事件按照危害程度划分为 5 级；其次，基于模糊识别理论的 K-means 算法模型结合距离测度来识别危害性排名前五的嫌疑人；然后我们研究了近三年来恐怖袭击事件发生的主要原因、时空特性、蔓延特性并运用博弈模型进行解释并对反恐斗争给出自己的见解；最后我们进行数据挖掘，使用模糊层次分析法建立城市遭受恐怖袭击的风险评估模型，并应用于对我国 15 个主要城市进行风险评估排序。

针对问题一，首先对原始数据进行预处理；然后对 1998-2017 年近 20 年以来的恐怖袭击事件的指标进行量化分级并使用 Fuzzy-C 均值聚类将事件聚成 5 类。其次，我们基于层次分析-因子分析混合的综合评分模型对事件进行评价从而得出 1998-2017 年以来危害程度最高的十大恐怖袭击事件，其中得分前三的事件为 200109110004、200109110005 与 201408090071，具体十大恐怖事件结果如表 4.11 所示。基于这两种模型我们对典型事件的危害级别进行评级，将恐怖事件按照危害程度分为 5 级，具体数据如表 4.12 所示。

针对问题二，我们综合考虑了 1998-2017 年以来已被声称负责的恐怖袭击事件、2015-2016 年度未被声称负责的恐怖袭击事件，首先对事件进行分析，包括分析其区位特征、受害者特征、袭击方式特征以及武器类型特征；然后构建了基于模式识别基础下的 K 均值算法。同时，我们运用信息熵手段得出指标权重，选取权重较为靠前的 8 个指标作为主要特征，进行距离测度，得出未被声称负责的案件的可能嫌疑人，排名最靠前的两位是 Islamic State of Iraq and the Levant (ISIL) 与 Taliban，具体情况见表 5.6 所示。最后我们根据表 5.6 的结果结合对案件和组织的特征分析，得出问题中表 2 典型事件的嫌疑度，如表 5.7 所示。

针对问题三，首先我们通过对 2015-2017 年恐怖袭击事件发生的主要原因、时空特征、蔓延特性、级别分布等规律进行分析探讨；然后得出对蔓延态势分析的结果，结果表明在 2015-2017 年间，恐怖袭击事件呈逐年下降的趋势，袭击成功率也有一定程度的下降，政府反恐力量的投入与应对及时性是制约恐怖主义活动的关键因素。最后结合误对策博弈模型，得出对于北美地区未来的反恐态势来说，当恐怖分子采取袭击策略的比例大于最优策略均衡临界值时，政府应采取防御策略；当恐怖分子采取袭击策略的比例小于最优策略临界值时，政府可以适宜降低防御程度，但不应放弃防御策略。同时，该模型也同样适用于国际反恐态势分析。

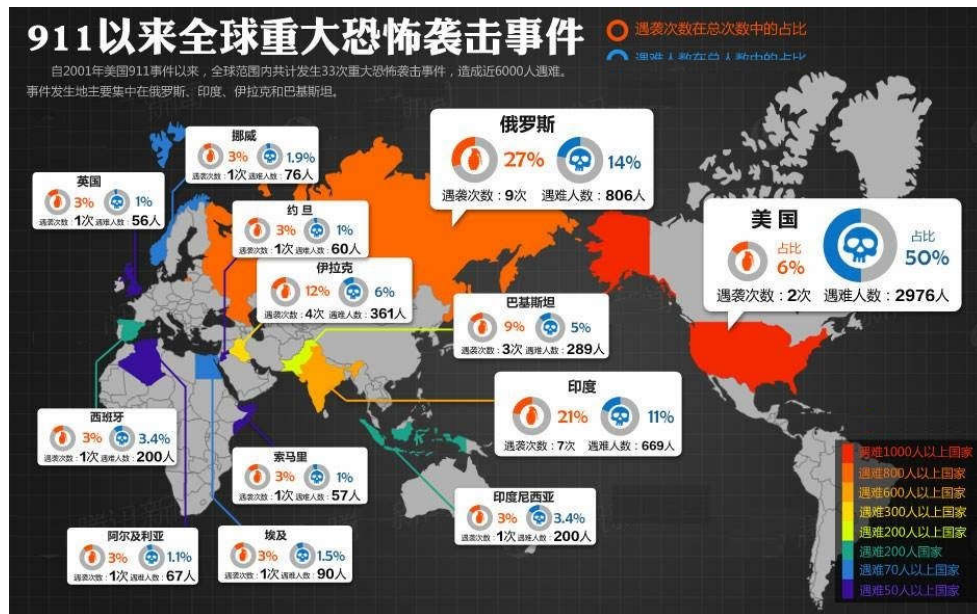
针对问题四，我们进行数据挖掘，使用模糊层次分析法建立城市遭受恐怖袭击的风险评估模型，并以我国 15 个城市为例，然后应用模型来预测这 15 个城市中最可能遭受恐怖袭击的城市，利用 ArcGIS 实现城市遭受恐怖袭击风险的专题图，将显示结果结合城市的地理位置与社会背景，最后我们从在这 15 个城市中得出最易遭受恐怖袭击的城市是乌鲁木齐与拉萨。

**关键词：**恐怖袭击；量化分级；距离测度；模糊聚类

# 1 问题重述

## 1.1 问题背景

恐怖袭击是指极端分子或组织人为制造的、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行为，它不仅具有极大的杀伤性与破坏力，能直接造成巨大的人员伤亡和财产损失，而且还给人们带来巨大的心理压力，造成社会一定程度的动荡不安，妨碍正常的工作与生活秩序，极大地阻碍经济发展。



恐怖主义的历史可以溯源到法国革命。在过去二百多年的发展历程中，恐怖主义经历了三次跌宕起伏的浪潮：

- 第一次浪潮是在 20 世纪初期以俄罗斯等国家无政府主义者实施的暗杀、爆炸为代表，在无政府主义思想的推波助澜之下，全球恐怖活动在 20 世纪二三十年代达到了第一次高潮

- 第二次浪潮起始于 20 世纪 60 年代末、70 年代初。以 1972 年慕尼黑奥运会人质事件为标志，恐怖主义在中东、西欧等地区狂飙突进。这一阶段的恐怖主义是在殖民主义体系瓦解、民族解放与民族自决运动风起云涌的大环境中产生的

- 第三次浪潮则以 2001 年的“9·11 事件”为标志，以宗教极端主义为核心思想体系的“全球圣战”运动成为这一阶段恐怖主义扩张的强力推手和主要表现。

自美国“9·11”恐怖袭击事件以来，人们越来越认识到恐怖主义带来的巨大经济损失、心理挫折及政治影响。这些恐怖袭击事件让各国人民都深受其害，生命财产安全、地区安全稳定都受到巨大威胁。仅 2017 年就发生了 22487 起恐怖袭击事件，超过上百个国家受到恐怖主义的伤害。自杀式袭击、劫机、绑架、爆炸等事件层出不穷，恐怖袭击的攻击方式变幻莫测，共计造成上万人死亡。

文献<sup>[1]</sup>归纳了极端意识形态指导下的恐怖袭击组织和原因如下：

极端意识形态	典型组织	原因分析
极端民族主义	英国“爱尔兰共和军”、巴解组织、西班牙“埃塔”组织	政治、民族、宗教
极端宗教主义	伊斯兰原教旨主义组织	宗教、民族、政治
政治极左主义	德国“红军旅”、意大利“红色旅”	政治
极端反人类主义	纳粹主义	政治、种族

由于受技术、社会等多种因素的限制，恐怖主义在 20 世纪主要表现为各种本地化的恐怖事件。一些事件（如 1972 年慕尼黑奥运会事件）背后虽然有复杂的国际因素，但是，这些事件本身在辐射范围和影响效果上看，基本还仅限于个别国家、少数地区，全球性的影响还不十分明显。进入 21 世纪之后，随着信息技术的发展及应用，恐怖主义全球化的怪象也悄然形成。恐怖主义的全球化意味着当今时代，那些看上去是本地化的恐怖袭击事件，其产生的原因和事后的效果都可能超越本地，其背后都会有国际的诱因和国际性的影响。也就是说，看上去是本地化的恐怖袭击事件在全球化时代都蕴含着全球性的意义。

恐怖主义是人类共同威胁，打击恐怖主义是每个国家应该承担的责任。对恐怖袭击事件相关数据的深入分析有助于加深人们对恐怖主义的认识，为反恐防恐提供有价值的信息支持。

## 1.2 问题重述

针对以上现状，依据危害性对恐怖事件进行分级并列举出危害程度比较高的恐怖袭击事件就更有理论和现实意义。同时，由于在历年来国际上的多起恐怖事件由于种种原因未能确定作案者，因此，依据事件特征来发现恐怖袭击事件制造者或者是嫌疑人对于提高恐怖袭击案件的破案效率和维护世界安定拥有重大意义。通过已知事件的属性来预测和分析未来反恐态势能够增加全球反恐斗争的效率。

本文以此为背景，基于全球恐怖主义数据库（GTD<sup>1</sup>）中 1998—2017 年世界上发生的恐怖袭击事件的记录数据，主要考虑以下几个问题：

问题一：依据危害性对恐怖袭击事件分级

迄今为止，尚未有与地震、交通事故和气象灾害类似的灾害等级划分标准，结合现在信息处理技术，也为了能够给出能够反映恐怖袭击自身在人数伤亡、经济损失、社会秩序、公共安全和交通状况等方面的特征，考虑建立基于数据分析的恐怖袭击量化分级模型，并对给定的恐怖袭击事件记录进行分级。

问题二：依据事件特征发现恐怖袭击事件制造者

针对可能的新发生的以及尚未被界定和明确组织者及其原因的恐怖袭击事件，如何利用现有的恐怖袭击数据库对未知发起者和起因的恐怖袭击事件进行有针对性的甄别无疑对于有效对抗恐怖组织和恐怖袭击具有重要的现实意义。

材料中有多起恐怖袭击事件尚未确定作案者，将可能是同一个恐怖组织或一个人在不同时间、不同地点多次作案的若干案件串联起来统一组织侦查，有助于提

<sup>1</sup> <http://www.start.umd.edu/gtd>

高破案效率并尽早发现新生或隐藏的恐怖分子。针对在 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，运用数学建模方法寻找上述可能性，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，对应的未知作案组织或个人标记不同的代号，并按该组织或个人的危害性从大到小选出其中的前 5 个，记为 1 号-5 号。并对列出的恐袭事件，按嫌疑程度对 5 个嫌疑人排序并填入表中。

问题三：对未来反恐态势的分析

建立量化分析模型的最终目的是为了有效预防可能的恐怖袭击以及有效对抗恐怖组织。依据数据资料并结合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，用图或表给出研究成果并提出对反恐斗争的见解和建议。：

问题四：数据的进一步利用

通过分析和挖掘现有恐怖袭击事件数据库的信息，除了给定的问题，我们还可以尝试进一步区域遭受恐怖袭击的风险评估等。

## 2 基本假设

考虑到现实情况，我们有如下假设：

- 1.假设提供的数据是真实可靠的；
- 2.假设数据中对地区等变量的划分标准是可靠的；
- 3.假设数据集中个别缺失数据对结果不会产生重大影响；

## 3 项目符号

$x_{in}$ ：为 AHP 方法中第  $i$  个事件的第  $n$  个指标对应的值

$Y_i$ ：记为 AHP 方法计算出的综合评价分

$F_i$ ：称为公共因子

$\varepsilon_i$ ：为  $X_i$  的特殊因子，

$a_{ij}$ ：为因子载荷

$v_k$ ：为每个聚类的中心特征向量

A1：1998-2017 年恐怖袭击数据中死亡总数

A2：1998-2017 年恐怖袭击数据中财产损失

A3：1998-2017 年恐怖袭击数据中武器类型

A4：1998-2017 年恐怖袭击数据中地区

A5：1998-2017 年恐怖袭击数据中受伤总数

A6：1998-2017 年恐怖袭击数据中攻击类型

A7：1998-2017 年恐怖袭击数据中是否超出人道主义

A8：1998-2017 年恐怖袭击数据中财产损失程度

## 4 问题一模型建立与求解

### 4.1 问题分析

对于问题一，需要建立基于数据分析的量化分级的模型，将附件 1 给出的事件按危害程度从高到低分为一至五级，本题这里使用模糊 C 均值(Fuzzy C-means, 简称为 FCM) 聚类方法，将事件分为了五类；对列出近二十年来危害程度最高的十大恐怖袭击事件这个问题，我们使用基于层次分析-因子分析混合的综合评分模型对事件进行评分，排序出前十。

### 4.2 指标选择及数据来源

进入 21 世纪以来，恐怖袭击案的多次发生一直在威胁着全世界人民的安全，因此在依据危害性对恐怖袭击进行分级的过程中必须要对恐怖袭击的影响因素进行分析，能够对恐怖袭击的分级奠定一定的基础。影响恐怖袭击的因素内部因素和外部因素，见下图 4.1 所示。

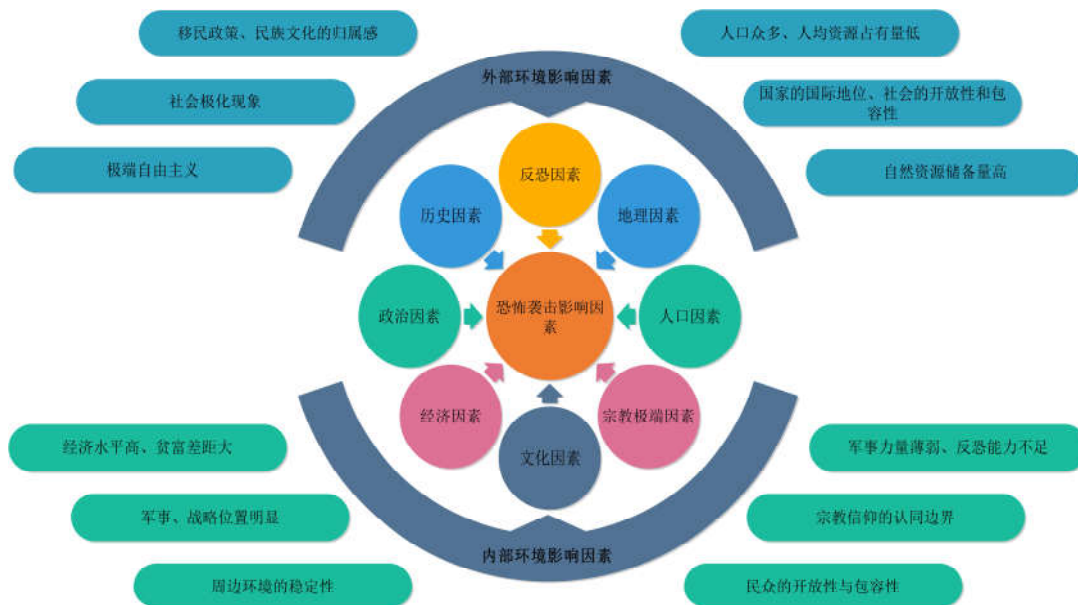


图 4.1 恐怖袭击的影响因素

本节数据来源为附件 1 中全球恐怖主义数据库（GTD）中 1998—2017 年世界上发生的恐怖袭击事件的记录数据。

问题一旨在依据危害性对恐怖袭击事件量化分级，因此指标选取应尽可能与恐怖袭击事件带来的危害高度相关，且各指标自身信息熵表现出的差异越大越好。结合上述恐怖袭击影响因素的背景选取指标，选出 20 个指标，并依据各指标反映出的危害性将其分为五个二级指标，分别为人员伤亡指标、财产损失指标、不良社会影响指标、袭击类型及规模指标以及蔓延范围和恐怖集团指标。如表

4.1 示：

表 4.1 恐怖袭击事件危害性指标体系

一级指标	二级指标	三级指标
恐怖袭击事件危害性	人员伤亡指标	死亡总数
		受伤总数
		受害者类型
		人质或绑架受害者总数
	财产损失指标	财产损失
		财产损失程度
	不良社会影响指标	恐怖袭击目的
		对人群社会影响
		是否超出人道主义
		是否为持续事件
	袭击类型及规模指标	自杀式袭击
		攻击类型
		武器类型
		成功的攻击
		抓获的凶手数量
	蔓延范围及恐怖集团指标	国家
		地区
		附近地区
		第一可疑、涉嫌犯罪集团 声称负责

### 4.3 数据预处理

预处理主要包括对数据中的缺失值、异常值进行处理，对变量指标无量纲化等等。具体处理方法及步骤如下：

#### (1) 异常数据的挖掘与处理

所谓数据挖掘是指按照既定目标，对大量的统计数据进行探索，揭示其隐藏的规律并将之模型化的一种先进有效的方法，其实质是对海量数据的识别过程。识别统计数据的第一步就是检验异常值的存在，常用的异常值挖掘方法有 $3\sigma$ 检测法、聚类分析法和回归分析等。

在本题中，利用 Stata 软件对选择的 20 个指标数据研究发现，部分指标对恐怖袭击事件的危害性的度量存在误导作用，比如“武器类型”这一指标并非按照武器杀伤力或危害程度进行排序，处理不当可能会给结果带来较大误差。综合比较几种检验方法后选择效果更好的 $3\sigma$ 准则检验异常值。在该准则下，若统计数据没有明显的上升或下降趋势，且都分布在其均值周围，标准差 $\sigma$ 就能很好地反映其离散程度。统计数据若是来自某一总体的样本，要求统计数据与其均值的偏差超过 $3\sigma$ 的比例不超过 $1/9$ ，否则将那些均值之差的绝对值超过 $3\sigma$ 的统计数据

视为异常数据。

常用的检验异常值是否剔除的方法有 t 检验、拉依达准则（ $3\sigma$  准则）、拉格布斯准则和迪克逊准则等，本题采用  $3\sigma$  准则剔除异常值。即根据每个样本点距离样本中心的距离判断是否将可疑数据  $x_d$  从该组数据中剔除，至于选择 3 倍的标准差（即  $3\sigma$ ）还是 2 倍的标准差与显著性水平  $\alpha$  有关。显著性水平  $\alpha$  表示的是检验出错的几率为  $\alpha$ ，或检验的置信度为  $1-\alpha$ ， $3\sigma$  相当于显著性水平为 0.01， $2\sigma$  相当于显著性水平为 0.05。

#### （2）缺失值的处理

本数据库中缺失数据较多，本题的数据库检测发现，多数变量都有较多的缺失值，而数据大量缺失很可能对恐怖主义事件危害性分级的定位有较大影响，因此本部分的处理相当重要。处理确实值的方法主要有删除元组、数据补齐、不处理。例如死亡人数(nkill)变量，约 4280 个数据缺失，对这样的缺失值数据我们采取的是数据补齐的方式。首先根据别其他相关字段进行预估，如事件摘要（summary）变量，该字段就会显示该事件伤亡人数预估量。若无法根据相关字段进行填充的，使用了 K 最近距离邻法，即根据欧式距离或相关分析来确定距离具有缺失数据样本最近的 K 个样本，将这 K 个值加权平均来估计该样本的缺失数据。

