

基于DRF-KPCA和非线性面板模型的地区涉恐安全风险评估与影响因素研究

黄苏莉, 陆秋君

上海理工大学理学院, 上海
Email: 2503569275@qq.com

收稿日期: 2021年3月15日; 录用日期: 2021年4月3日; 发布日期: 2021年4月20日

摘 要

地区涉恐安全风险评估与影响因素研究能够研判地区未来的反恐态势, 提高反恐斗争的针对性和资源的管理效率。运用空间统计方法对全球恐怖袭击事件发生的分异特征进行分析, 选取恐怖袭击事件数、恐怖袭击死亡人数、恐怖袭击受伤人数和恐怖袭击财产损失程度等指标构建地区涉恐安全风险评估指标体系, 提出基于密度敏感稳健模糊核主成分分析的多维面板数据的分级模型。与其他的主成分聚类算法相比较, 所提算法的合理性和准确度都有提升。结合主成分方法提取出来的风险水平因子, 运用最大似然估计方法建立随机效应面板顺序Logit模型, 考察宏观经济和社会发展因素对地区涉恐安全风险的影响, 最后对回归模型展开预测绩效评价。

关键词

涉恐安全风险, 多维面板数据, 密度敏感稳健模糊核主成分分析, 顺序Logit回归

Regional Terrorism-Related Safety Risk Evaluation and Its Determinants Research Based on DRF-KPCA and Nonlinear Panel Model

Suli Huang, Qiujun Lu

College of Science, University of Shanghai for Science and Technology, Shanghai
Email: 2503569275@qq.com

Received: Mar. 15th, 2021; accepted: Apr. 3rd, 2021; published: Apr. 20th, 2021

Abstract

The research on regional terrorism-related safety risk evaluation and its determinants can judge the future situation of regional anti-terrorism, and improve the pertinence of anti-terrorism fight and efficiency of resource management. In this paper, the spatial statistical method is used to analyze the differentiation characteristics of global terrorist attacks. We construct the regional terrorism-related safety risk evaluation index system according to the variables such as number of terrorist attack, number of deaths, number of the injured and extent of property damage in terrorist attacks, and put forward a multi-dimensional panel data classification model based on density sensitive robust fuzzy kernel principal component analysis. Compared with other principal component clustering algorithms, the rationality and accuracy of the proposed algorithm are improved. The influence of macroeconomic and social development factors on regional terrorism-related safety risk is considered by presenting the random effect panel order Logit model based on maximum likelihood estimation method and the risk level factors obtained from principal component method. Finally, the prediction performance evaluation of the regression model is carried out.

Keywords

Terrorism-Related Safety Risk, Multi-Dimensional Panel Data, Density Sensitive Robust Fuzzy Kernel Principal Component Analysis, Order Logit Regression

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

作为一种影响巨大、变数多样的袭击活动,恐怖主义现象在国外管理科学界得到了高度关注,涌现了大量相关研究成果,研究范围包括理论研究、事例研究以及恐怖事件类型研究,其中跨学科介入恐怖主义研究的成果也比较多[1]。关于恐怖主义的研究主要集中于三个方面,一是对恐怖主义定义及其内涵的研究,二是从不同学科视角对恐怖主义进行研究,再有就是对不同恐怖袭击方式以及恐怖主义发生区域的研究。

定性研究是国内研究恐怖主义的主流方法,通常将恐怖主义与经济学、社会学、心理学或政治理论相结合,集中在文献分析法、历史研究法和比较研究法上。Ordu [2]探讨伊斯兰博科圣地和尼日尔三角洲武装分子在尼日利亚的趋势和模式,分析社会结构和政治经济对恐怖袭击的影响。李丽华等[3]以2017年英国发生的暴恐事件为研究对象,从舆情传播主体、信息特征以及传播特征角度得出预防和控制结论。刘乐[4]以伊斯兰国组织为代表的恐怖主义力量对于主流社会意义体系的侵蚀消解和冲击动摇为案例,给出国际社会重构反恐时代意义体系的根本战略。周秋君[5]追溯恐怖主义在当代欧洲的发展轨迹,对恐怖主义发展的新态势及其变化的原因展开分析。Oliveira等[6]采用内容分析方法研究巴塞罗那和坎布里尔斯的目的地营销组织对于恐怖袭击采取的行动。