#### （3）无量纲化

无量纲化是使不同规格的数据转换到同一规格。本题变量的量纲众多，如变量死亡总人数、财产损失价值及攻击类型这样的分类变量等。常见的无量纲化方法有标准化、归一化和区间缩放法等。标准化的前提是特征值服从正态分布，标准化后，将其转换成标准正态分布；归一化是对原始数据进行线性变换，将其映射到[0,1]之间，该方法也被称为离差标准化；区间缩放法利用了边界值信息，将特征的取值区间缩放到某个特点的范围，如[0,1]。

本题对变量进行 min-max 归一化，变换函数为：

$$x' = \frac{x - \min}{\max - \min} \quad (4.1)$$

其中 min 是样本中最小值，max 是样本中最大值。

## 4.4 模型的建立及求解

### 4.4.1 基于 AHP—FA 的综合评价模型

#### 1. 基于层次分析法（AHP）的综合评价

AHP 是一种定性与定量相结合的方法，通过分析各指标之间的关系、建立对比矩阵，将复杂问题分解为若干层次和若干指标，在各指标之间进行比较和计算，从而可较为合理地确定指标权重。

层次分析法（AHP）是 20 世纪 70 年代由美国运筹学家 Santy 提出的，经过半个世纪的发展已经发展成一种较为比较成熟的解决多目标的复杂问题的有效方法。它的基本原理是对评价体系的各种要素分解成目标层、中间层和决策层，逐层分析形成一个从上到下的递阶的层次关系，但必须注意遵循个因素之间的相对独立性和可操作原则。应用层次分析法一般经过以下 4 个步骤：

（1）建立层次结构模型。将决策的目标、考虑的因素和决策对象按它们的相互关系分为目标层、中间层和决策层，同一层的因素从属于上一层因素的同时



又支配下一层因素。

(2) 构建判断矩阵。在确定各层次各因素之间的权重时，如果只是定性的结果，则很难被别人接受；因此采用一致矩阵法，将因素两两相互比较而不是放在一起比较，采用相对尺度，以尽可能减少性质不同的因素相互比较的困难，提高准确度。

(3) 层次单排序及一致性检验。所谓层次单排序是指，对于上一层某因素而言，本层次各因素的重要性排序。即对于矩阵  $A$  计算满足  $AW = \lambda_{\max} W$  的特征根与特征向量，式中  $\lambda_{\max}$  为矩阵  $A$  的最大特征根， $W$  为对应于  $\lambda_{\max}$  的正规化的特征向量， $w$  的分量  $w_i$  即是相应元素的权值。判断矩阵通常是不一致的，但是为了能用它的对应于特征根的特征向量作为被比较因素的权重量，其不一致程度应在容许的范围内，对矩阵进行一致性检验，计算一致性指标  $CI$  (Consistency Index)：

$$CR = \frac{CI}{RI} \tag{4.2}$$

式中， $CR$  为一致性比率，当  $CR < 0.1$  时认为不一致程度在允许的范围内，而

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{4.3}$$

式中， $CI$  为一致性特征数 ( $CI=0$  时  $A$  一致， $CI$  接近于零有满意的一致性， $CI$  越大，不一致性越严重)； $n$  为成对比较的因子个数。

$RI$  为随机一致性指标，其值参考表 4.2 与 4.3：

**表 4.2 判断矩阵标度定义**

标度	含义
1	表示两个因素相比，具有相同重要性
3	表示两个因素相比，前者比后者稍重要
5	表示两个因素相比，前者比后者明显重要
7	表示两个因素相比，前者比后者强烈重要
9	表示两个因素相比，前者比后者极端重要
2,4,6,8	表示上述相邻判断的中间值
倒数	若因素 $i$ 与因素 $j$ 的重要性之比为 $a_{ij}$ ,那么因素 $j$ 与因素 $i$ 重要性之比为 $a_{ji}=1/a_{ij}$

**表 4.3 平均随机一致性指标  $RI$**

阶数 $n$	1	2	3	4	5	6	7	8	9	10	11
$RI$	0	0	0.52	0.89	1.12	1.24	1.36	1.41	1.46	1.49	1.52

(4) 层次总排序及一致性检验。原理与层次单排序及一致性检验一样，利用式 (4.3) 进行总排序一致性检验：

$$CR = \frac{a_1 CI_1 + a_2 CI_2 + \cdots + a_n CI_n}{a_1 RI_1 + a_2 RI_2 + \cdots + a_n RI_n} = \frac{\sum_{i=1}^n CI_i a_i}{\sum_{i=1}^n RI_i a_i} \quad (4.4)$$

$CR < 0.1$  则可按照总排序权向量表示的结果进行决策, 否则需要重新考虑模型或重新构造那些一致性比率  $CR$  较大的成对比较矩阵。

(5) 确定  $n$  个指标的权重  $(\alpha_1, \alpha_2, \dots, \alpha_n)$ , 则:

$$Y_i = a_1 x_{i1} + a_2 x_{i2} + \cdots + a_n x_{in} \quad (4.5)$$

其中  $x_{in}$  为第  $i$  个事件的第  $n$  个指标对应的值,  $Y_i$  记为 AHP 方法计算出的综合评价分。

**表 4.4 基于 AHP 的恐怖主义风险量化指标与权重**

一级指标	权重	二级指标	权重	三级指标	权重	整体权重
恐怖袭击事件危害性	1	人员伤亡 F1	0.2408	死亡总数 F11	0.4066	0.0979
				受伤总数 F12	0.2669	0.0643
				受害者类型 F13	0.0607	0.0146
				人质或绑架受害者总数 F14	0.2658	0.0640
		财产损失 F2	0.1272	财产损失 F21	0.5639	0.0717
				财产损失程度 F22	0.4361	0.0555
		不良社会影响 F3	0.1674	恐怖袭击目的 F31	0.3131	0.0524
				对人群社会影响 F32	0.1105	0.0185
				是否超出人道主义 F33	0.3726	0.0624
				是否为持续事件 F34	0.2038	0.0341
		袭击类型及规模 F4	0.262	自杀式袭击 F41	0.2074	0.0543
				攻击类型 F42	0.2643	0.0692
				武器类型 F43	0.2745	0.0719
				成功的攻击 F44	0.2333	0.0611
				抓获的凶手数量 F45	0.0205	0.0054
		蔓延范围及恐怖集团 F5	0.2026	国家 F51	0.1234	0.0250
				地区 F52	0.3469	0.0703
				附近地区 F53	0.2619	0.0531
				第一可疑、涉嫌犯罪集团 F54	0.0593	0.0120
				声称负责 F55	0.2085	0.0422

## 2. 基于因子分析 (FA) 的综合评价

因子分析 (factor analysis) 是主成分分析的推广, 也是利用降维的思想, 由研究原始变量相关矩阵或协方差矩阵的内部依赖关系出发, 把一些具有错综复杂

关系的多个变量归结为少数几个综合因子的统计分析方法。

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i, \quad (i=1,2,\cdots,p) \quad (4.6)$$

上式中的  $F_1, F_2, \dots, F_m$  称为公共因子,  $\varepsilon_i$  为  $X_i$  的特殊因子,  $a_{ij}$  为因子载荷,

记  $X_i$  为 FA 方法计算出的综合评价分。

该模型可用矩阵表示为:

$$X = AF + \varepsilon$$

这里

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} = (A_1, A_2, \cdots, A_m), \text{称矩阵 } A \text{ 为因子载荷矩阵}$$

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

且满足:

$$(1) \quad m \leq p;$$

$$(2) \quad \text{Cov}(F, \varepsilon) = 0, \quad \text{即公共因子与特殊因子是不相关的};$$

$$(3) \quad D_F = D(F) = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \\ & & & 1 \end{bmatrix} = I_m, \quad \text{即各个公共因子不相关且方差为 1};$$

$$(4) \quad D_\varepsilon = D(\varepsilon) = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ & & \ddots \\ 0 & & & \sigma_p^2 \end{bmatrix}, \quad \text{即各个特殊因子不相关, 方差不要求相等}.$$

因子分析的逻辑框架图如下:

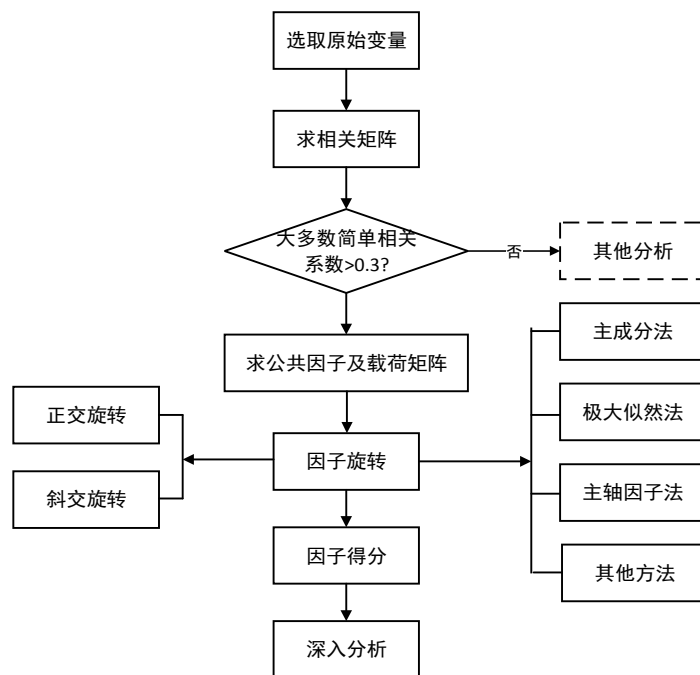


图 4.2 因子分析框架图

具体分析如下：

(1) KMO 与 Bartlett 检验

表 4.5 KMO-Bartlett 球形检验结果

Kaiser-Meyer-Olkin 测量取样适当性		0.672
Bartlett 的球形鉴定	卡方	735814.38
	df	0.351
	显著性	0

Barlett 球形检验表明：Bartlett 值为 735814，P 接近于 0，若显著性水平为 0.01，则拒绝相关矩阵为单位矩阵的原假设，相关矩阵与单位矩阵存在显著差异故考虑进行因子分析。取样足够的 Kaiser-Meyer-Olkin 检验是用于比较观测相关系数值与偏相关系数值的一个指标，其值越逼近 1，表明对这些变量进行因子分析的效果越好。从表 4.5 可见，KMO 值为 0.672，因而可以对指标变量进行因子分析。

(2) 各指标描述性统计

对样本指标提取公共因子前，先对各指标的数据分布有个基本了解，如下表所示：

表 4.6 各指标描述性统计结果

样本指标	均值	标准偏差	最小值	最大值	偏度	峰度
死亡总数	2.45	12.21	0	1570	57.93	5879.39
受伤总数	3.52	38.99	0	8191	173.03	35040.04

人质或绑架受害者总数	0.88	24.85	0	5350	126.99	22270.57
...	...	...	...	...	...	...
抓获的凶手数量	0.12	2.04	-9	406	109.05	17752.38
附近地区	0.09	0.28	0	1	2.91	6.48
第一可疑、涉嫌犯罪集团	0.07	0.31	0	1	2.6	4.76
声称负责	0.16	0.37	0	1	1.82	1.32

表 4.6 的信息反映了各指标的均值、标准差、最大值、最小值、偏度以及峰度情况。从死亡总数来看，均值为 2.45，但极差却高达 1570，说明存在重大恐怖袭击事件带来如此严重的人员死亡；同样地，受伤总数的均值为 3.52，但最大值却高至 8191 人，说明绝大部分的恐怖袭击事件带来的伤亡人数相对较少，但极端恐怖袭击事件所造成的损失却是不可估量的。再看财产损失程度，均值为 3.7，介于 3 和 4 之间，在对照表里分别对应较小的和未知的财产损失，说明大多数的恐怖袭击事件带来的损失不是特别大，但其偏度和峰度分别为-1.02 和 -1.05，呈现明显的左偏尖峰分布，在左侧有长尾说明有少数存在损失程度为 1（即灾难性的财产损失）的恐怖袭击事件发生。

而对于恐怖袭击目的、对人群社会影响和是否超出人道主义（3 个入选标准）来说，均接近于 1，说明数据集中大部分记录在案数据符合恐怖袭击标准。可以发现，符合自杀式袭击的样本较少，同时是否为持续事件指标也反映出绝大部分恐怖袭击事件表现出突发性特征，很少有持续性表现。再者对成功的攻击这一指标来说，均值接近于 1，说明大多数的恐怖袭击事件都达到其预定目的，会给社会安全稳定带来严重威胁。

据此，有理由认为我们所搭建的指标体系可以较为全面地反映出恐怖袭击事件的危害性，在通过 KMO 检验之后进一步进行提取出公共因子并求出因子得分，其方差贡献率如下表所示：

**表 4.7 公共因子方差贡献率**

元件	起始特征值			循环平方和载入		
	总计	变异的 %	累加 %	总计	变异的 %	累加 %
1	2.78	12.08	12.08	2.72	11.82	11.82
2	2.16	9.38	21.46	2.04	10.86	22.68
3	2.08	9.06	30.52	1.92	10.33	33.01
4	1.7	7.37	37.89	1.91	10.28	43.29
5	1.22	5.32	43.21	1.27	7.52	50.81
6	1.17	5.07	48.28	1.21	7.27	58.08
7	1.14	4.95	53.23	1.18	7.12	65.2
8	1.05	4.55	57.77	1.05	6.58	71.78

由表 4.7 可知，这 8 个公共因子解释的累计方差贡献率达到 71.78%，且经正交旋转后公共因子所解释的累积总方差不变，只是分配在各指标的方差贡献度有所变化，丢失信息量较少，足以代表原统计指标所含大部分信息，用它们来代替原有指标对恐怖袭击事件危害性进行测度是可行的。

其特征值碎石图如下所示：

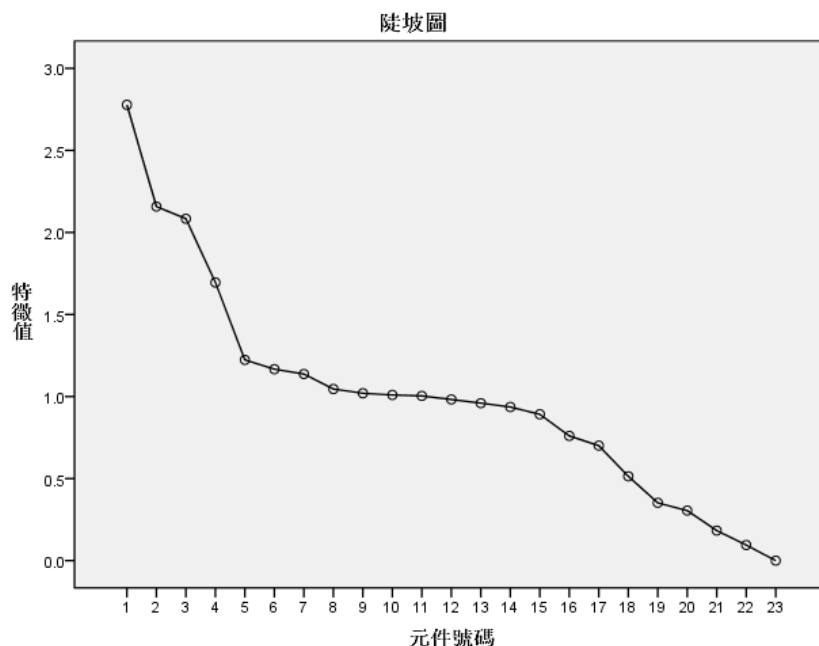


图 4.3 特征值碎石图

碎石图是表征因子个数与特征值关系的图形，横轴为因子数量，纵轴为特征值。从图 4.3 可以看出前 8 个因子特征值变化非常明显，之后趋于平缓。所以前 8 个因子对恐怖袭击事件危害性评价具有显著性作用，这与上面方差贡献度的分析一致。

### (3) 因子旋转

因子分析的目的不仅是要找出公共因子，更重要的是知道每个公共因子的具体经济意义。为便于对公共因子进行解释，一般需对因子载荷矩阵进行旋转。因子模型的表达式为：

$$X_i = a_{i1} F_1 + \dots a_{im} F_m + a_i u_i \quad (4.7)$$

其中， $F_1, \dots, F_m$  为公共因子， $a_{ij}$  ( $j=1, 2, \dots, m$ ) 为因子载荷，它是第  $i$  个变量在第  $j$  个公共因子上的负荷； $u_i$  为特殊因子， $a_i$  为特殊因子载荷。对于由因子模型矩阵得到的初始因子载荷矩阵，如果因子载荷之间相差不大，对因子的解释就不是很明确，因此要通过旋转因子坐标轴，以使每个因子载荷在新的坐标系中能按列和行向 0 和 1 两极分化。本报告选取最常用的因子载荷旋转方法，即方差最大化的正交旋转。经旋转后的原变量只在某一个因子上具有较大的载荷，而在其它因子上载荷很小，这就可实现公共因子含义的清晰化，能够直观地看出每个因子所代表的原经济变量，从而实现模型结构的简化。

表 4.8 给出了提取出的 8 个公共因子在各指标上的权重情况，如下：

表 4.8 旋转后因子载荷矩阵

公共因子								
	1	2	3	4	5	6	7	8

武器类型	0.964	0.013	-0.032	0.034	0	0.002	0.019	0.015
攻击类型	0.853	-0.006	0.056	0.013	-0.013	0.104	-0.022	-0.005
受伤人数	-0.006	0.894	0.027	-0.017	0.048	-0.031	-0.023	0.014
死亡总数	-0.027	0.709	0.013	0.035	0.188	0.362	0.081	0.025
财产损失	-0.012	0.017	0.934	-0.057	0.036	-0.042	0.005	0.002
...	...	...	...	...	...	...	...	...
地区	0.007	0.003	-0.04	0.064	0.135	0.042	0.77	0.014
国家	0.008	0.044	0.073	0.085	-0.322	-0.03	0.603	-0.107
附近地区	-0.009	-0.003	0.002	0.187	-0.067	0.03	0.035	0.594
入选标准 1	0.038	-0.031	-0.035	-0.234	0.34	-0.118	0.222	0.56
抓获的凶手数量	0.009	0.033	-0.001	-0.012	0.054	-0.048	-0.054	0.095

从上表中可以看出：

(1) 武器类型、攻击类型在因子 1 上具有较大的载荷，结合表 1 的指标体系分类，可将因子 1 视为袭击类型因子。

(2) 受伤人数、美国人质或绑架受害者人数及死亡总数在因子 2 上具有较大载荷，因子 2 可以充分反映这三个变量所包含的信息，可视为人员伤亡因子。

(3) 财产损失及财产损害程度指标在因子 3 上载荷较大，同时观察到两者系数的符号恰恰相反，这是因为我们定义的财产损失程度，数值越接近于 1 其危害性越大，可将因子 3 视为财产损失因子。