风险评估和数学建模理论为恐怖主义量化分析或模型构建提供了方法,包括博弈论优化建模、复杂社会网络规划建模、可视化技术与计量建模等[7]。付举磊等[8]使用社会网络理论构建“事件-时间-空

间”关系网络模型, 结合地理信息系统技术, 分析东突恐怖活动的时空特征。王雷[9]等提出基于和声搜索算法优化的支持向量机的分级模型, 对我国 2008~2013 年的突发暴恐事件进行分级研究。Estrada 等[10]提出恐怖行动的三个阶段, 引入风险指标估计可能影响区域经济绩效的潜在恐怖事件因素。郭璇等[11]提出基于事件树和概率风险分析的民航机场恐怖袭击风险评估模型, 以降低民航机场恐怖袭击事件发生的概率和风险。Buscema 等[12]使用基于拓扑的方法对受重大系统性恐怖主义活动影响并具有全球地缘政治影响的国家进行测试, 预测恐怖袭击的空间模式。Lanouar 等[13]使用马尔可夫机制转换模型研究恐怖袭击和政治暴力对突尼斯入境游客和过夜人数的影响。何晶等[14]提出一种基于多局域的恐怖组织网络择优增长演化模型以模拟恐怖组织网络的动态演化规律。宋艳等[15]研究了恐怖袭击事件下, 政府和极端组织作为参与者基于现存设施布局的博弈选址问题。

以上工作的研究主要是特定场景下恐怖袭击事件的分析或风险评估, 以及恐怖组织的行为分析和预测等问题。关于恐怖袭击风险问题, 从研究背景来看, 主要集中在事件信息的危害性评估和分级上, 较少有基于面板数据框架的地区涉恐安全风险的研究。本文依据全球恐怖主义数据库(GTD), 运用空间统计方法对全球恐怖袭击事件发生的分异特征进行分析。从恐怖袭击危害性的角度, 选取恐怖袭击事件数、恐怖袭击死亡人数、恐怖袭击受伤人数和恐怖袭击财产损害程度等指标构建地区涉恐安全风险评估指标体系, 提出基于密度敏感稳健模糊核主成分分析(DRF-KPCA)的多维面板数据的分级模型。结合主成分方法提取出来的风险水平因子, 运用最大似然估计方法建立随机效应面板顺序 Logit 模型, 考察宏观经济和社会发展因素对地区涉恐安全风险的影响, 并对回归模型进行预测绩效评价。

本文余下部分的结构安排如下: 第 2 节给出地区涉恐安全风险评估与影响因素研究的方法; 第 3 节是地区概况; 第 4 节的实证研究中给出基于 DRF-KPCA 技术提取的地区涉恐安全风险分级结果, 并建立非线性面板数学模型进行影响因素研究; 第 5 节是结论。

2. 研究方法

2.1. 密度敏感稳健模糊核主成分分析算法

密度敏感稳健模糊核主成分分析(DRF-KPCA)算法扩展自 RKF-PCA 算法, 通过引入相对密度获取初始隶属度, 以及优化重构误差算法自适应地对样本进行隶属度加权, 同时利用核方法对数据中非线性信息提取的优势, 得到更有效率、更稳健地特征提取[16]。

设样本集 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\} \subset \mathbf{R}^n$, 核函数通过变换 Φ 将样本 \mathbf{x}_i 映射到特征空间中的 $\Phi(\mathbf{x}_i)$, 赋予隶属度 μ_i 后可得空间中的协方差矩阵 \mathbf{C} 为

$$\mathbf{C} = \frac{1}{\rho} \sum_{i=1}^m (\mu_i^{p/2} \Psi(\mathbf{x}_i)) (\mu_i^{p/2} \Psi(\mathbf{x}_i))', \quad (1)$$

其中 p 是模糊度系数, $\rho = \sum_{i=1}^m \mu_i^p$, $\Psi(\mathbf{x}_i) = \Phi(\mathbf{x}_i) - \frac{1}{\rho} \sum_{k=1}^m \mu_k^p \Phi(\mathbf{x}_k)$ 。

令 \mathbf{v} 是 \mathbf{C} 属于特征值 λ 的特征向量, 即有

$$\mathbf{C}\mathbf{v} = \lambda\mathbf{v}. \quad (2)$$

通常情况 $\Psi(\mathbf{x}_i)\Psi(\mathbf{x}_j)'$ 难以直接通过计算获取, 可定义一个 $m \times m$ 的矩阵 $\bar{\mathbf{K}}$, 如式(3)所示

$$\bar{\mathbf{K}} = [\bar{K}_{ij}], \quad (3)$$

其中 $\bar{K}_{ij} = (\mu_i^{p/2} \Psi(\mathbf{x}_i))' (\mu_j^{p/2} \Psi(\mathbf{x}_j))$, $i, j = 1, 2, \dots, m$ 。综合式(1)和(2)得

$$\bar{\mathbf{K}}\bar{\boldsymbol{\alpha}} = \rho\lambda\bar{\boldsymbol{\alpha}}, \quad (4)$$

$$\mathbf{v} = \sum_{i=1}^m a_i \mu_i^{p/2} \Psi(\mathbf{x}_i), \quad (5)$$

这里 $\bar{\mathbf{a}} = (a_1, a_2, \dots, a_m)'$, $a_i = \frac{1}{\lambda \rho} (\mu_i^{p/2} \Psi(\mathbf{x}_i))' \mathbf{v}$, $i = 1, 2, \dots, m$ 。

解式(3)、(4)和(5)就能得到特征值和特征向量, 测试样本在特征空间中的投影为

$$(\mathbf{v}, \Psi(\mathbf{x}_j)) = \sum_{i=1}^m a_i \mu_i^{p/2} (\Psi(\mathbf{x}_i))' (\Psi(\mathbf{x}_j)), \quad j = 1, 2, \dots, m. \quad (6)$$

在重构误差意义上, DRF-KPCA 可以表示为以下模型

$$\min \sum_{i=1}^m \left(\mu_i \|\Psi(\mathbf{x}_i) - \mathbf{V}\mathbf{V}'\Psi(\mathbf{x}_i)\|^2 + \sigma^2 (\mu_i \log \mu_i - \mu_i) \right), \quad \text{s.t. } \mathbf{V}\mathbf{V}' = \mathbf{I}, \quad (7)$$

其中 \mathbf{V} 为由前 q 个最大特征值对应的特征向量构成的投影矩阵, $\mathbf{I} \in \mathbb{R}^{q \times q}$ 为单位矩阵, σ^2 为正则化控制参数。

通过优化理论中的拉格朗日方程, 隶属度 μ_i 有式(8)的解析形式

$$\mu_i = \exp\left(-\frac{1}{\sigma^2} \|\Psi(\mathbf{x}_i) - \mathbf{V}\mathbf{V}'\Psi(\mathbf{x}_i)\|^2\right), \quad i = 1, 2, \dots, m. \quad (8)$$

分析式(7)和(8)可以发现, 对重构误差较大的样本, 其隶属度随着重构误差增大而取值越小, 而重构误差较小的样本其隶属度取值较大。

为解决 PCA 算法对初始值敏感的问题, 文献[16]引入相对密度获取样本初始隶属度, 训练样本 \mathbf{x}_i 的相对密度 d_i 定义如下

$$d_i = \exp\left\{\omega \times \frac{\text{Par}(\mathbf{x}_i)}{\varsigma}\right\}, \quad i = 1, 2, \dots, m, \quad (9)$$