(4) 疑似恐怖主义和是否超出人道主义（标准 3）指标在因子 4 上载荷较大，可将其视为不良社会影响因子。同样地，两者的符号也呈相反状态。

(5) 声称负责和自杀式袭击指标在因子 5 上载荷较大，可将其视为恐怖集团因子。

(6) 第 6 个因子上人质或绑架受害者总数、是否为持续事件及成功的攻击所占比重较大，可视为危害发生因子。

(7) 第 7 个因子上地区和国家所占比重较大，可视为地理位置因子。

(8) 第 8 个因子所包含指标最多，有附近地区、恐怖袭击目的（入选标准 1）、第一可疑或涉嫌犯罪集团以及抓获的凶手数量，定义为蔓延范围因子。

这样，前面提取的 8 个公共因子就有了具体的含义，并且所包含的信息量足以替代原有指标，实现了评价体系的简化。

**表 4.9 因子得分协方差矩阵**

因子	1	2	3	4	5	6	7	8
1	1	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0
4	0	0	0	1	0	0	0	0
5	0	0	0	0	1	0	0	0
6	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0	1	0
8	0	0	0	0	0	0	0	1

为了使提取的因子具有不同的评价维度，还要使各因子之间不存在相关关系，从而避免因子得分综合评价的多重共线性。表 4.9 给出的一种得分协方差矩阵表明该矩阵等价于单位矩阵，8 个因子之间不相关，可以代表不同的评价维度。

### (3) 计算因子得分

在 SPSS 中，利用回归法可以得出各因子得分。

取各因子的方差贡献度为因子得分权重，可得出各上市公司的综合因子得分，各因子系数得分如表 4.10 所示。综合得分计算方法为：

$$W_i = (0.1182 \times F_1 + 0.1086 \times F_2 + 0.1033 \times F_3 + 0.1028 \times F_4 + 0.0752 \times F_5 + 0.0727 \times F_6 + 0.0712 \times F_7 + 0.0658 \times F_8) / 0.7178 \quad (4.8)$$

表 4.10 因子得分系数矩阵

	因子							
	1	2	3	4	5	6	7	8
是否为持续事件	0.091	-0.055	-0.07	-0.036	-0.055	0.417	-0.037	-0.02
附近地区	-0.015	-0.002	0.014	0.123	-0.054	0.023	0.026	0.571
入选标准 1	0.026	-0.029	-0.03	-0.132	0.289	-0.115	0.194	0.531
...	...	...	...	...	...	...	...	...
国家	0.002	0.042	0.052	0.037	-0.272	-0.023	0.515	-0.101
地区	0	-0.012	-0.009	0.002	0.097	0.036	0.652	0.018
攻击类型	0.313	-0.005	0.043	-0.002	-0.001	0.025	-0.017	-0.013
武器类型	0.361	0.013	-0.001	-0.001	0.011	-0.071	0.015	0.007

### 3. 基于算术平均方法确定权重的组合 AHP-FA 综合评价模型

Step1 根据上述两个模型，可得到两组综合评分  $X_i$ ， $Y_i$ ；

Step2 根据算术平均方法确定两组综合评分的权重，即令：

$$l_i = \frac{1}{m}, \quad i = 1, 2, \dots, m \quad (4.9)$$

显然

$$\sum_{i=1}^m l_i = 1, \quad l_i \geq 0, \quad i = 1, 2, \dots, m$$

Step3 恐怖袭击事件危害程度  $Z_i$  为：

$$Z_i = l_1 X_i + l_2 Y_i, \quad i = 1, 2, \dots, m \quad (4.10)$$

根据（4.8）式的因子分析计算出的综合得分进行评估分析可以得出 1998-2017 近 20 年来十大恐怖事件的结果，如表 4.11 所示：



表 4.11 近 20 年来十大恐怖事件

GTD 标志号	综合得分
200109110004	30.52
200109110005	27.45
201408090071	17.55
201406150063	14.52
199808070002	7.97
200109110006	7.62
201406100042	6.39
200409010002	6.33
201710140002	4.83
200403210001	4.83

#### 4.4.2 模糊 C 均值 (Fuzzy C-means) 聚类<sup>[9]</sup>

通过对附件 1 的数据进行分析研究,可以采用模糊 C 均值 (Fuzzy C-means, 简称为 FCM) 聚类方法。FCM 聚类是一种比较典型的模糊聚类算法,在 1981 年被 Bezdek 提出,主要用于将多维数据空间分布的数据点分成特定数目的类。在模糊聚类中,每一个数据点以某种程度属于第一类,他用隶属度来表示每个数据点属于每个聚类的程度。运用 FCM 聚类方法对恐怖袭击事件进行聚类分析,将聚类结果运用到依据危害性对恐怖事件进行分级的问题中,能够有效解决分级问题。对于 FCM 聚类的具体描述如下:

设  $n$  个样本的数据集  $X=\{X_1, X_2, \dots, X_n\}$ , 将样本分为  $C$  类,  $1 < C < n$ , 对任意  $X_i \in X$ , 其特征矢量  $X_i = \{x_{i1}, x_{i2}, \dots, x_{ip}\}$ ,  $x_{ip}$  表示第  $i$  个样本的第  $P$  个特征属性。用隶属度函数定义的目标函数可以写为:

$$J(U, V) = \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 \quad (4.11)$$

约束条件为:

$$\begin{cases} 0 \leq u_{ij} \leq 1 \\ \sum_{i=1}^c u_{ij} = 1 \end{cases} \quad (4.12)$$

在式 (4.12) 中  $u_{ij}$  为第  $i$  个样本对于第  $j$  类的隶属度函数  $1 \leq i \leq n$ ,  $1 \leq j \leq C$ ;

$U = \{u_{ij}\}$  为  $c \times n$  阶模糊分类矩阵;  $V = [v_1, v_2, \dots, v_c]$  为  $p \times c$  阶聚类中心矩阵;  $v_k$  为每个聚类的中心特征向量;  $m$  为模糊加权指数; 样本到聚类中心的欧氏距离为:

$$(d_{ij})^2 = \|x_j - v_i\|^2 = (x_j - v_i)^T A (x_j - v_i) \quad (4.13)$$

$J(U, V)$  表示各类样本到聚类中心的加权值, 根据聚类准则求得:

$$\min \{J(U, V)\} = \min \left\{ \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 \right\} \quad (4.14)$$

采用拉格朗日乘数法，构造式：

$$F = \sum_{i=1}^c (u_{ij})^m (d_{ij})^2 + \lambda (\sum_{i=1}^c u_{ij} - 1) \quad (4.15)$$

则，优化条件为：

$$\begin{cases} \frac{\partial F}{\partial \lambda} = (\sum_{i=1}^c u_{ij} - 1) = 0 \\ \frac{\partial F}{\partial u_{ij}} = [m(u_{ij})^{m-1} (d_{ij})^2 - \lambda] \end{cases} \quad (4.16)$$

由式 (4.12) 可得：

当  $I_k = \emptyset$  时：

$$u_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij} / d_{kj})^{\frac{2}{m-1}}} \quad (4.17)$$

$u_{ik} = 0$  ,  $\forall i \in \bar{I}_k$  且满足  $\sum_{i \in I} u_{ik} = 1$  ,  $I_k \neq \emptyset$  时：

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x^k}{\sum_{k=1}^n (u_{ik})^m} \quad (4.18)$$

在 (4.12) 式中， $I_k = \{1 \leq i \leq c; d_{ik} = 0\}$  ,  $\bar{I}_k = \{1, 2, \dots, c\} - I_k$ 。

根据以上步骤对恐怖事件危害性进行 FCM 聚类分析，聚类图如图 4.5 所示，根据图中信息反映，可大致分为五个类别，分别用不同的颜色表示出来，如下：

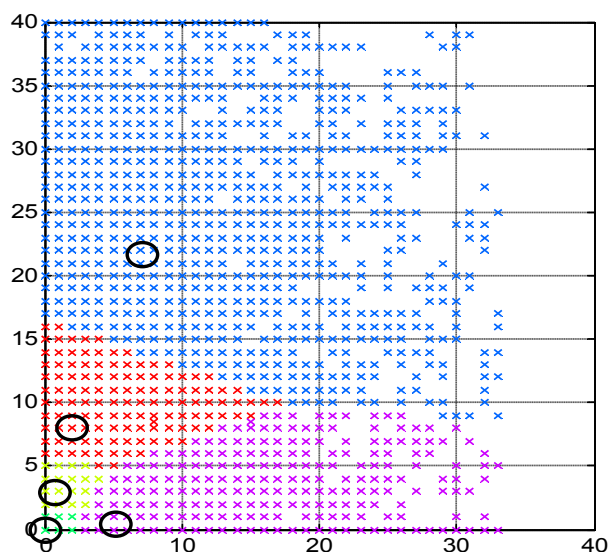


图 4.5 FCM 聚类分析图

再者，根据综合评价结果进行分级，结果如下表所示：

表 4.12 危害级别结果分析图

危害级别	恐怖袭击事件数量	聚类分析结果
1 级	654	第一类
2 级	3874	第五类
3 级	39807	第二类
4 级	4118	第四类
5 级	65730	第三类

表 4.12 中依据危害性给出了对恐怖事件分级的结果。我们发现，危害级别为一级的恐怖事件主要表现特征为受伤人数多、受伤人数也很多、有一定的财产损失，因此危害程度评为 1 级；危害级别为二级的恐怖袭击事件主要特征为死亡人数多，受伤人数相对较小，有一定的财产损失，因此，危害级别为 2 级；危害级别为三级的事件死亡人数中等、受伤人数很多，有一定的财产损失；危害级别为 4 级的恐怖袭击事件死亡人数中等、受伤人数中等；危害级别为 5 级的恐怖袭击事件有较少的死亡人数，受伤人数基本为 0，财产损失也较少。对于恐怖事件分为 5 级，各级事件占比情况如图 4.5 所示：

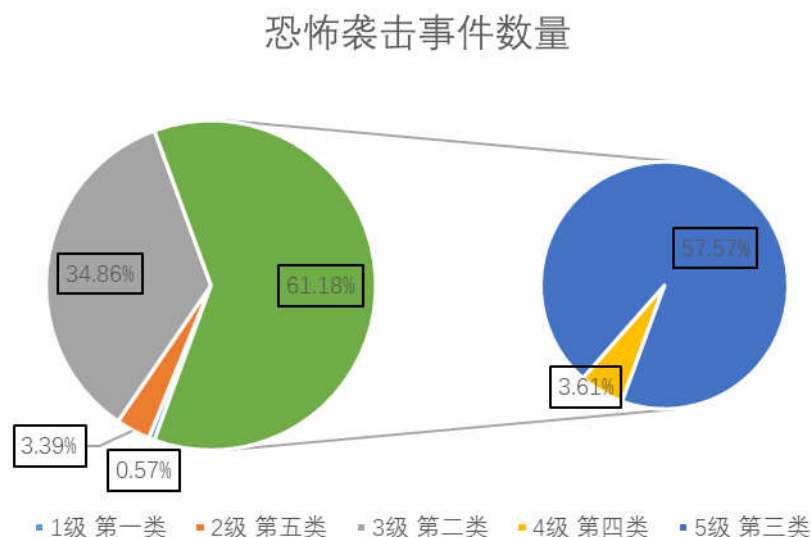


图 4.5 各级别恐怖事件数量

据图 4.5 可以看出，恐怖事件发生的级别在 3、4、5 两级占比达到了 96.04%，其中，第 3 级占比数为 34.86%，此级别的恐怖袭击事件对社会造成的危害程度相当显著。

综合表 4.12 与图 4.5 来看，并结合因子分析的结果分析对本题表一中的恐怖袭击事件的危害性进行分级，危害性级别情况情况如表 4.13 所示：

表 4.13 典型事件危害级别

GTD 标志号	危害级别
200108110012	一级
200511180002	三级
200901170021	一级
201402110015	五级
201405010071	四级
201411070002	五级
201412160041	二级
201508010015	五级
201705080012	三级
200108110012	一级

## 5 问题二模型建立与求解

### 5.1 问题分析

这一题的诉求点是依据事件特征发现恐怖袭击事件制造者，逻辑主线很清晰，要想发现未宣告负责的恐怖袭击事件的制造者，首先要对各类型的恐怖袭击事件的特征作出判断。题目要求针对在 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，建立数学建模方法寻找上述可能性，实质上是对具有同类型特征的恐怖袭击事件的匹配和判别过程。

### 5.2 恐怖袭击事件特征分析

附件 1 中给出了近几十年的恐怖袭击事件各项数据信息，那么是否要利用如此之多的信息呢？答案当然是否定的，随着社会的发展，恐怖袭击事件的手段、地点、人群、恐怖组织等也在随之改变。故我们选取了 2006—2016 年的恐怖主义袭击事件的相关指标进行分析。

首先要明确的是，题目要问的关键在于找出未宣告负责恐怖袭击事件的制造者，那么对于幕后的事件制造者来说，其相关指标信息是无法从正面获取的，既然如此，我们可以利用已知的宣告负责的事件制造者的信息，获取一定得相关系数去与已发生的但制造者未知的恐怖主义事件相匹配，同时结合逆向思维利用缺失信息计算相离系数来推测未知的恐怖袭击事件制造者行动的可能性大小。因而，首先我们要对已知恐怖袭击事件的特征作出分析，如下。

#### 5.2.1 恐怖袭击事件的区位特征

由于我们想要获取的恐怖袭击事件制造者并未宣称负责，故无法知晓其属于哪一组织，但恐怖事件发生的地理位置在每个恐怖袭击事件发生后准确无误的有效信息，故我们分别筛选出这十年中宣称负责的和未宣称负责的恐怖袭击事件发生的地区，记录其频数，如表 5.1 所示：

**表 5.1 恐怖袭击事件区位分布频数**

序号	地区	宣称负责	未宣称负责
1	北美	50	87
2	中美洲和加勒比海	0	4
3	南美	67	268
4	东亚	13	23
5	东南亚	241	1909
6	南亚	2185	6042
7	中亚	4	23
8	西欧	113	493
9	东欧	63	754
10	中东和北非	2328	9822
11	撒哈拉以南的非洲	745	3298
12	澳大利亚和大洋洲	1	23

**数据来源：由附件 1 数据集计算得出**

由表 5.1 可以看出，无论是宣称负责还是未宣称负责的恐怖袭击事件制造者，就袭击事件发生的频数来说，中东和北非以及南亚无疑是恐怖主义活动最为活跃的爆发区，几近占据了近年来恐怖袭击事件次数的大半；而对于中美洲和加勒比海、东亚、中亚、澳大利亚和大洋洲这些地区而言，其恐怖主义活动极为少见；但与亚洲接壤的非洲，却是匿名恐怖袭击事件制造者的“天堂”，该地区发生的多起孔壁袭击事件均以无人宣称负责而收尾。

为了更直观地展示宣称负责的组织与未宣称负责的组织的情况，给出环状对

比图如图 5.1 所示：

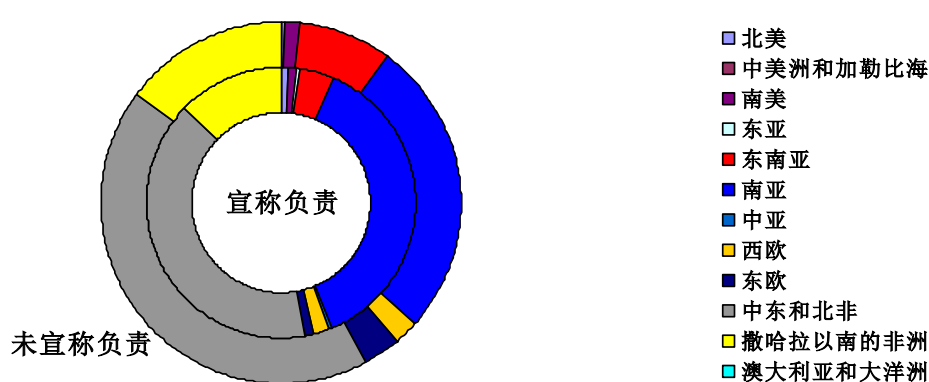


图 5.1 恐怖袭击事件宣称负责与否对比环状图

图 5.1 中，中心的圆环表示宣称负责的恐怖主义袭击事件发生的频数，外侧的圆环代表的是未宣称负责的恐怖主义袭击事件发生的频数。不难看出，灰色、蓝色和黄色区域分别对应中东和北非、南亚、撒哈拉以南的非洲，其中又以中东和北非恐怖袭击事件最为高发；而美洲和欧洲则是恐怖袭击事件鲜有发生的区域。

从整体上看，绝大部分哪个地区，其未宣称负责的幕后恐怖袭击事件制造者都要比宣称负责的恐怖组织多。这说明虽然区位因素是挖掘恐怖主义袭击事件幕后黑手的一把钥匙但其仅仅是打开了无数大门的第一扇门，还需要对更多的信息进行探索。

5.2.2 恐怖袭击事件的受害者特征

既然仅凭区位因素无法获知谁才是幕后的事件制造者，那么，我们自然会提出第二个疑问：“事件的目标受害者是谁？”

每一起恐怖主义袭击事件的背后都有其原因，有基于政治、经济、宗教抑或是社会因素的，意识形态的不同是引发恐怖主义袭击事件的根源。而对于这样具有特定目的的事件来说，其目标受害人也一般是具有固定定位的人群。据此，我们对样本期内目标受害者的类型作出了统计分析，其数据分布如表 5.2 所示：

表 5.2 恐怖袭击事件目标受害者分类情况

受害者类型	宣称负责	未宣称负责	受害者类型	宣称负责	未宣称负责
商业	376	1685	非政府组织	14	94
政府（一般）	450	1725	其他	9	24
警察	1168	2654	公民自身和财产	1087	7470
军事	1679	3889	宗教人物或机构	157	544
流产有关	1	5	电信	26	66
机场和飞机	24	39	恐怖分子	182	455
政府外交	74	170	游客	6	7
教育机构	59	447	运输	111	506

食物或水供应	3	25	未知	129	1894
新闻记者	53	267	公用事业	152	454
海事	10	29	暴力政党	40	297

数据来源：由附件 1 数据集计算得出

表 5.2 给出的是宣称负责的可以看出，恰如我们所预想的一样，在众多的恐怖主义袭击事件中，往往基于特定的目的而导致了特定的受害人群。其中，警察、军事属于恐怖袭击事件受害频率最高的职业，这两个指标包括对警察部队或警务设施的攻击以及攻击部队、巡逻队、营房、车队，吉普车和飞机的行为，因其需要战斗在反恐斗争的第一线上；而公民自身和私有财产这一指标紧随其后，该指标反映的恐怖袭击事件包括对个人，一般的公众的攻击，及对公共领域，包括市场、商业街、繁忙的十字路口和步行街的攻击，说明有相当一部分的恐怖袭击事件发生在人群密集的地方，反映了恐怖主义分子企业引发社会秩序混乱的险恶用意。

对于其他的诸多行业来说，恐怖主义者都有可能将其作为攻击目标。结合本题的内容来看，在已发生的恐怖袭击事件中，未宣称负责的恐怖主义事件制造者归属于这些行业的可能性很大程度上比警察、军事、公民自身和私有财产这种指标的概率要小得多。