其中 $\text{Par}(\mathbf{x}_i) = \frac{1}{m} \sum_{j=1}^m \left(\frac{1}{\sqrt{(2\pi)^n s}} \exp\left(-\frac{1}{2s} \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \right)$, $\varsigma = \frac{1}{m} \sum_{i=1}^m \text{Par}(\mathbf{x}_i)$, ω 为权重, s 是 Parzen-window 的平滑参数。

2.2. 基于 DRF-KPCA 的多维面板数据的聚类分级

为保证重构空间准确地描述系统状态的演化轨迹, 通常会选择多维面板数据为研究对象。当多个指标变量所代表的状态间相关性较强时, 需要先对面板数据进行降维后才能进行聚类分析[17]。

假定有 m 个研究对象组成聚类集合 $\{\mathbf{X}(1), \mathbf{X}(2), \dots, \mathbf{X}(T)\}$, 其中矩阵 $\mathbf{X}(t) = [x_{ij}(t)]$, 这里 $x_{ij}(t)$ 表示第 i 个研究对象在第 t 个时间内的第 j 个指标数值 ($i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, $t = 1, 2, \dots, T$)。通常, 不同指标数值之间差异明显, 于是令

$$x_{ij}^*(t) = \frac{x_{ij}(t) - \bar{x}_{.j}(t)}{\sqrt{\text{var}_{.j}(t)}}, \quad i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n, \quad t = 1, 2, \dots, T, \quad (10)$$

其中 $\bar{x}_{.j}(t) = \frac{1}{m} \sum_{i=1}^m x_{ij}(t)$, $\text{var}_{.j}(t) = \frac{1}{m-1} \sum_{i=1}^m (x_{ij}(t) - \bar{x}_{.j}(t))^2$ 。即用标准化数据代替原始尺度, 以消除各指标间量纲的影响, 则第 t 个时间内研究对象的指标矩阵为

$$\mathbf{X}^*(t) = \begin{bmatrix} x_{11}^*(t) & x_{12}^*(t) & \cdots & x_{1n}^*(t) \\ x_{21}^*(t) & x_{22}^*(t) & \cdots & x_{2n}^*(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}^*(t) & x_{m2}^*(t) & \cdots & x_{mn}^*(t) \end{bmatrix}. \quad (11)$$

通过运行 DRF-KPCA 算法获得 $\mathbf{X}^*(t)$ 在低维上的映射

$$\mathbf{Y}(t) = \begin{bmatrix} y_{11}(t) & y_{12}(t) & \cdots & y_{1z}(t) \\ y_{21}(t) & y_{22}(t) & \cdots & y_{2z}(t) \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1}(t) & y_{m2}(t) & \cdots & y_{mz}(t) \end{bmatrix}. \quad (12)$$

这里对每个截面矩阵取 z 个指标主成分, 使得累计方差贡献率达到 85% 以上。根据多维面板数据的聚类思想, 计算各个时间截面的样本综合评价函数, 如下所示

$$F_i(t) = w_1(t)y_{i1}(t) + w_2(t)y_{i2}(t) + \cdots + w_z(t)y_{iz}(t), \quad (13)$$

其中 $F_i(t)$ 为第 i 个研究对象在第 t 个时间内的综合评价函数, 指标主成分贡献率为

$$w_j(t) = \frac{\lambda_j(t)}{\sum_{i=1}^z \lambda_i(t)}, \quad j = 1, 2, \dots, z, \quad t = 1, 2, \dots, T, \quad (14)$$

这里 $\lambda_j(t)$ 为协方差矩阵的特征值, 并且将其排列, 即 $\lambda_1(t) \geq \lambda_2(t) \geq \cdots \geq \lambda_z(t)$ 。

最后构建综合评价函数序列矩阵

$$\mathbf{F}_i = (F_i(1), F_i(2), \dots, F_i(T)), \quad i = 1, 2, \dots, m. \quad (15)$$

选择欧式距离 $d(\mathbf{F}_i, \mathbf{F}_j) = \sqrt{\sum_{t=1}^T (F_i(t) - F_j(t))^2}$ 来表示第 i 个对象与第 j 个对象之间的差异, 从而对面板数据进行层次聚类, 基于 DRF-KPCA 的多维面板数据的聚类算法流程见表 1。