先简单地对宣称负责的恐怖袭击事件次数做面积图可以很直观地得出上述结论，如图 5.2 所示：

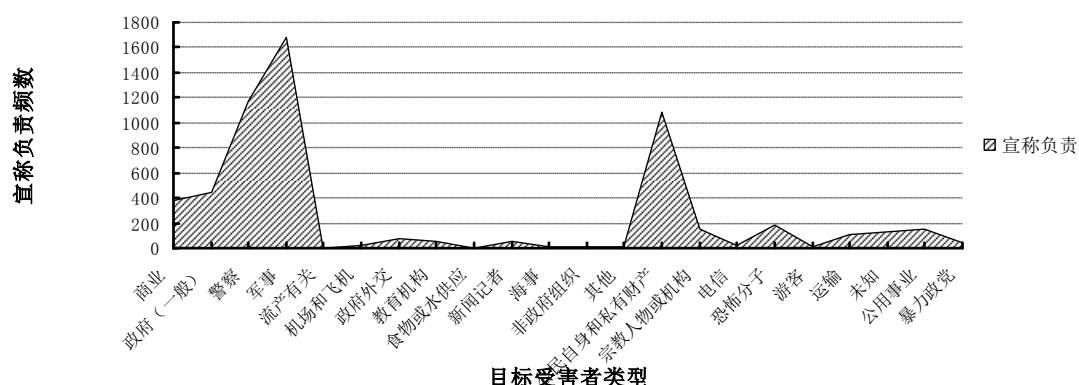


图 5.2 恐怖袭击事件目标受害者类型分布

结合图 5.2，从概率的角度出发，我们可以判断出，在对恐怖袭击制造者未知但袭击事件及受害者类型明确的情况下，我们很有理由相信最可疑的恐怖袭击事件制造者应归属于常见的攻击受害者这一类。

同时可以发现，在诸多的恐怖袭击事件中，个别行业指标占据了大部分的频数，有很多行业显得重要性很低。故我们从全面考虑的角度，试图将各个有用信息都包含近来，将频数较少的行业信息统一归类为“其他”这一分类中，如图 5.3 所示：

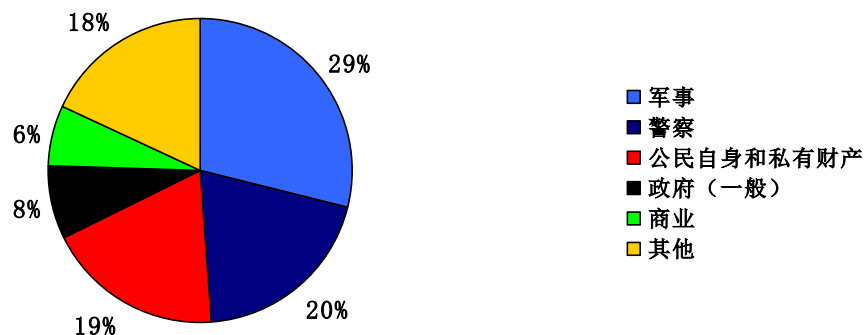


图 5.3 恐怖袭击事件受害人一览图

最后，我们给出在统计样本区间内所有已宣称负责的与未宣称负责的恐怖袭击事件次数对比的折线图，如图 5.4 所示：

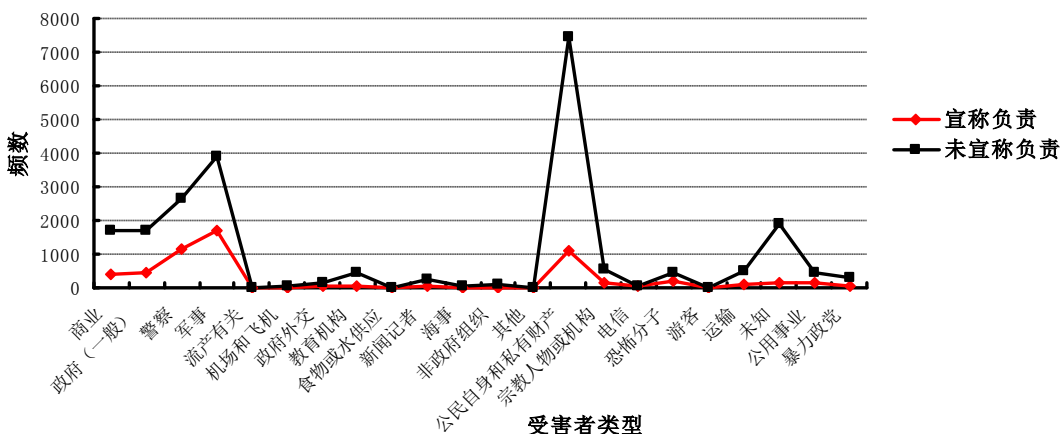


图 5.4 已宣称与未宣称负责恐怖袭击事件次数对比

由上图可以很清晰地看出，无论哪个行业范畴，其未知事件制造者的恐怖袭击频数都有高于已知事件制造者的频数，因此，虽然可以借助目标受害者的信息加以判断，但却不能得出最为精确的判断，这也就是本题希望我们建模模型加以精准判断的原因。

### 5.2.3 恐怖袭击事件袭击方式特征

以上分别分析了恐怖袭击事件发生的区域和受害者信息，但却并不能精准地定位出袭击事件制造者的身份，这时候我们考虑从袭击事件制造者的角度是搜寻蛛丝马迹。

既然袭击事件制造者并未发出声明，那应该如何根据已有信息去判断其身份？我们首先想到的是恐怖主义分子惯用的手段，也就是其一贯的袭击方式。表 4.3 给出了对恐怖主义事件袭击方式的统计。



表 5.3 恐怖袭击的袭击方式情况统计

袭击方式	宣称负责	未宣称负责
暗杀	371	1375
武装袭击	1350	4752
轰炸或爆炸	2916	11855
劫持	16	65
劫持人质（路障）	66	77
劫持人质（绑架）	446	1885
基础设施攻击	199	1202
徒手攻击	17	149
未知	427	1386

数据来源：由附件 1 数据集计算得出

很明显可以看出，已发生的恐怖袭击事件中，武装袭击与轰炸或爆炸是最常见的方式，那么我们有理由把已知恐怖袭击事件和袭击方式后关联的事件制造者信息提取出来，与我们所要判断的未宣称负责事件进行匹配，从而进一步判断出其身份信息。

给出更为明了的图形化展示，如图 5.5 所示：

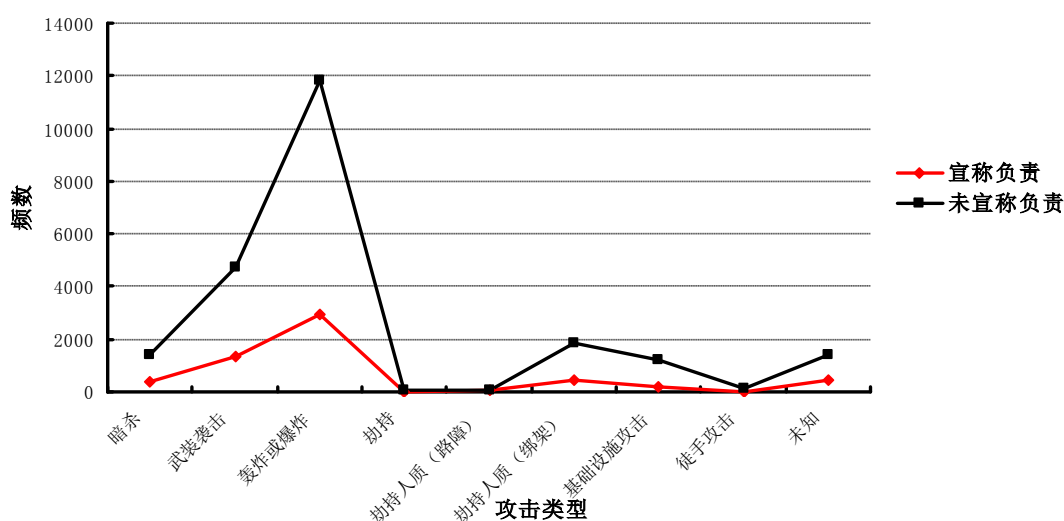


图 5.5 恐怖袭击事件袭击方式特征分析

根据图 5.4 可以看出，在已发生的恐怖袭击事件中，武装袭击与轰炸或爆炸这两个方式是最为常见的，处在一个很高的频率；基础设施攻击和劫持人质事件时有发生，但采用路障劫持人质的案件却很少。这样的信息没用么？当然不是，若是问题中需要匹配的那项恰好有在袭击方式的指标是路障劫持，那么当机立断筛选出少数几个经常性采取劫持人质（路障）方式的恐怖组织，这便是逆向思维在该问题中的运用。

#### 5.2.4 恐怖袭击事件武器类型特征

既然考虑到恐怖袭击事件袭击方式的特点，那么恐怖分子进行恐怖主义活动时的帮凶也就是其使用武器，从某种意义上说也可以用来推断问题所求解的嫌

疑者的身份。因为大多数的恐怖主义分子以团体方式作案，其使用武器具有固定的黑市来源，故某一恐怖集团可能多次作案虽然并未宣称其身份，但可根据其常用武器追溯到袭击事件始作俑者。

由此我们统计了各类型武器在已发生袭击事件中出现的频数，如表 4.4 所示：

表 5.4 恐怖袭击事件武器类型特征

对应编号	武器类型	宣称负责	未宣称负责
2	化学武器	7	45
5	轻武器	1587	5848
6	爆炸物	3312	12672
8	燃烧武器	178	1151
9	致乱武器	121	601
10	交通工具	8	39
11	破坏设备	—	18
12	其他	4	14
13	未知	593	2354

数据来源：由附件 1 数据集计算得出

观察表 5.4 可以看到，其对应编号中间有空缺的部分，其中，1、3、4、7 分别对应生物武器、放射性武器、核武器、假武器，前三类武器目前条件下尚未有恐怖集团公开使用，故不做展示，而假武器无法导致真正的恐怖袭击事件发生，也不做统计。

绘制面积图如图 5.6 所示，方便观察：

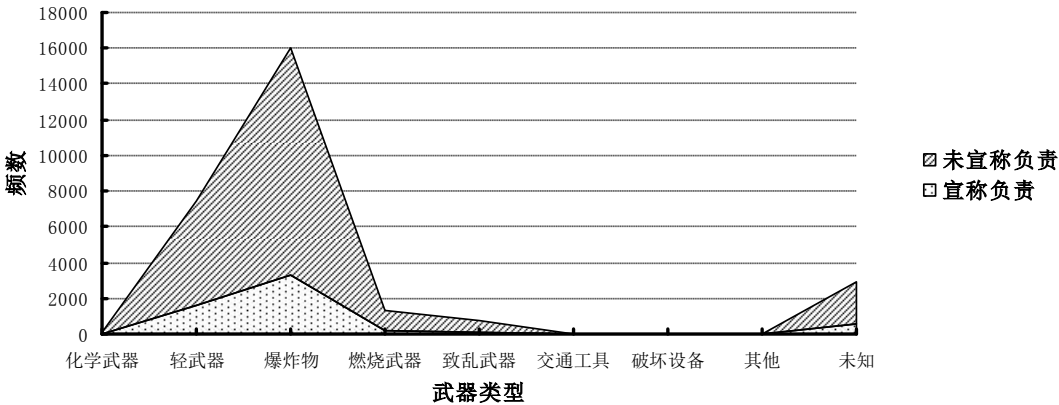


图 5.6 袭击事件武器类型特征分析

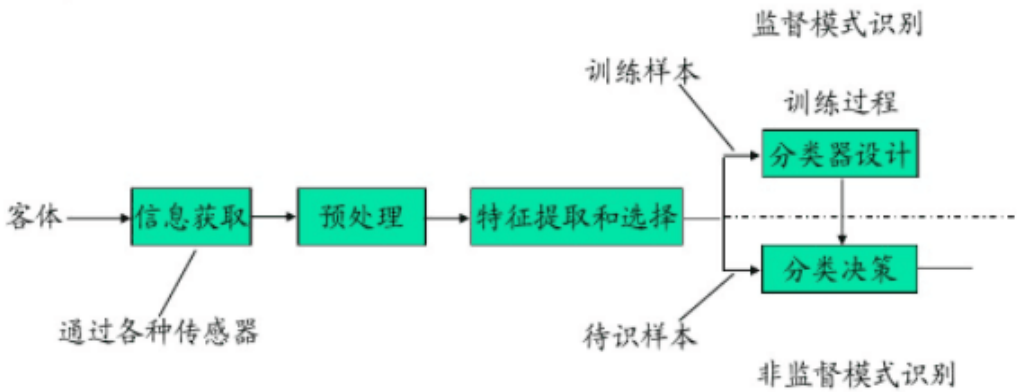
据图 5.6 很明显可以看出，在未宣称负责的恐怖袭击事件中，使用爆炸物或者轻武器的事件数最多，可以第一时间搜集经常使用某规格的轻武器或爆炸物的恐怖主义组织，从而整合信息对本题所要求解问题进行匹配分析。

### 5.3 模型的建立和求解

#### 5.3.1 基于模糊识别理论的 K-Means 算法模型<sup>[10]</sup>

1. 模式识别的理论知识

根据近 20 年来已被声称负责并且已经知道犯罪集团名称的数据来提取和选择特征，模式识别系统的基本构成如图 5.7 所示：



5.7 模式识别系统的基本构成

通过模式识别系统处理获取数据，对 15-16 年未被声称负责的恐怖事件进行处理具有一定的现实意义。模式识别系统的具体运行过程如图 5.8 所示：



图 5.8 模式识别系统

说谓模式就是对样本所具有的特征进行描述，特征向量的每一个元素即为一个特征。在模式识别过程中，模式识别就是研究在特征空间中如何将样本点进行

分类，我们发现，选择特征数量不同，出现的结果也是不同的，如图 5.9 所示：

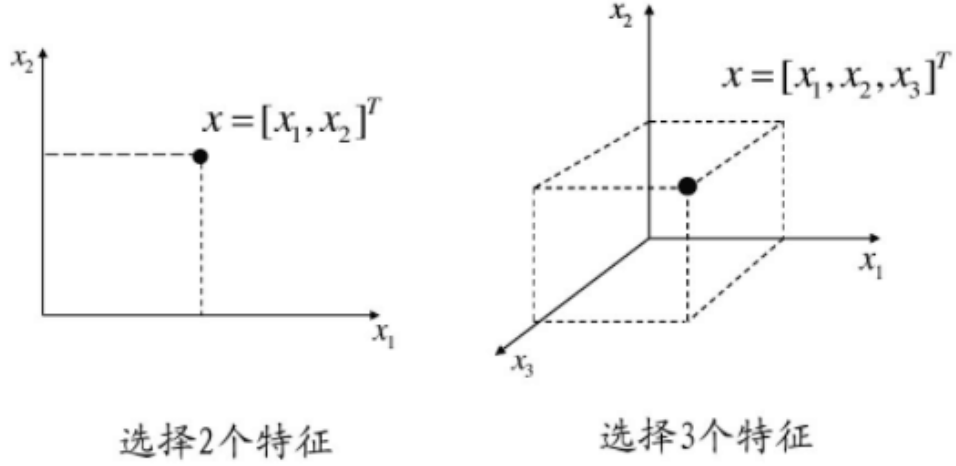


图 5.9 模式的特征向量表示

## 2. 信息熵<sup>[11]</sup>确定指标体系

在本模型中，特征提取环节主要提取的特征需要能够对模型有相关性，考虑到任务二的特殊性，我们使用信息熵确定权重的方法来确定相关指标，信息熵确定权重的步骤如下：

**Step1:** 将各个指标进行标准化处理，对于  $X_1, X_2, \dots, X_k$  这几个给定的指标，其中  $X_i = \{x_1, x_2, \dots, x_n\}$ ，假设对各指标数据标准化之后的值为  $Y_1, Y_2, \dots, Y_k$  则有：

$$Y_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (5.1)$$

**Step2:** 根据信息熵的定义，一组数据的信息熵

$$E_j = -\ln(n)^{-1} \sum_{i=1}^n p_{ij} \ln p_{ij} \quad (5.2)$$

$$\text{其中, } p_{ij} = \frac{Y_{ij}}{\sum_{i=1}^n Y_{ij}} \quad (5.3)$$

如果  $p_{ij} = 0$ ，则定义  $\lim_{p_{ij} \rightarrow 0} p_{ij} \ln p_{ij} = 0$ 。

**Step3:** 根据信息熵的计算公式，计算出各个指标的信息熵为  $E_1, E_2, \dots, E_k$ 。

通过信息熵计算各指标的权重：  $w_i = \frac{1 - E_i}{k - \sum E_i}$ 。

根据以上步骤，运用 Matlab 软件运行程序，计算得出各个指标的权重，权

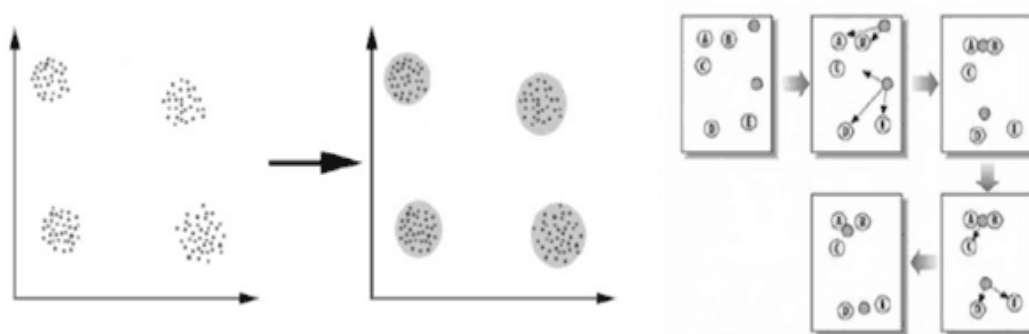
重为前 8 的指标如表 5.5 所示，因此我们主要选取这 8 个指标来描述恐怖事件各个属性。

**表 5.5 前 8 项指标权重表**

指标体系	字母表示	权重
死亡总数	A1	0.0979
财产损失	A2	0.0719
武器类型	A3	0.0717
地区	A4	0.0703
受伤总数	A5	0.0692
攻击类型	A6	0.0643
是否超出人道主义	A7	0.0624
财产损失程度	A8	0.0555

### 3. K-Means (K 均值算法)

K-Means 聚类算法就是基于距离的聚类算法，主要是通过不断的取离种子点最近均值的算法，如图 5.10 所示为两个中心点的 K 均值聚类算法。



**图 5.10 两个中心点的 K 均值算法**

### 4. K 均值算法

在恐怖袭击事件中，将模式识别中经过 K-Mean 聚类算法得出的结果进行距离测度，在任务 2 中，我们采用 K-Means( K 均值 )算法，具体算法思想如下：

假设有  $n$  个点，需要聚成  $k$  个簇。选择  $k$  个点作为簇的中心点，然后算法进行多次迭代处理并调整中心点位置，直到达到最大迭代次数，或者中心收敛于固定的点不在移动。每次迭代分为两个步骤。第一步，找到距离中心最近的点，并将这些数据点赋给相应的簇。第二步，试着用各簇中所有点的坐标均值更新中心位置。算法过程如下：

**Step1:** 从已筛选过后的近 20 年的已知犯罪集团并已被声称负责的恐怖案件中选取适当个数的犯罪集团作为中心点。

**Step2:** 对剩余的事件测量其到每个中心点的距离，并把它归到最近的中心点的簇。

**Step3:** 重新计算已经得到的各个簇的中心点。

**Step4:** 迭代 Step2 与 Step3 直到新的中心点与原中心点相等或小于指定阈值。

K-Means( K 均值 )算法的工作流程图如图 5.11 所示:

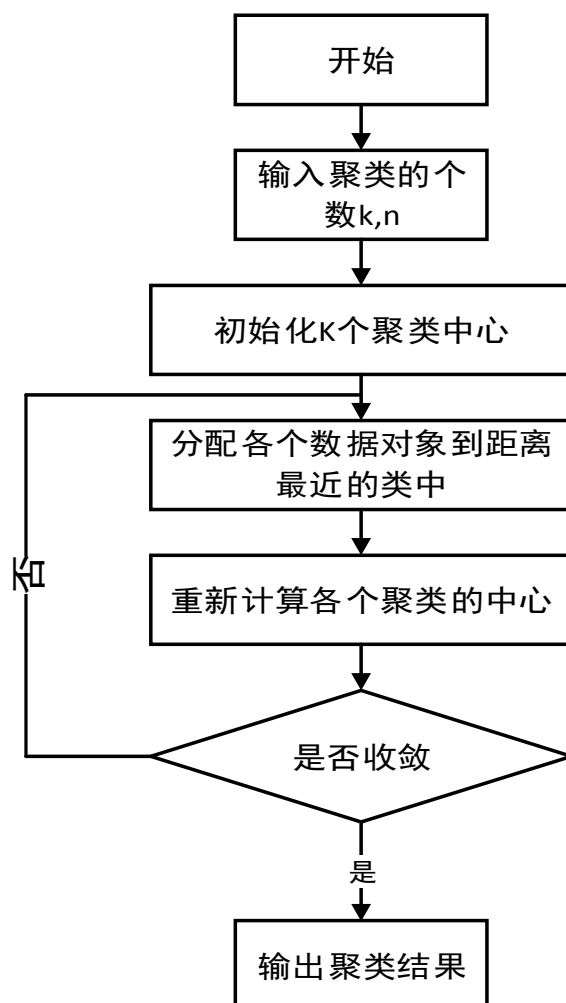


图 5.11 K 均值算法流程图

### 5.3.2 欧式距离测度<sup>[8]</sup>

根据 K-均值聚类结果，用欧式距离测度方法（Euclidean Distance）测度方法，欧式距离测度是所有距离测度中最简单最直观的，符合我们通常情况下对距离的理解。

假定给定的数据集  $X = \{x_m | m = 1, 2, \dots, total\}$ ，X 中样本用 d 个描述属性  $A_1, A_2, \dots, A_d$  来表示。数据样本  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$ ， $X_j = (X_{j1}, X_{j2}, \dots, X_{jd})$  分别是样本  $X_i$  和  $X_j$  对应 d 个描述属性  $A_1, A_2, \dots, A_d$  具体取值。样本  $X_i$  和  $X_j$  之间的相似度通常用两者之间的距离  $d(X_i, X_j)$  来表示，距离越小说明两者相似度越

高，差异性越小。欧式距离的公式如下：

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (5.4)$$

对于本题，若两个恐怖事件的属性对应的取值用欧式距离计算，对于 A、B 两个样本，每个样本对应有 8 个描述属性的具体取值，即  $A = (A_1, A_2, \dots, A_8)$  和  $B = (B_1, B_2, \dots, B_8)$ ，则两者之间的距离为：

$$d = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2 + \dots + (A_8 - B_8)^2} \quad (5.5)$$

根据距离测度的结果，并结合特征向量的均值或者属性，由 1998-2017 年已被声称负责的恐怖事件中通过模式识别针对 2015 年度与 2016 年度发生的、尚未被声称负责的数据中识别出危害性最大的 5 个嫌疑人或组织如表 5.6 所示。其中，5 号嫌疑人代表危害性较低的犯罪组织，1 号嫌疑人代表危害性最高的犯罪嫌疑人。

**表 5.6 欧式距离测度结果**

Top-5 teams	欧式距离测度	嫌疑人
Al-Qaida in Iraq	3.621593281643159	5 号
Al-Shabaab	3.065248021430621	4 号
Tehrik-i-Taliban Pakistan (TTP)	2.462114958328325	3 号
Taliban	1.129856270154337	2 号
Islamic State of Iraq and the Levant (ISIL)	1.026942015261932	1 号

### 5.3 结果分析

在使用欧氏距离进行距离测度之后，我们可以根据测度的数值所得出的危害性大小的前五位嫌疑人。但是，在对任务二中给出的表 2 里面的恐怖袭击事件填写嫌疑人信息时，需要考虑攻击类型与死亡人数的关系，并且，恐怖集团的攻击类型是具有一定的相似性的，对攻击类型的分析能够帮助判断嫌疑人。可以很明显的在图 5.12 中看出恐怖分子实施的轰炸类型的面积最大，占比最大。攻击类型与死亡人数的关系如图 5.12 所示：



图 5.12 攻击类型与死亡人数

此外，在对恐怖袭击事件进行分析时，可以看出，在总共 114183 起恐怖袭击案件中，在城市本身发生的概率达到了 91.19%，在非城市本身发生的概率只有不到 9% 如图所示。因此，在对未被声称负责的案件进行分析时，要考虑地区与附近地区对犯罪组织的影响。

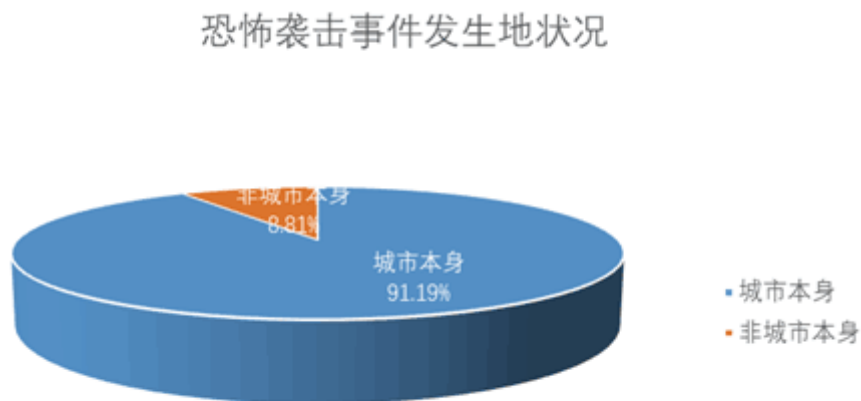


图 5.13 恐怖事件发生地状况

**结果：**根据表 5.13 欧式距离测度的结果可以得出五个犯罪集团的危害性大小，具体数字如表 5.7 所示：



**表 5.7 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	1	—	2	4
201702210037	4	3	—	—	2
201703120023	—	—	2	5	1
201705050009	—	—	2	4	1
201705050010	—	—	2	5	4
201707010028	—	2	4	1	—
201707020006	5	1		3	2
201708110018	—	3	4	1	—
201711010006	—	4	3	—	2
201712010003	5	1	—	—	4

表 5.7 代表了恐怖分子关于 2017 年 10 起典型事件的嫌疑度,我们利用 K 均值聚类算法和欧式距离测度结合各类嫌疑的属性特征得出了典型事件的嫌疑人。

## 6 问题三模型建立与求解

### 6.1 问题分析

该问题指出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,试图结合网上相关信息,建立适当的模型对近三年来发生的恐怖袭击事件进行研究,同时分析其主要原因、时空特性、蔓延特性、级别分布等规律,进而研判下一年全球或某些重点地区的反恐态势。我们拟从近三年恐怖袭击事件各项指标数据中的信息出发,从已有数据中提炼出恐怖袭击事件的潜在特征与规律,寻找出对未来反恐态势影响最大的一个或几个综合指标,建立科学的未来反恐态势分析评估模型。

### 6.2 恐怖袭击事件态势分析<sup>[4]</sup>

要对未来反恐态势进行合理的评估分析,首先要对近几年发生的恐怖袭击事件的成因、危害性、时空特性、蔓延性等有个基本的了解,故我们先利用附件 1 中给出的数据提取出有用的信息。

#### 6.2.1 袭击事件发生原因分析

参考相关资料,我们发现恐怖袭击事件发生的原因有很多,无论是基于意识形态、领土斗争、资源掠夺最终都可以归结为政治、经济、宗教和社会因素这几类。在附件 2 给出的指标体系中,从恐怖袭击事件动机的角度给出如下三个区分依据:

- (1) Crit1: 政治、经济、宗教或社会目标;
- (2) Crit2: 意图胁迫、恐吓或煽动更多的群众;
- (3) Crit3: 超出国际人道主义法律范畴。

我们统计了在 2015—2017 年各年发生的恐怖袭击事件中分别符合上述区别依据的频数,并绘出如图 6.1 所示柱状图:

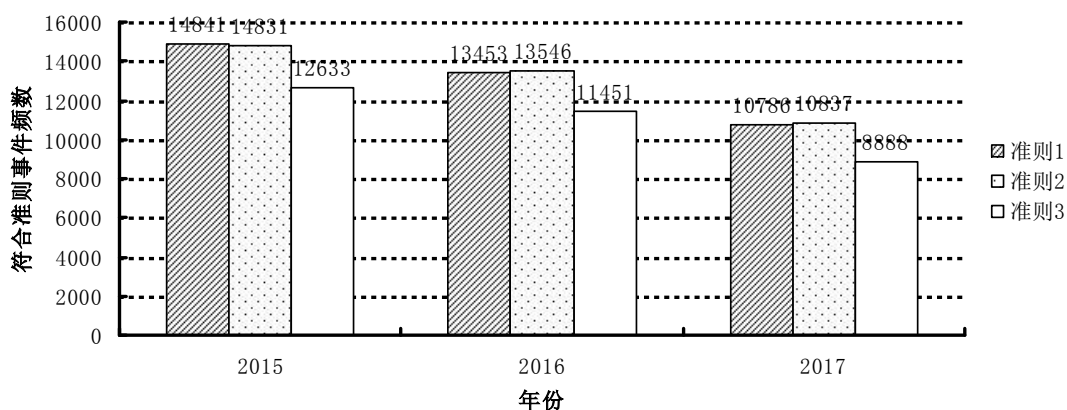


图 6.1 2015—2017 年恐怖袭击事件成因对比

据图 6.1 反映，无论是 2015、2016 还是 2017 年，三类引发恐怖袭击事件的动机数量均无明显的差异。其中第一类的政治、经济、宗教或社会因素与第二类的意图胁迫、恐吓或煽动更多群众的统计数量基本一致，但第三类超出人道主义法律范畴的事件数量均比其余两类低 2000 起左右。

再从时间维度来看，从 2015 至 2017 年，无论成因如何，其整体袭击事件发生数量都呈现出逐年递减的趋势，结合相关资料可大胆猜测这种现象可能与国际反恐力量投入有关，以期在下面建立的模型中得以证实。

抛开时间维度，仅从三年累计发生恐怖袭击事件来看（如图 6.2 所示），三种成因基本形成了均等划分的情况，其中第一类与第二类占比相同，都为 35%，符合上述对柱状图的解读。

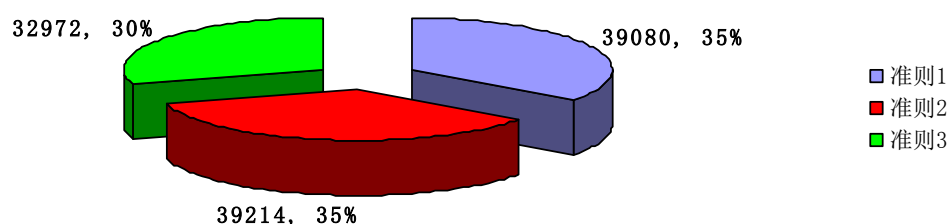


图 6.2 2015—2017 年恐怖袭击事件成因占比分析

### 6.2.2 袭击事件伤亡情况分析

对未来对恐怖袭击事件的成因有了基本的了解后，恐怖袭击造成的伤亡情况是最能反映当前袭击事件态势的直观参照。

故以年度为单位分别统计各年死亡总数、受伤总数、成功的攻击数量并分别

计算三者的年增长率，如表 6.1 所示：

表 6.1 2015—2017 年恐怖袭击伤亡情况统计

	死亡总数	受伤总数	成功的攻击	死亡增长率	受伤增长率	成功攻击增长率
2015	38861	44069	12675	-0.08	-0.07	-0.08
2016	34887	40001	10981	-0.11	-0.09	-0.13
2017	26429	24927	8648	-0.24	-0.38	-0.21
合计	100177	108997	32304	—	—	—

数据来源：根据附件 1 中 2015-2017 年数据计算得出

从数值上看，各年因恐怖袭击导致的死亡数都高达 25000 人，受伤人数更是高至 40000 人，而成功的攻击次数平均仅不到 10000 起，如此看来平均每起成功的攻击就会带来 6 人的伤亡情况，反映出恐怖袭击给国际社会带来的恶劣影响。

观其各项指标增长率情况，从 2015 至 2017 年，不仅各项增长率均为负向数据，其走势也呈现出下降趋势，进一步印证了 6.2.1 中对于国际反恐力量投入的猜测的可能性。为了更直观地给出比对情况，绘制如下折线图：

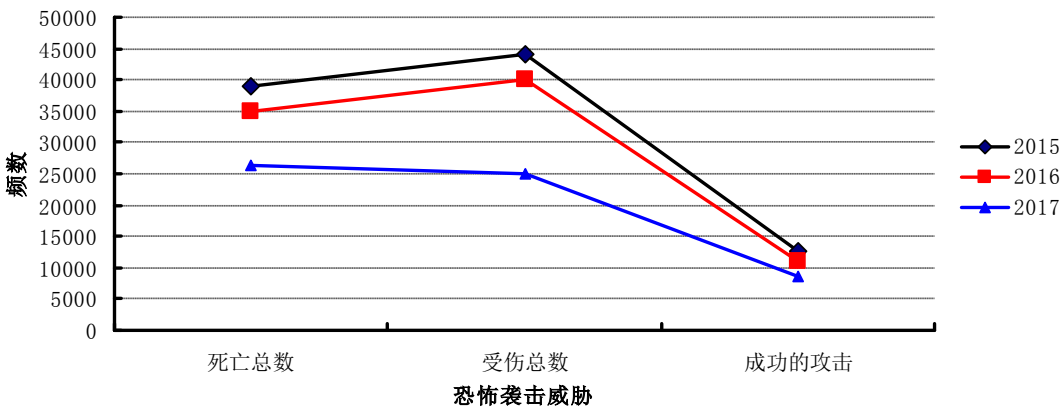


图 6.3 恐怖袭击伤亡情况年度比较

由图 6.3 可更清晰地看出，从 2015 至 2017 年，死亡总数、受伤总数以及成功的攻击者三项反映袭击事件伤亡情况的指标呈明显的逐年递减趋势，且其递减的幅度有逐年增大的趋势。据此我们做出假设：恐怖袭击事件是否在时间分布上也会呈现出某种态势，并进一步深入分析。

### 6.2.3 袭击事件时间分布态势

根据上述伤亡情况分析，考虑到我们的考察对象是 2015—2017 三年间的所有数据信息，可按照季度或月度对事件维度进行精细化处理，本文采用季度处理方法对伤亡情况进一步进行划分，探寻恐怖袭击事件在时间分布上的态势，如表 6.2 所示：

表 6.2 恐怖袭击伤亡情况季度数据

		死亡总数	受伤总数	成功的攻击
2015 年	一季度	9371	10381	3428
	二季度	11161	10952	3227
	三季度	10620	12877	3185
	四季度	7709	9859	2835
2016 年	一季度	8553	11153	2782
	二季度	9524	10639	2947
	三季度	7952	9762	2632
	四季度	8858	8447	2620
2017 年	一季度	6747	6241	2171
	二季度	7380	6767	2350
	三季度	6190	5860	2253
	四季度	6112	6059	1874

数据来源：根据附件 1 中 2015-2017 年数据计算得出

表 6.2 数据详细记录了各项伤亡指标在各年不同季度的发生数，初步可以看出每年第四季度发生伤亡数量比其他季度少，做出如图 6.4 所示柱状图以便观察：

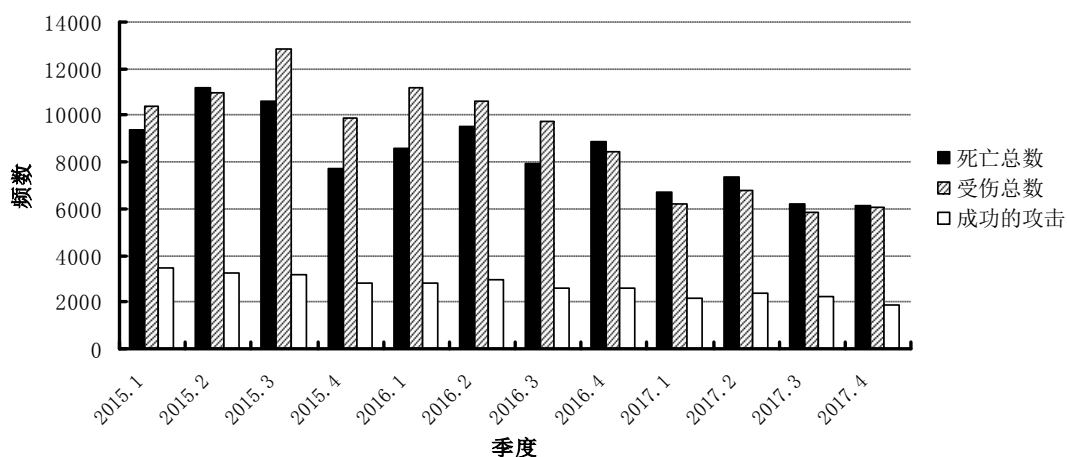


图 6.4 袭击伤亡情况季度对比柱状图

观察图 6.4 发现，2015 年前三个季度，其死亡总数与受伤总数一直高居不下，但在第四个季度出现了明显的滑坡，初步推测在国际大环境下这一年遭受了太多起恐怖袭击，各个国家反恐组织整合力量集中应当恐怖袭击事件，出现了第四季度明显下降的现象；而 2016 年四个季度中，其死亡总数一直在 8000 人左右上下波动，但其受伤总数却呈现出明显的下滑趋势；一直到 2017 年，受伤总数呈明显下降趋势，死亡总数呈波动下降趋势，成功攻击的概率也在 2017 年出现较大幅度的下跌。