Table 1. Algorithm flow of “DRF-KPCA + Hierarchical Clustering”

表 1. “DRF-KPCA + 层次聚类” 算法流程

<p>输入: 成员集合 $\{\mathbf{X}(1), \mathbf{X}(2), \dots, \mathbf{X}(T)\} \subset \mathbf{R}^{m \times n}$, 核函数 K 及其参数, 模糊度系数 p, 保留维度 q, 正则化控制参数 σ^2, Parzen-window 的平滑参数 s, 权重 ω, 隶属度阈值 η 和迭代的最大次数 k_{\max}。</p> <p>输出: 聚类树状图及聚类结果。</p> <p>步骤 1 根据式(10)计算标准化矩阵 $\mathbf{X}^*(t)$, $t = 1, 2, \dots, T$;</p> <p>步骤 2 对 $\mathbf{X}^*(t)$ 运行 DRF-KPCA 算法得到 $\mathbf{Y}(t)$, $t = 1, 2, \dots, T$;</p> <p>步骤 2.1 令 $k = 0$, 通过式(9)求出相对密度 d_i, 赋值给初始隶属度 $U^0 = \{d_1, d_2, \dots, d_m\}$;</p> <p>步骤 2.2 根据式(3)、(4)和(5)求出特征值 $\lambda_1^k, \lambda_2^k, \dots, \lambda_q^k$ 和对应的特征向量 $\mathbf{v}_1^k, \mathbf{v}_2^k, \dots, \mathbf{v}_q^k$;</p> <p>步骤 2.3 令 $k = k + 1$, 根据式(7)和(8), 计算重构误差和新的隶属度 $U^k = \{\mu_1^k, \mu_2^k, \dots, \mu_m^k\}$;</p> <p>步骤 2.4 如果满足收敛条件 $\max_{1 \leq i \leq m} \mu_i^{k-1} - \mu_i^k < \eta$ 或 $k > k_{\max}$, 则迭代结束, 否则转向步骤 2.2;</p> <p>步骤 2.5 根据式(6)得到研究变量在低维空间中的表示。</p> <p>步骤 3 根据式(13)、(14)和(15)计算综合评价函数序列矩阵 \mathbf{F}_i, $i = 1, 2, \dots, m$;</p> <p>步骤 4 选择矩阵间的距离公式, 对 m 个研究对象进行聚类分级。</p>

2.3. 基于 DRF-KPCA 的面板顺序 Logit 模型

由基于 DRF-KPCA 的多维面板数据的聚类分级, 我们得到关于研究对象的定性分类结果, 如果我们还关心影响分类结果的因素贡献, 可以引入离散变量构建面板多值选择模型。Logit 回归是一种分类模型,

采用 Sigmoid 函数作为后验概率分布函数对数据集进行分类。面板 Logit 模型由于同时使用截面数据和时序序列数据, 既可以控制个体异质性, 也可以对风险进行预警, 从模型中分析影响风险的因素[18]。

假设风险状态被设定为一个不可观测的潜在变量 y_{it}^* , 而 y_{it} 为可观察到的风险的分类状态, r_1, r_2, \dots, r_J 为每一风险状态下的阈值。因此, 分类状态 y_{it} 可以看作潜在变量 y_{it}^* 和阈值 r 共同生成的离散变量, 构建的顺序 Logit 模型如下

$$y_{it} = \begin{cases} 0, & y_{it}^* \leq r_1, \\ 1, & r_1 < y_{it}^* \leq r_2, \\ 2, & r_2 < y_{it}^* \leq r_3, \\ \vdots & \vdots \\ J, & y_{it}^* \geq r_J, \end{cases} \quad (16)$$

$$y_{it}^* = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + \varepsilon_{it}, \quad (17)$$