同样地，我们将各指标信息在时间维度上以平滑线的形式展现，如图 6.5 所示：

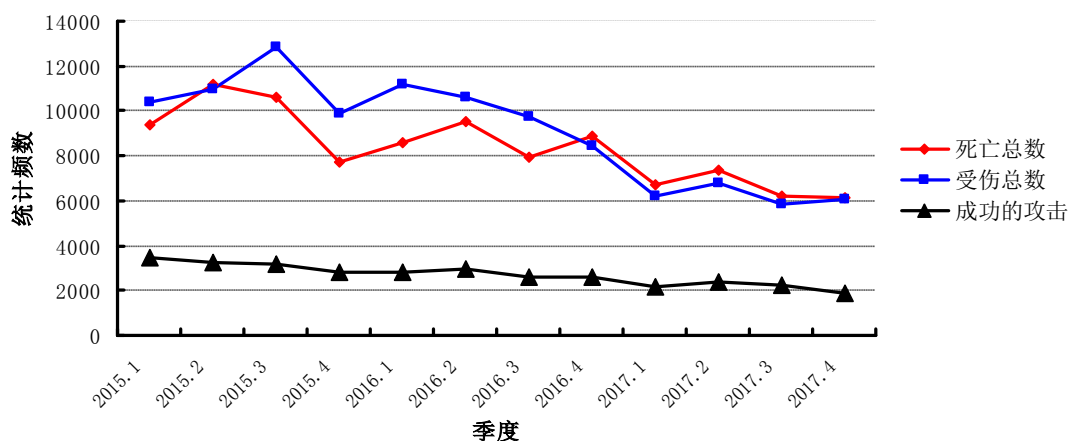


图 6.5 2015—2017 年恐怖袭击伤亡情况季度比较

从图 6.5 中可清晰地看出攻击的成功率在以一个缓慢的速度下降，反映出各国在应对恐怖袭击事件上加强了防范措施，而死亡人数和伤亡人数基本在同一时间点出现波峰波谷，说明反恐措施及时与否和恐怖袭击活动带来的危害损失程度紧密相关。这些现象表明，对于未来恐怖袭击的态势分析，结合反恐组织力量作为博弈面更为合理可靠。

#### 6.2.4 袭击事件空间态势

由上述 6.2.3 发现，恐怖袭击事件发生的时间态势体现出一定的规律性，但并不是十分显著。由此我们试图刻画其地理位置信息，从空间分布上解读其分布态势。

在本题中，我们选取了 2015-2017 年所有袭击事件样本，记录下其经度和维度信息，把维度记为 X 轴，经度记为 Y 轴，利用 Matlab 软件将其整合在平面坐标系中，并根据其集中度指标将其初步划分为五个集中区域。结果如图 6.6 所示：

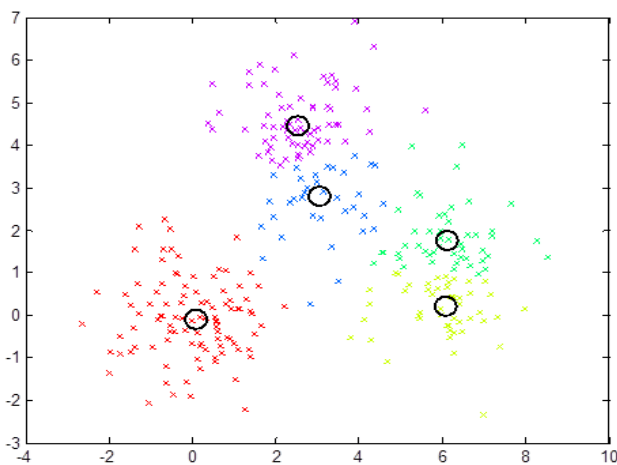


图 6.6 袭击事件空间分布特征图

据图 6.6 信息所示（由于地理位置信息太多，未直接在图表上展示），这三

年内发生的袭击事件的区域位置基本可以分为如下五类：

- (1) 紫色：对应北美地区，主要包括美国、加拿大、墨西哥等地；
- (2) 蓝色：对应中东和北非地区，主要包括伊朗，伊拉克，以色列，约旦，科威特，黎巴嫩，利比亚，摩洛哥，北也门，卡塔尔，沙特阿拉伯，也门南部，叙利亚，土耳其，阿联酋等地；
- (3) 绿色：对应东亚地区，主要包括中国大陆，香港，日本，澳门，朝鲜，韩国，中国台北 等地；
- (4) 黄色：对应东南亚和南亚，主要包括文莱，柬埔寨，东帝汶，印度尼西亚，老挝，马来西亚，缅甸，菲律宾，新加坡，泰国，越南，阿富汗，孟加拉国，不丹，印度，马尔代夫，毛里求斯，尼泊尔，巴基斯坦，斯里兰卡等地；
- (5) 红色：对应中美洲和加勒比海地区，主要包括安提瓜和巴布达、巴哈马、巴巴多斯、伯利兹、开曼群岛、哥斯达黎加、古巴、多米尼加、多米尼亚共和国，萨尔瓦多，格林纳达，瓜德罗普岛，瓜地马拉，海地，洪都拉斯，牙买加，马提尼克，尼加拉瓜，巴拿马，圣基茨和尼维斯，圣露西亚，特立尼达，多巴哥等地。

不难看出，在北美和东亚地区，恐怖袭击事件发生较为集中，说明这里可能会有长期盘踞的恐怖主义集团；而恐怖袭击事件活动范围较广的明显位于中美洲和加勒比海地区，这里因其政治因素也正是恐怖主义活动的高发区；其他各地区也基本呈现出集中趋势，只有少数恐怖主义事件发生偏离了中心范围，属于小规模作案。

#### 6.2.5 袭击事件蔓延态势分析

在 6.2.4 中研究中发现，恐怖主义袭击事件在国家层面上有明显的集中趋势，我们对附件 1 数据库中的相关信息发现，恐怖事件高发区域除了表现在国家层面，在城市这一层级上也反映出其对外的蔓延性与扩张性。为此我们记录了这三年来恐怖袭击发生于城市自身还是城市附近的相关信息，如图 6.7 所示：

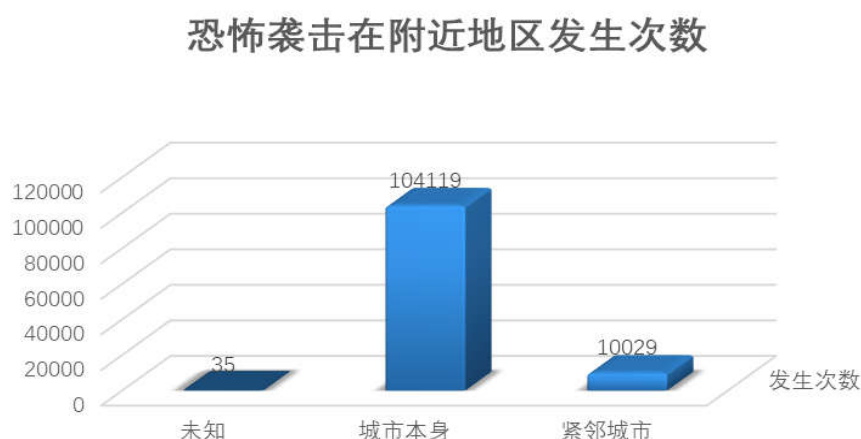


图 6.7 恐怖袭击事件在附近区域发生的次数

据图 6.7 可以看出，大多数的恐怖袭击事件具有很明显的针对性和很强的目

的性，其主要集中于各个国家的城市中，极少数发生于城市附近地区，这恰恰印证了我们在 6.2.1 中对于恐怖主义袭击事件发生原因的分析，其主要目的出于政治、经济、宗教或者社会目的，而人流量最为密集的城市必然成为首选目标。

既然恐怖袭击事件主要集中于城市自身发生，那么是否说明其蔓延性很弱？这需我们从另一个角度去考虑，虽然恐怖袭击事件的发生并没有波及到其附近地区，但其引发恐怖袭击的动机却可以远远达到传播恐怖主义的目的。

基于该视角，我们在对 2015 年至 2017 年共计 39453 个样本的城市数据信息进行配对后，选出了在相近时间不同地点主动声明承担恐怖袭击责任的组织数量，其统计数据如表 6.3 所示：

表 6.3 相近时间不同地点恐怖组织发布声明情况

年份	第一声称负责	第二声称负责	第三声称负责	争相负责声明
2015	2950	84	17	60
2016	2860	128	43	86
2017	2640	75	12	69
总数	8450	287	72	215

数据来源：根据附件 1 中 2015-2017 年数据计算得出

由表 6.3 可以看出，第一声称负责的恐怖集团三年间逐渐减少，但争相负责的恐怖集团却有增长的迹象，其原因主要在于各国反恐力量对于国内恐怖主义势力的打击不断加强。其中，比较特殊的是 2016 年，这一年无论是第一声称负责的恐怖集团还是第二第三个，其数量都较其他两年多，推测与 2016 年的国际反恐形势有很大关联。

首先将这三年的累计频数折算成百分比的状况，如图 6.8 所示：

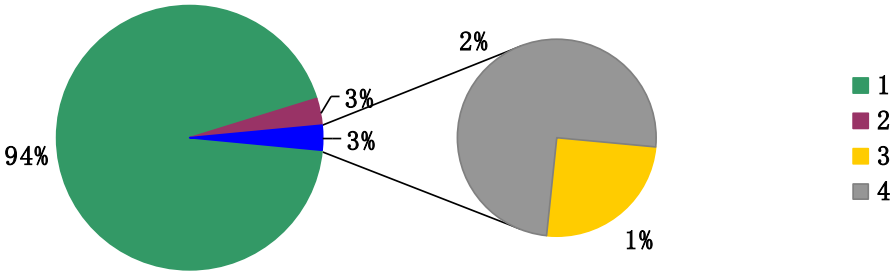


图 6.8 2015—2017 年声称负责恐怖集团比较

图 6.8 中，1、2、3、4 分别代表第一、二、三声明对最近恐怖袭击事件负责的集团数量和争相声明对最近恐怖袭击事件负责的集团数量。其中**争相声明**对最近恐怖袭击事件负责的集团数量这一指标最能反映近几年来国际恐怖主义的态势，而各级宣称对最近恐怖袭击事件负责的集团数量在年度的变化情况也可以反映出当前恐怖主义活动的基本走势，拟采用折线图绘制其走势。



绘制过程中发现,由于第一声明对最近恐怖袭击事件负责的集团数量较其余两类数量高出很多,故对原始数据进行对数化处理绘制成如图 6.9 所示对数曲线图:

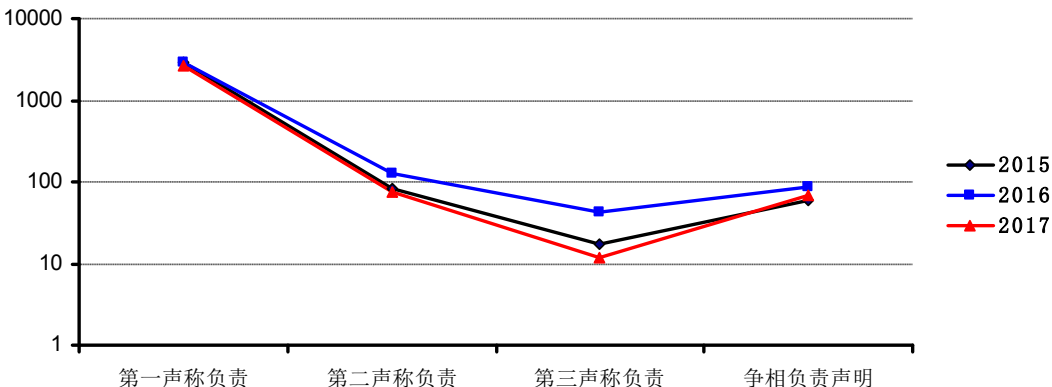


图 6.9 2015—2017 年声称负责走势变化

据图 6.9 可以很直观地看出,2017 年各级别的声明对最近恐怖袭击事件负责的集团数量都处于最低位置,说明在 2017 中各国以及国际上投入的反恐力量取得了很大成效;2015 年各项数据指标基本与 2017 年持平,只有第三声称负责的恐怖主义组织比 2017 年稍高;但是 2016 中各项数据显示,这一年的恐怖主义活动较为活跃,各级声明对最近恐怖袭击事件负责的集团数量都处于一个较高的范围,结合 2016 年的重大国际事件来看,该现象可能产生于国际恐怖主义和国际社会的矛盾激化。

### 6.2.6 恐怖袭击事件态势分析<sup>[5]</sup>结论

**结论:** 根据上述对恐怖袭击事件发生的原因、时间特性、空间特性以及蔓延特性,得出的主要结论是: 2015—2017 年间,恐怖袭击事件基本呈逐年下降的趋势,袭击成功几率也有一定幅度的下跌,在反恐的环境下,政府反恐力量的投入与应对及时性是制约恐怖主义活动的关键因素<sup>[6]</sup>。

## 6.3 模型的建立

### 6.3.1 误对策模型<sup>[2]</sup>的建立

根据上述对于国际恐怖袭击事件的态势分析,要建立模型对近三年来发生的恐怖袭击事件进行研究,我们拟采用从博弈的角度对“反恐态势”建立一个涵括了政府反恐力量与恐怖分子的相对评估模型。

恐怖袭击事件中博弈双方高度不对称和存在认知差异,误对策方法提供了合适的分析框架。本文尝试建立误对策分析模型,去分析不同情境下策略冲突的平衡结果,模型构建示意图如图 6.10 所示:



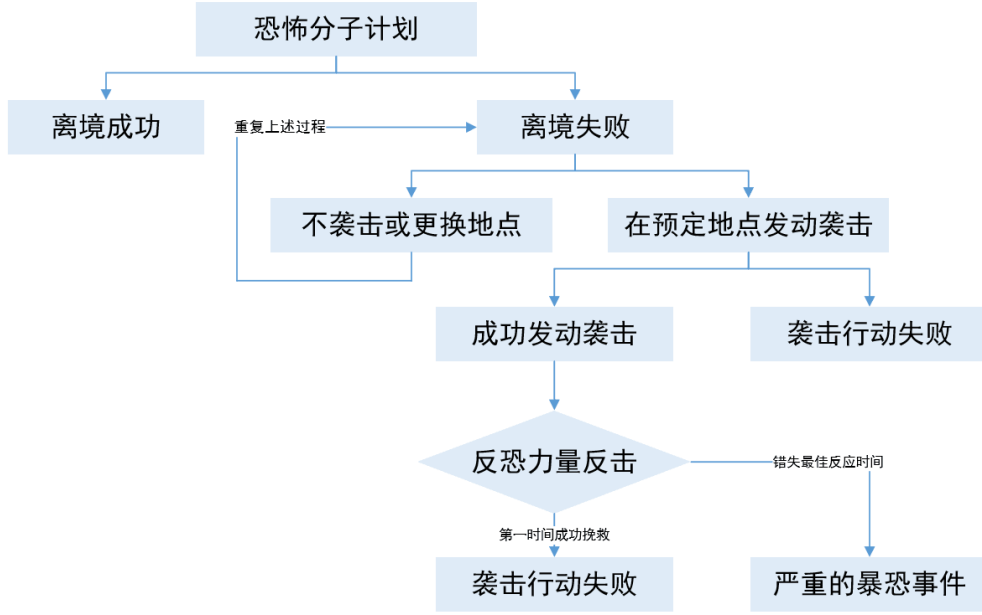


图 6.10 基于安全策略的恐怖主义态势分析误对策模型示意图

误对策模型的基本定义如下：

**定义 1：** 政府反恐力量  $G$  和恐怖分子  $T$  组成的参与者集合  $N = (G, T)$ ，且  $\Phi_G$  与  $S_G$  分别为政府反恐力量  $G$  的纯策略和混合策略， $\Phi_{TG}$  和  $S_{TG}$  分别为政府反恐力量  $G$  感知到的恐怖分子  $T$  采取的纯策略和混合策略。

**定义 2（一阶误对策和二阶误对策）：** 一阶误对策是指至少有一个参与者的感知存在着偏差，但是双方均未意识到，一阶误对策实质上双方在进行不同的冲突对策，二阶误对策是指至少有一个参与者意识到面临着一场一阶误对策，即意识到对方存在着误判。

**定义 3（解释函数）：** 定义映射  $F_{TG} : \Phi_T \rightarrow \Phi_{TG}$  为政府反恐力量  $G$  对于恐怖分子  $T$  的解释函数，对于恐怖分子的策略  $s_T \in S_T$ ，政府反恐力量的混合策略解释函数  $f_{TG} : S_T \rightarrow S_{TG}$  定义为：
$$f_{TG}^k(s_T) = \sum_{h \in \{h \in \Phi_T | F_{TG}(h) = k\}} s_T^h, k \in \Phi_{TG}$$

其中， $f_{TG}^k$  是构成混合策略  $f_{TG}$  的第  $k$  个纯策略。

类似的，我们可以定义恐怖分子  $T$  对于政府反恐力量  $G$  的混合策略解释函数。

**定义 4（误对策纳什均衡）：** 对于参与者集合  $N = (G, T)$ ，设  $f_{TG} : S_T \rightarrow S_{TG}$  为政府反恐力量的解释函数，定义恐怖分子的最优策略记为  $s_T^* \in S_T$ ，以及政府反恐力量感知到得恐怖分子最优策略记为  $s_{TG}^*(S_G) = f_{TG}(s_T^*)$ ，则策略组合  $s_T^* \in S_T$

是一个误对策纳什均衡，当满足下列条件时：

$$\forall s_G \in S_G, U(s_G^*, s_{TG}^*) \geq U(s_G, s_{TG}^*) \quad (1)$$

$$\forall s_T \in S_T, U(s_T^*, s_{GT}^*) \geq U(s_T, s_{GT}^*) \quad (2)$$

如果对于所有的  $s_T \neq s_T^*$  和  $s_G \neq s_G^*$ ，上述不等式（1）和（2）为严格不等式，则策略组合  $s^*$  称为严格误对策纳什均衡。

**定义 5（误对策的平衡结局）：**参与者  $i$  感知的对策平衡结局集合记为  $E_i$ ，

则误对策的平衡结局  $q = (s_T^*, s_G^*)$  包括如下两种情况：

（1）如果  $q \in (E_T \cap E_G)$ ，则  $q$  是一个可维持的误对策平衡结局；

（2）如果  $q \notin (E_T \cap E_G)$ ，则  $q$  是一个失稳的误对策平衡结局。

在上述 5 个基本定义下，我们建立误对策模型来对恐怖主义态势进行分析，具体过程如下。

### 6.3.2 误对策模型中政府反恐力量感知到的博弈<sup>[3]</sup>

在对 2015—2017 年恐怖袭击事件梳理中，我们发现北美地区在这三年中是恐怖主义事件的高发地区，我们提取出发生于北美地区的恐怖袭击事件相关文本及数据信息作为案例，进行误对策态势分析，过程如下。