其中 $i=1, 2, \dots, m$, $t=1, 2, \dots, T$, $r_1 < r_2 < \dots < r_J$, \mathbf{x}_{it} 是用来描述第 i 个研究对象在第 t 个时间内可观测特征的解解释向量, $\boldsymbol{\beta}$ 是回归系数。 u_i 为不随时间改变及不可观测且对潜在变量有影响的个体效应, 用于刻画不同对象的个体异质性。 ε_{it} 为随机干扰项, 服从 Logistic 分布, 其累积分布函数为

$$F(\varphi) = \frac{\exp(\varphi)}{1 + \exp(\varphi)}. \quad (18)$$

假设 u_i 的密度函数为 $g(u_i)$, 第 i 个对象的条件分布为

$$f(y_{i1}, y_{i2}, \dots, y_{iT} | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \prod_{t=1}^T \prod_{j=0}^J \pi_{ij}^{\tau_{ij}}, \quad (19)$$

其中

$$\pi_{i0} = P(y_{it} = 0 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \frac{\exp(r_1 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)}{1 + \exp(r_1 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)}, \quad (20)$$

$$\pi_{i1} = P(y_{it} = 1 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \frac{\exp(r_2 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i) - \exp(r_1 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)}{[1 + \exp(r_2 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)][1 + \exp(r_1 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)]}, \quad (21)$$

$$\pi_{i2} = P(y_{it} = 2 | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \frac{\exp(r_3 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i) - \exp(r_2 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)}{[1 + \exp(r_3 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)][1 + \exp(r_2 - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)]}, \quad (22)$$

⋮

$$\pi_{iJ} = P(y_{it} = J | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) = \frac{1}{1 + \exp(r_J - \mathbf{x}_{it}'\boldsymbol{\beta} - u_i)}, \quad (23)$$

$$\tau_{ij} = \begin{cases} 1, & y_{it} = j, \\ 0, & \text{其它}. \end{cases} \quad (24)$$

这里 $i=1, 2, \dots, m$, $t=1, 2, \dots, T$, $j=0, 1, \dots, J$ 。于是, $(y_{i1}, y_{i2}, \dots, y_{iT})$ 的概率密度

$$f(y_{i1}, y_{i2}, \dots, y_{iT}) = \int_{-\infty}^{\infty} f(y_{i1}, y_{i2}, \dots, y_{iT} | \mathbf{x}_{it}, \boldsymbol{\beta}, u_i) g(u_i) du_i. \quad (25)$$

假设不同研究对象的观测值相互独立, 则构建顺序 Logit 模型的对数似然函数为

$$\log L(\boldsymbol{\beta}) = \sum_{i=1}^m \log f(y_{i1}, y_{i2}, \dots, y_{iT}). \quad (26)$$

施加概率和为 1 的约束 $\sum_{j=0}^J \pi_{ij} = 1 (i = 1, 2, \dots, m, t = 1, 2, \dots, T)$, 最大化此似然函数, 即得到对 β 的随机效应 Logit 估计量。上述积分没有解析解, 可以使用 Gauss-Hermite quadrature 方法进行数值积分。

3. 地区概况

全球各地区¹恐怖活动随时间变化趋势明显, 呈现先起后伏状态, 多在 2014 年出现峰值, 如图 1 所示。南亚、中东和北非是恐怖袭击频发地区, 发生数约占全球的 70%。2004~2014 年间, 各地区恐怖袭击事件发生数差距开始逐渐拉大, 撒哈拉以南的非洲年均涨幅为 51%, 中美洲和加勒比海区、中亚以及澳大利亚和大洋洲年均涨幅不足 5%。近年来, 随着各国提升对恐怖活动的应对能力与打击力度, 恐怖活动及伤亡人数逐渐减少, 2014~2018 年中东和北非恐怖袭击事件发生数呈现逐年下降的趋势, 年均降幅为 19%, 伤亡人数年均降幅为 30%。