在恐怖袭击事件发生之前，除了恐怖分子知道自己要采取恐怖袭击的策略外，政府反恐力量往往并不清楚该情况政府反恐力量的策略几个包括如下 2 种策略：

$A_1'$ ：无论是否有恐怖袭击事件发生，时刻保持戒备，采取全部动员或者轮班交接的部分动员方式；

$A_2'$ ：只是例行巡逻或者根本就不动员，直到事故发生之后才广泛地动员。

上述策略为政府反恐力量的自身防御策略，但是其对于恐怖分子的攻击策略判断则存在着较大的差异，从政府反恐力量感知的角度来看，恐怖分子可能采取的策略为：

$K_1'$ ：制造恐怖袭击事件，例如聚众闹事打砸抢、发动自杀性袭击、滥杀无辜等；

$K_2'$ ：无恐怖袭击事件，或者仅为恐怖分子虚张声势的威胁。

在恐怖袭击事件发生之前，相关反恐力量主要用于处置大型人质绑架、航空器劫持等类似的恐怖案件，对于人群密集场所的疏于防范，政府机构决策者一般认为，首先，内地恐怖袭击事件发生的概率很小；其次，由于爆炸物品和枪支弹药的严格管理，较为容易获得的刀具、斧头和汽油成为了恐怖分子的首选，简单的武器难以造成严重的伤害结果；第三，考虑到火车站、汽车站和商场等人群密

集场所进行大范围防御部署的防御成本，当无恐怖袭击发生的征兆或情报预警时，主要有安保部门进行常规监控。

基于上述分析，政府反恐力量面对不同策略组合的收益偏好次序为：

(1) 给定恐怖分子发动袭击的策略  $K_1'$  下，政府反恐力量进行主动防御  $A_1'$  获得的支付，要高于被动应对  $A_2'$  获得的支付；

(2) 给定恐怖分子不发动恐怖袭击  $K_2'$  下，考虑到大范围防御的成本，政府反恐力量不进行防御动员  $A_2'$  获得的支付，要高于进行防御  $A_1'$  获得的支付。

派出所和公共场所以公安机关为主的安保力量通常能吹好小规模袭击事件，而政府反恐力量认为内地很少发生大规模的恶性恐怖袭击事件，即使个别恐怖分子发动了恐怖袭击事件，现有安保力量也会在较短时间内制服并控制局势。因此，政府反恐力量认为，恐怖分子对于不同策略组合的收益偏好应满足：

(1) 给定政府反恐力量进行防御动员  $A_1'$ ，此时恐怖分子发动袭击  $K_1'$  获得的收益收益最小；

(2) 给定政府反恐力量例行巡逻或者不动员  $A_2'$ ，此时恐怖分子发动袭击  $K_1'$  获得的收益要高于放弃袭击机会  $K_2'$  的收益；

(3) 考虑到公共场所进行大范围防御动员带来的社会成本，以及由此产生的社会恐惧情绪，恰是恐怖分子想要达到的目标之一，因此，在反恐力量大范围地开展防御动员  $A_1'$  下，恐怖分子却躲避起来不发动袭击  $K_2'$  的策略是恐怖分子最偏好的。

根据上述政府反恐力量感知到的双方策略集合和收益排序，可以看出恐怖分子的收益偏好具有明确顺序，但是政府反恐力量的收益偏好存在六种可能的情况，如表 6.4 所示，为了简化分析，这里收益偏好采用 Bennett 等学者提出的序数效用，并不影响均衡分析结果，其中，数值越大表示越偏好。

**表 6.4 政府反恐力量感知到的恐怖袭击博弈策略和收益偏好**

策略组合	政府反恐力量收益偏好顺序						恐怖分子收益偏好
A1, K1	2	3	4	3	4	4	1
A2, K1	1	1	1	2	2	3	3
A1, K2	3	2	2	1	1	1	4
A2, K2	4	4	3	4	3	2	2

### 6.3.3 误对策模型中恐怖分子感知到的博弈

在 6.2 中的恐怖袭击事件特征分析中我们发现，大部分袭击事件集中发生在城市自身所在区域，针对性极强，尤其是部分核心城市，人口集中、媒体发达，信息传播速度极快，一旦遭受恐怖袭击，其危害会更广，影响会更大。因此，恐怖分子在极端思想的蛊惑下，经过精心策划，如果将攻击目标确定为火车站和早

市等防范薄弱、容易造成重大伤亡的公共场所；将发动时机确定为国际重大事项前夕的敏感时刻，以扩大事件的影响。

在恐怖分子看来，政府反恐力量的策略集合包括：

$A_1$ ：在国际重大事项等敏感时刻，针对公共场所进行大范围的防御动员；

$A_2$ ：在国际重大事项等敏感时刻维持例行国际反恐巡逻或者不动员。

恐怖分子采取的策略为：

$K_1$ ：如果政府反恐力量在国际重大事项等敏感时刻不进行大范围的防御动员，恐怖分子将择机发动袭击；否则，可能推迟发动袭击的时间，或者改变别地发动袭击，或者放弃袭击；

$K_2$ ：恐怖分子在国际重大事项之前等敏感时刻，放弃发动袭击。

恐怖分子感知到双面面对着不同策略组合存在着各自的收益偏好，其中，恐怖分子选取国际重大事项前夕等敏感时期发动袭击，其收益偏好满足如下条件：

1) 如果反恐力量不进行大范围的防御  $A_2$ ，则恐怖分子在国际重大事项前夕等敏感时期发动袭击  $K_1$  最偏好；

2) 如果政府反恐力量没有进行防御动员  $A_2$ ，恐怖分子却错失时间、没有发动袭击  $K_2$ ，该结果最差；

3) 如果政府反恐力量已经大范围防御动员  $A_1$ ，此时恐怖分子推迟或取消袭击  $K_2$  更好。

在恐怖分子看来，政府反恐力量并不清楚是否发生、何时发生恐怖袭击事件；而且因为城市公共场所太多，现有反恐资源难以做到全部而且有效地覆盖。因此，面对不同策略组合政府反恐力量的收益偏好为：

1) 不管是否发生恐怖袭击，针对公共场所进行大范围防御动员  $A_1$ ，对于政府部门而言都将造成难以承受的压力；

2) 如果恐怖分子放弃袭击  $K_2$ ，此时政府反恐力量将更偏好维持例行巡逻  $A_2$ ，而不进行大范围的防御动员  $A_1$ 。

同理，根据上述恐怖分子感知到的双方策略组合和相应的收益偏好次序，可以得到恐怖分子感知到的恐怖袭击博弈，如表 5 所示。其中，政府反恐力量的收益偏好存在着三种可能次序。

表 6.5 恐怖分子感知到的恐怖袭击博弈策略和收益偏好

策略组合	政府反恐力量收益偏好顺序			恐怖分子收益偏好
A1, K1	1	1	2	2
A2, K1	3	2	1	4
A1, K2	1	1	2	3
A2, K2	2	3	3	1

#### 6.3.4 引入安全策略优化模型

但是，上述误对策模型并不适用于所有情况。当决策者对面临的局势一无所知时，不考虑恐怖分子采取何种最优的攻击策略造成的不确定性，此时恐怖分子所有可能的攻击策略都将被同等采用。因此我们在模型中引入安全策略对纳什均衡策略进行模型修正。

定义 4 给出了误对策的纳什均衡，其适用于博弈双方的策略是相互影响的情况下做出的最优决策，即给定自己感知到的对方最优策略后，自己也将选取最优策略。尽管纳什均衡概念作为博弈论的理论基石得到了广泛的应用，但是，当政府反恐力量等决策者面临着高度不确定的环境时，即使是从自身角度出发，其也很难判断出恐怖分子的最优策略是什么。故而在这里采用另一个均衡策略，即安全策略。

**定义 6 (安全策略):** 在策略式表达的博弈  $(X_1 \dots X_n; \mu_1 \dots \mu_n)$  中，任何策略  $x$  如果满足如下问题的解，则称之为安全策略 (security strategy)，又称为最大最小策略或防守策略：

$$\max_{x_i \in X_i} \min_{x_{-i} \in X_{-i}} \mu(x_i, x_{-i})$$

**定义 7 (零和博弈在纯策略下的安全策略):** 一个策略表达式表述的两人有限零和博弈可以表示为  $(X_1, X_2; A)$  其中，双方的策略集合  $X_1 = \{a_1, a_2, \dots, a_m\}$  和  $X_2 = \{\beta_1, \beta_2, \dots, \beta_n\}$ ， $A = (a_{ij})_{m \times n}$  是参与者 1 努力最大化而参与者 2 努力最小化的支付矩阵。如果下面等式成立，则纯策略组合  $(a_i^*, \beta_j^*)$  为两人有限零和博弈在纯策略下的最优解：

$$\max_{x_i \in X_1} \min_{x_j \in X_2} a_{ij} = \min_{x_i \in X_2} \max_{x_j \in X_1} a_{ij} = a_{i^* j^*}$$

**定义 8 (零和博弈在混合策略下的安全策略):** 当两人有限零和博弈在纯策略下不存在安全策略，即  $\max_{x_i \in X_1} \min_{x_j \in X_2} a_{ij} \neq \min_{x_i \in X_2} \max_{x_j \in X_1} a_{ij}$  时，此时存在着最优混合的安全策略，当满足如下条件时：

$$\max_{x \in S_1} \min_{y \in S_2} E(x, y) = \min_{y \in S_2} \max_{x \in S_1} E(x, y)$$

其中， $x$  和  $y$  为混合策略， $E(x, y)$  为混合策略的支付函数，

$$E(x,y)=x^T Ay=\sum_i\sum_j a_{ij}x_iy_j。$$

### 6.4 误对策博弈的安全策略分析与结论

在文本分析的恐怖袭击误对策博弈中，双方收益偏好次序不同，不能严格满足零和博弈的限制条件，因此安全策略并不等于最优的纳什均衡策略，由于政府反恐力量与恐怖分子之间存在着严重的认知差异，政府决策者难以判断恐怖分子的最优攻击策略，因此本文将采取保守的安全策略而不是纳什策略，分析双方的均衡行为。具体步骤如下：

**Step1:** 在表 6.4 所示的政府反恐力量感知到的恐怖袭击博弈中，根据政府反恐力量不同策略组合的偏好存在六种可能的均衡结果，如表 6.6 所示。其中，如果采用传统的纳什均衡的稳定性概念，则唯一的纳什均衡策略为 $\{A_2, K_2\}$ ，即政府反恐力量认为双方的最优策略组合是不袭击和不动员。

表 6.6 政府反恐力量感知的恐怖袭击博弈模型

	K1	K2	
A1	<u>2</u> (3, 4, <u>3</u> ,4,4) , 1	3 (2,2,1,1,1) , <u>4</u>	[2(2,2,1,1,1)]
A2	1 (1, 1,2,2,3) , 3	<u>4</u> (4,3,4,3,2) , 2	[1(1,1,2,2,2)]
	[1]	[2]	

**Step2:** 根据表 6.5 所示的恐怖分子感知到的恐怖袭击博弈，根据恐怖分子感知到的政府反恐力量不同策略组合偏好次序，存在三种可能的均衡结果。前两种如表 6.7 所示，第三种如表 6.8 所示：

表 6.7 反恐力量感知的恐怖袭击博弈(防御成本较高)

	K1	K2	
A1	1(1), <u>2</u>	1(1), <u>3</u>	[1(1)]
A2	<u>3</u> (2), 4	<u>2</u> (3), 1	[2(2)]
	[2]	[1]	

表 6.8 反恐力量感知的恐怖袭击博弈(防御成本可承受)

	K1	K2	
A1	<u>2</u> ,2	2, <u>3</u>	[2]
A2	1, <u>4</u>	<u>3</u> ,1	[1]
	[2]	[1]	

**结论 1:** 在双人零和博弈中，由于结果必然是帕累托最优的，此时安全策略称之为“最优的安全策略”，即纳什均衡策略。

根据对恐怖袭击事件误对策博弈的安全策略分析，可以得出以下两种情况：

**第一种情况:** 当政府反恐力量针对公共场所进行大范围防御和动员的成本较高时，无论政府反恐力量还是恐怖分子均认为政府部门将采取例行巡逻或不防御

的安全策略，此时政府反恐力量误认为恐怖分子不会采取恐怖袭击  $K_2$ ，但是恐怖分子自身的安全策略却是选择敏感时刻发动袭击  $K_1$ 。

**第二种情况：**当政府反恐力量对于公共场所进行大范围的防御和动员的成本可以接受时，无论政府反恐力量还是恐怖分子均认为政府部门的安全策略是采取防御措施，此时政府反恐力量误认为恐怖分子不会采取恐怖袭击  $K_2$ ，但是恐怖分子自身的安全策略却是选择敏感时刻发动袭击  $K_1$ 。

**结论 2：**综合来看，对于北美地区未来的反恐态势来说，当恐怖分子采取袭击策略的比例大于最优策略均衡临界值时，政府应采取防御策略；当恐怖分子采取袭击策略的比例小于最优策略临界值时，政府可以适宜降低防御程度，但不应放弃防御策略。同时，该模型也同样适用于国际反恐态势分析。

## 7 问题四模型的建立与求解

### 7.1 问题分析

当前，随着 IS 的崛起，中东局势逐步恶化，恐怖主义活动越发频繁。城市作为国家经济、政治、文化、信息交流的枢纽地，商品、资金、技术、人口等社会要素的集散地，越来越成为恐怖主义袭击的主要对象。

结合附件 1 的 GTD 数据库的数据，我们进行数据挖掘，使用模糊层次分析法建立城市遭受恐怖袭击的风险评估模型，从而对国家城市反恐力量的规划、科学预防当前城市恐怖袭击提供辅助决策支持。

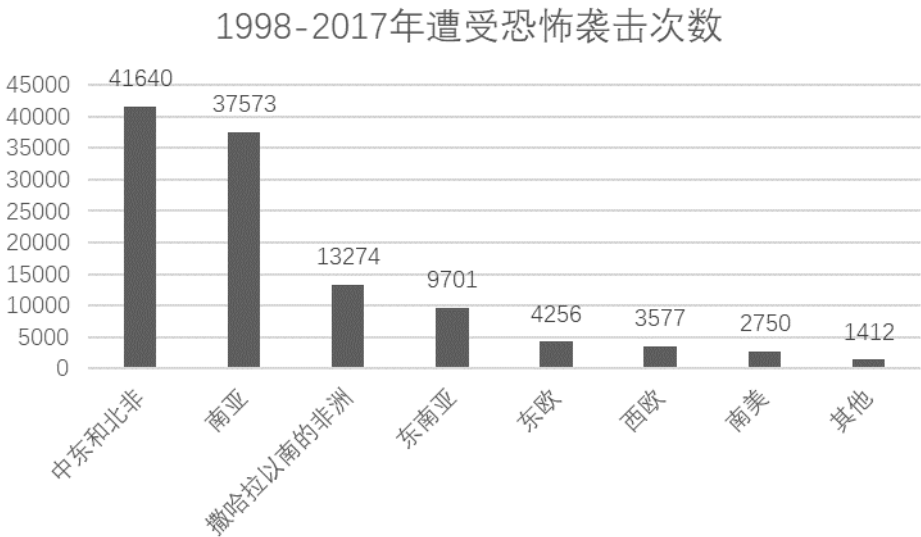


图 7.1 各地区 1998-2017 遭受攻击总次数

由图 7.1 可知，在发生恐怖袭击事件的 12 类地区中，南亚、中东和北非、撒哈拉以南的非洲分别为发生恐怖袭击事件最多的地区的前三类。其中，中东、北非地区发生恐怖袭击事件次数最多。在这三类地区包含的国家中，绝大多数的国家经济较为落后，政治环境动荡不安。由此，我们可以推测，恐怖事件的发生与地区的经济发展息息相关，与一个地区的防御力量也有必然联系。因此，反恐

活动必须关注经济欠发达国家的经济发展，同时对于政治环境动荡的国家也要保持较高的关注。

7.2 指标选取及数据来源

恐怖分子选择袭击某城市主要包括意识形态、组织实施、期望产生，3 个方面影响<sup>[7]</sup>。综合三方面，影响恐怖分子选择城市的因素有政治因素、民族因素、宗教因素、经济因素、人口因素、交通因素、公共安全、种族因素，如图 7.1 所示。

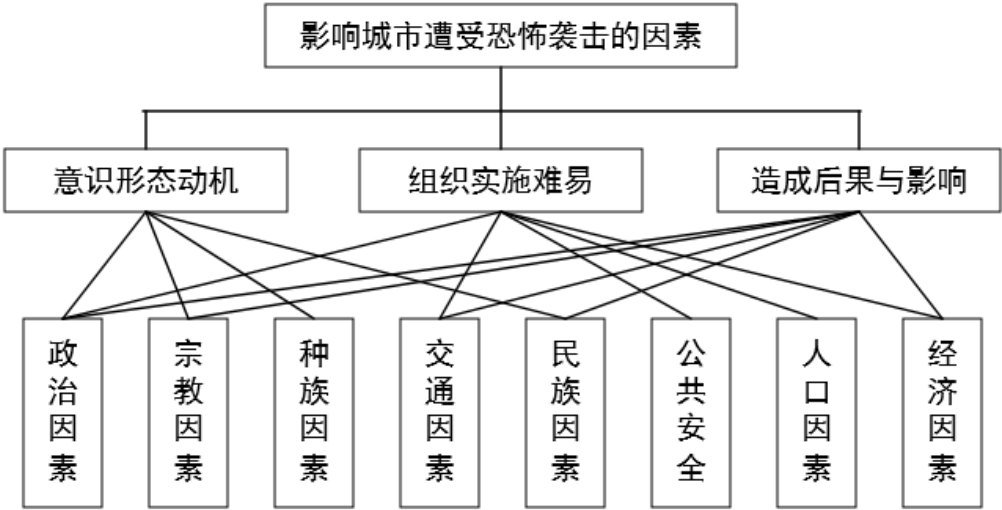


图 7.1 影响城市遭受恐怖袭击的因素

结合这些影响因素以及问题一中建立的恐怖袭击事件的危害性影响因素指标体系，选取评估城市遭受袭击的影响因素的指标，建立指标体系，指标体系如下表 7.1 所示：

表 7.1 城市遭受袭击影响因素指标体系

一级指标	二级指标	三级指标
意识形态动机	自身情况	受害者类型
		犯罪集团
组织实施难易	外界支持情况	武器类型
		攻击类型
		声称负责
造成后果与影响	损失情况	财产损失
		财产损失程度
	舆论影响	是否超出人道主义
		对人群社会影响
	人员伤亡情况	死亡总数
		受伤总数
		人质或绑架受害者总数



### 7.3 基于模糊层次分析法建立城市遭受恐怖袭击风险评估模型

模糊综合分析法是将模糊数学的理论运用到层次分析中，将该方法引入模糊理论，将难以量化的定性问题量化表达，为评价与决策提供依据，一般过程如下：

#### (1) 建立层次结构模型

模糊层次分析法的层次结构模型与一般层次分析法的层次结构建立方法相同，分为目标层、准则层（子准则层）、方案层。

$$(2) \text{ 构建模糊矩阵 } A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \quad a_{ij} = (l_{ij} m_{ij} u_{ij})$$

上式中， $a_{ij} = (l_{ij} m_{ij} u_{ij})$  是以  $m_{ij}$  为中值的闭区间，而  $m_{ij}$  就是层次分析中所采用的1~9标度中的整数模糊矩阵，仍为正负反矩阵，即

$$a_{ji} = a_{ij}^{-1} \approx \left( \frac{1}{u_{ij}} \frac{1}{m_{ij}} \frac{1}{l_{ij}} \right)$$

#### (3) 计算各指标权重：

1) 元素  $i$  的综合模糊值  $D_i^k$ （初始权重）。计算方式如下：

$$D_i^k = \sum_{j=1}^n a_{ij}^k / \sum_{i=1}^n \sum_{j=1}^n a_{ij}^k, \quad i = 1, 2, \dots, n$$

2) 去模糊化，求出最终权重。

##### ① 去模糊化计算

$$v(D_i^k \geq D_j^k) = u(d) =$$

$$\begin{cases} 1 & m1 \geq m2 \\ \frac{l2 - u1}{(m1 - u1) - (m2 - l2)} & m1 \leq m2, u1 \geq l2(3) \\ 0 & \text{otherwise} \end{cases}$$

$$i = 1, 2, \dots, n_k$$

##### ② 最终计算权重

$$w_{ih}^k(A_i^k) = \min v(D_i^k \geq D_j^k)$$

上式中， $A_i^k$  表示第  $k$  层第  $i$  个因素， $i = 1, 2, \dots, n_k, j = 1, 2, \dots, n_k$ ，归一化后

得到  $w_h^k = (w_{1h}^k, w_{2h}^k, \dots, w_{n_k h}^k)^T$ ，表示第  $k$  层次各因素对第  $k-1$  层次上第  $h$  个因素的最终权重。

3) 使用模糊判断矩阵确定其它层次、其他指标的权重，得到总权重。

(4) 利用 MATLAB 软件计算得出最终解，对所得结果进行评估，并选择最佳方案。

### 7.4 案例分析

以中国 15 个城市为例建立恐怖袭击风险评估模型，步骤如下：

Step1 层次结构构建

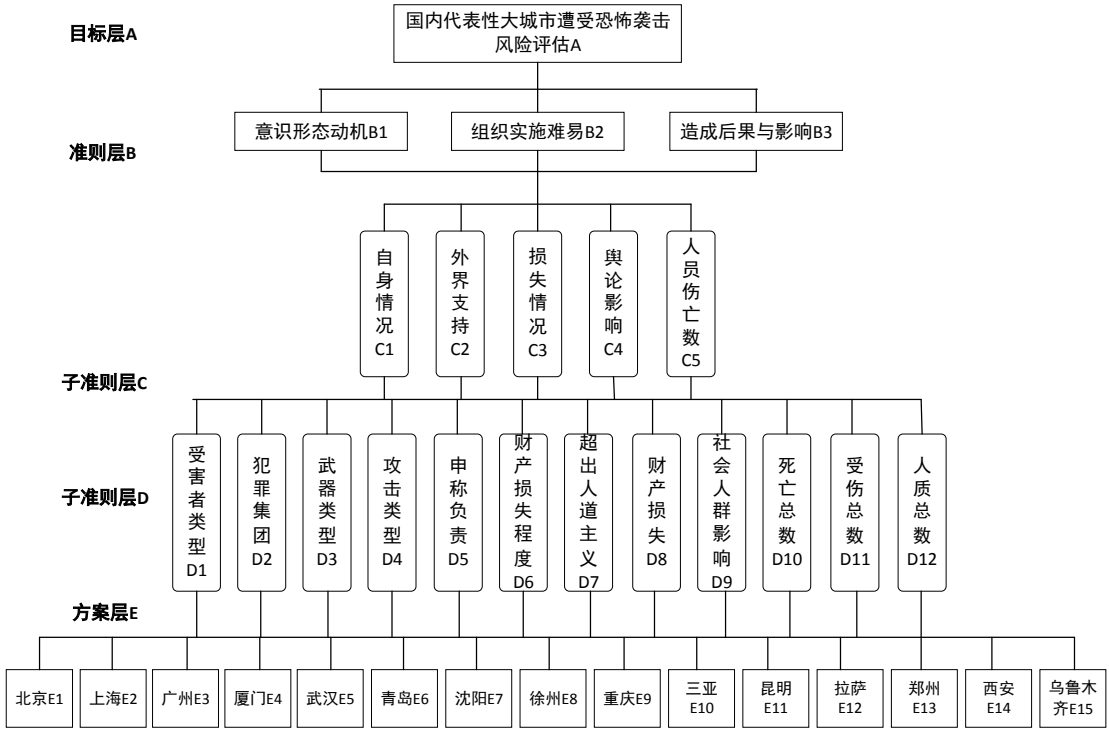


图 7.2 风险评估层次结构

具体城市遭受恐怖袭击风险评估层次结构模型如图 7.2 所示。

Step2 构造并求解模糊判断矩阵

模糊判断矩阵即确定各层次元素的相对重要性权重，包括对准则层的因素权重确定与子准则层的因素权重确定。在对城市遭恐怖袭击的因素进行定性分析基础上，采用专家评估法，运用三角模糊数来表示指标权重，从而得到上述城市遭恐怖袭击的因素进行定性分析基础上，从而得到上述城市遭恐怖袭击风险评估的模糊判断矩阵，并按照以下步骤进行求解：

- 1) 一级指标处理。处理专家评估矩阵，构建各城市  $E_i (i=1,2,\dots,15)$  的  $D_j (j=1,2,\dots,19)$  模糊判断矩阵，并求解初始权重。
- 2) 二级指标处理。处理专家评估矩阵，构建  $C_m (m=1,2,\dots,8)$  的模糊判断矩阵，并求解  $C_m$  的初始权重。
- 3) 三级指标处理。经过专家评估，构建  $B_n (n=1,2,3)$  的模糊判断矩阵，并求

解初始权重。

4) 计算  $A$  的模糊判断矩阵，求解  $A$  的初始权重。

5) 经归一模糊化计算，一次求解  $D_j$ 、 $C_m$ 、 $B_n$  及  $A$  的最终权重；归一化后得到总权重，并得出最终结果。

Step3 风险排序及专题图展示

在构建城市遭受恐怖袭击风险评估层次结构模型基础上，构造模糊判断矩阵，并进行一系列综合权重的计算，对所选的 15 个主要城市进行风险评估排序。即 E1-E15 遭受恐怖袭击的风险排序，在归一化处理后，得到结果如下表 7.2 所示:

表 7.2 城市风险评估权重表

城市	权重
北京	0.041057624
上海	0.02861496
广州	0.025247025
厦门	0.007651522
武汉	0.011294798
青岛	0.009823392
沈阳	0.023628433
徐州	0.0040139
重庆	0.011726244
三亚	0.016196301
昆明	0.048904893
拉萨	0.309621863
郑州	0.002677231
西安	0.00323265
乌鲁木齐	0.456309162

在以上数据处理的基础上，利用 ArcGIS 实现城市遭受恐怖袭击风险的专题图显示，如图 7.3 所示。

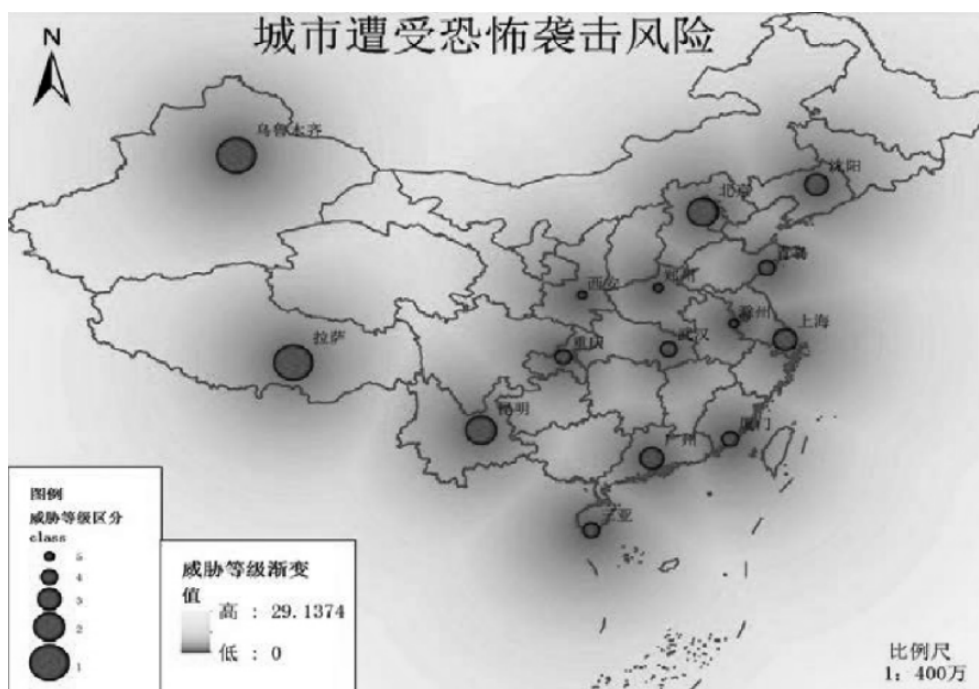


图 7.3 城市遭受恐怖袭击风险地图

根据表 7.2 和图 7.3 可见，最易遭受恐怖袭击的城市是乌鲁木齐，其次是拉萨。结合这两个城市的地理位置和社会背景，我们不难验证这个模型的可靠性。对此，我们可以根据数据，评估城市的遭受袭击的风险量，从而对国家城市反恐力量的规划、科学预防当前城市恐怖袭击提供辅助决策支持。同时对因素及反恐斗争的现实需要制定应对策略, 以期恐怖活动的风险评估提供有价值的情报决策支持。

## 8 模型评价

### 8.1 本文所建立模型的优点

本文针对不同的问题，综合运用模糊聚类、模式识别、误对策分析模型、模糊层次分析等统计分析工具，模糊聚类在样本量及指标众多的情况下具有优异的属性，可以将各项非数值型数据进行量化处理，有效利用文本和字符串变量，在保存大部分样本信息的前提下得出较为合理的结果。模式识别是解决匹配问题较为常用的分析方法，可以为缺失指标信息找到合理的信息匹配，且准确率较高。同时，本文的误对策分析模型添加了安全策略模型进行修正，使得博弈结果更加贴近实际情况，合理可靠。模糊层次聚类可以有效再次提炼附件 1 的数据集中的有效信息。我们在对各问题建模求解以及量化处理的过程中，综合运用 SPSS、Excel、Matlab、R 等统计分析软件对数据进行处理，是的各项聚类、相关性测度等运算过程简便化，运算结果真实可信，另外，我们对几乎所有的模型都进行了检验和对比分析。

### 8.2 本文所建立模型的缺点

由于附件 1 中指标高达一百多个，初始样本量更是高达十万条，给数据量化处理工作带来很大困难，因此在筛选过程中会遗漏部分可能存在重要作用的变量信息，并且我们在求解权重的时候，可能方法不尽合理，会对结果造成一定程度

的影响。最后，由于指标提取后，丢弃一些指标会造成小信息的丢失，所以得到的结论可能会存在一些可接受的误差。

### 参考文献:

- [1] 滑腾飞, 白玲, 刘大庆. 城市遭受恐怖袭击的风险评估模型研究[J]. 测绘与空间地理信息, 39(5):38-40, 2016,.
- [2] 刘德海, 周婷婷. 基于认知差异的恐怖主义袭击误对策分析[J]. 系统工程理论与实践, 10:2646-2656, 2015.
- [3] 刘德海等, 基于心智模型的恐怖主义袭击扩展式演化博弈分析[J]. 中国管理科学, 8:71-78, 2018.
- [4] 王雪梅, 恐怖主义犯罪发展特点分析[J]. 环球法律评论, 1:21-33, 2013.
- [5] 李益斌, 欧洲恐怖主义的新态势及原因分析[J]. 情报杂志, 3:55-63, 2018.
- [6] 傅子洋等, 基于贝叶斯网络的恐怖袭击预警模型研究[J]. 灾害学, 3:184-189, 2016.
- [7] 夏一雪等, 基于 logistic-ABC 的恐怖活动风险因素识别与指标评级模型研究[J]. 情报杂志, 8:1-9, 2018.
- [8] 黄力华. 从9.11恐怖袭击事件看航空旅客运输承运人责任[J]. 法学, 10:67-74, 2002.
- [9] 杨国勇, 陈超, 高树林, 冯波. 基于层次分析-模糊聚类分析法的导水裂隙带发育高度研究[J]. 采矿与安全工程学报, 2015, 32(02):206-212.
- [10] 徐泽水. 区间直觉模糊集相似性测度及其在模式识别中的应用(英文)[J]. Journal of Southeast University, 2007(01):139-143.
- [11] 管清云等, 基于距离熵的应急决策层信息融合方法[J]. 系统工程理论与实践, 35(1): 216-227, 2015.

## 程序及代码：

### 1. 模糊聚类

```
clc;
clear;
close all;

%% Problem Definition

data = load('mydata');
X = data.X;
k = 5;

CostFunction=@(m) ClusteringCost(m, X);    % Cost Function

VarSize=[k size(X,2)]; % Decision Variables Matrix Size

nVar=prod(VarSize);    % Number of Decision Variables

VarMin= repmat(min(X),k,1);    % Lower Bound of Variables
VarMax= repmat(max(X),k,1);    % Upper Bound of Variables

%% GA Parameters

MaxIt=200;    % Maximum Number of Iterations

nPop=100;    % Population Size

pc=0.8;    % Crossover Percentage
nc=2*round(pc*nPop/2); % Number of Offsprings (Parnets)

pm=0.3;    % Mutation Percentage
nm=round(pm*nPop); % Number of Mutants

gamma=0.2;

mu=0.02;    % Mutation Rate

beta=8;    % Selection Pressure

%% Initialization

empty_individual.Position=[];
empty_individual.Cost=[];
```

```

empty_individual.Out=[];

pop= repmat(empty_individual,nPop,1);

for i=1:nPop

    % Initialize Position
    pop(i).Position=unifrnd(VarMin,VarMax,VarSize);

    % Evaluation
    [pop(i).Cost, pop(i).Out]=CostFunction(pop(i).Position);

end

% Sort Population
Costs=[pop.Cost];
[Costs, SortOrder]=sort(Costs);
pop=pop(SortOrder);

% Store Best Solution
BestSol=pop(1);

% Array to Hold Best Cost Values
BestCost=zeros(MaxIt,1);

% Store Cost
WorstCost=pop(end).Cost;

%% Main Loop

for it=1:MaxIt

    P=exp(-beta*Costs/WorstCost);
    P=P/sum(P);

    % Crossover
    popc=repmat(empty_individual,nc/2,2);
    for k=1:nc/2

        % Select Parents Indices
        i1=RouletteWheelSelection(P);
        i2=RouletteWheelSelection(P);
    end
end

```



```

    % Select Parents
    p1=pop(i1);
    p2=pop(i2);

    % Apply Crossover
    [popc(k,1).Position, popc(k,2).Position]=...
        Crossover(p1.Position,p2.Position,gamma,VarMin,VarMax);

    % Evaluate Offsprings
    [popc(k,1).Cost,
popc(k,1).Out]=CostFunction(popc(k,1).Position);
    [popc(k,2).Cost,
popc(k,2).Out]=CostFunction(popc(k,2).Position);

end
popc=popc(:);

% Mutation
popm= repmat(empty_individual,nm,1);
for k=1:nm

    % Select Parent
    i=randi([1 nPop]);
    p=pop(i);

    % Apply Mutation
    popm(k).Position=Mutate(p.Position,mu,VarMin,VarMax);

    % Evaluate Mutant
    [popm(k).Cost, popm(k).Out]=CostFunction(popm(k).Position);

end

% Create Merged Population
pop=[pop
      popc
      popm]; %#ok

% Sort Population
Costs=[pop.Cost];
[Costs, SortOrder]=sort(Costs);
pop=pop(SortOrder);

```

```

% Update Worst Cost
WorstCost=max(WorstCost,pop(end).Cost);

% Truncation
pop=pop(1:nPop);
Costs=Costs(1:nPop);

% Store Best Solution Ever Found
BestSol=pop(1);

% Store Best Cost Ever Found
BestCost(it)=BestSol.Cost;

% Show Iteration Information
disp(['Iteration ' num2str(it) ': Best Cost = '
num2str(BestCost(it))]);

% Plot Solution
figure(1);
PlotSolution(X, BestSol);
pause(0.01);

end

%% Results

figure;
plot(BestCost,'LineWidth',2);
xlabel('Iteration');
ylabel('Best Cost');
grid on;

```

## 2. 熵权法

```

[code]function [s,w]=shang(x)
% 函数shang.m, 实现用熵值法求各指标(列)的权重及各数据行的得分
% x为原始数据矩阵, 一行代表一个国家, 每列对应一个指标
% s返回各行得分, w返回各列权重
[n,m]=size(x); % n=23个国家, m=5个指标
%% 数据的归一化处理
% Matlab2010b,2011a,b版本都有bug,需如下处理. 其它版本直接用
[X,ps]=mapminmax(x',0,1);即可
[X,ps]=mapminmax(x');
ps.ymin=0.002; % 归一化后的最小值

```

```

ps.ymax=0.996; % 归一化后的最大值
ps.yrange=ps.ymax-ps.ymin; % 归一化后的极差,若不调整该值,则逆运算会出错
X=mapminmax(x',ps);
% mapminmax('reverse',xx,ps); % 反归一化,回到原数据
X=X'; % X为归一化后的数据, 23行(国家), 5列(指标)
%% 计算第j个指标下,第i个记录占该指标的比重p(i,j)
for i=1:n
    for j=1:m
        p(i,j)=X(i,j)/sum(X(:,j));
    end
end
%% 计算第j个指标的熵值e(j)
k=1/log(n);
for j=1:m
    e(j)=-k*sum(p(:,j).*log(p(:,j)));
end
d=ones(1,m)-e; % 计算信息熵冗余度
w=d./sum(d); % 求权值 w
s=w*p'; % 求综合得分 [\code]

```

### 3. 层次分析法

```

clc;
clear;
A = [...];
[m,n]=size(A); %获取指标个数
RI=[...];
R=rank(A); %求判断矩阵的秩
[V,D]=eig(A); %求判断矩阵的特征值和特征向量, V 特征值, D
特征向量;
tz=max(D);
B=max(tz); %最大特征值
[row, col]=find(D==B); %最大特征值所在位置
C=V(:,col); %对应特征向量
CI=(B-n)/(n-1); %计算一致性检验指标 CI
CR=CI/RI(1,n);
if CR<0.10
    disp('CI=');disp(CI);
    disp('CR=');disp(CR);
    disp('对比矩阵 A 通过一致性检验,各向量权重向量 Q 为: ');
    Q=zeros(n,1);
    for i=1:n
        Q(i,1)=C(i,1)/sum(C(:,1)); %特征向量标准化
    end
    Q %输出权重向量

```

```
else
    disp('对比矩阵 A 未通过一致性检验，需对对比矩阵 A 重新构造');
end
```