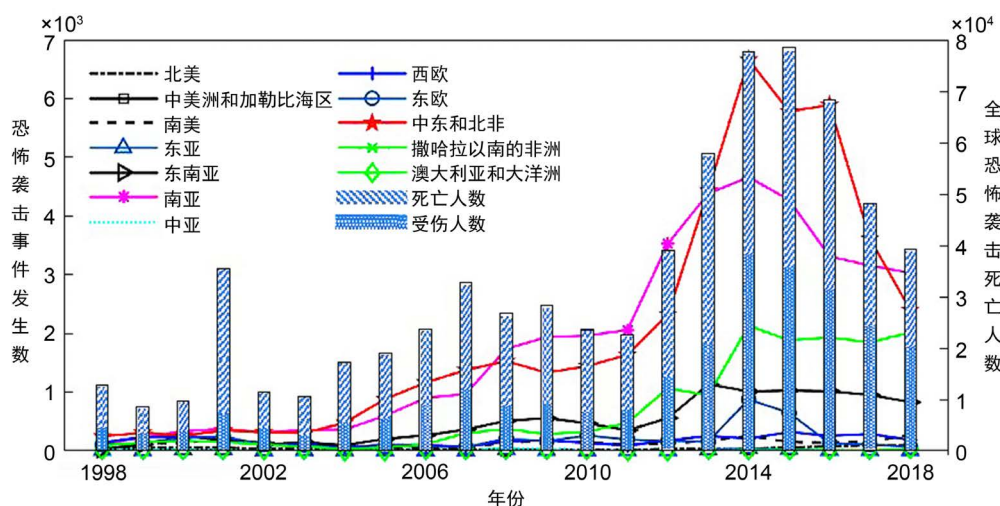


Figure 1. Statistics of terrorist attacks from 1998 to 2018
图 1. 1998~2018 年恐怖袭击事件统计

本文运用 ArcGIS 10.5 中的核密度分析工具, 对 1998~2007 年和 2008~2018 年两个不同时期恐怖袭击事件发生的空间分布特征进行可视化处理, 将核密度值按 ArcGIS 自然分级为低值区、较低值区、中值区、较高值区和高值区五类, 值区代表颜色分别为白色、绿色、黄色、橙色和红色, 见图 2。

从核密度分布图看, 全球恐怖袭击事件发生的总体空间分布格局没有明显变动的经历, 主要分布在中东和北非、南亚和撒哈拉以南的非洲。1998~2007 年较高值区和高值区集中在中东和北非、南亚, 完全处于低值分布的地区有东亚、中亚以及澳大利亚和大洋洲, 如图 2(a)所示。2008~2018 年核密度值出现不同幅度的上升, 已没有完全处于低值分布的地区, 处于低值和较低值分布的地区有中美洲和加勒比海区、东亚、中亚以及澳大利亚和大洋洲, 中值、较高值和高值区数量增加, 东南亚和东欧逐渐出现高值与较高值区域, 见图 2(b)。

其次, 基于 ArcGIS10.5, 利用标准差椭圆法定量分析 1998~2018 年恐怖袭击事件发生的总体轮廓和其所主导的方向, 地图可视化结果见图 3(a), 参数结果见图 3(b)。

1998~2006 年标准差椭圆重心表现为东西方向扩张强度明显大于南北方向, 总位移约为 3417 km, 其

¹ 本文采用的地区数据来自 GTD 全球恐怖主义数据库, 涵盖了 12 类地区, 1 = 北美, 2 = 中美洲和加勒比海区, 3 = 南美, 4 = 东亚, 5 = 东南亚, 6 = 南亚, 7 = 中亚, 8 = 西欧, 9 = 东欧, 10 = 中东和北非, 11 = 撒哈拉以南的非洲, 12 = 澳大利亚和大洋洲。

中向东移动约 3439 km, 向南移动约 111 km。在这一阶段, 椭圆方位角为向北偏西转移约 6°, 长轴与短轴均减小, 椭圆总体分布范围在缩小。2007~2018 年标准差椭圆重心的空间移动范围较小, 总位移约为 665 km, 表明恐怖袭击活动空间聚集性较强, 椭圆方位角表现为向北偏东转移约 10°, 长轴与短轴均增加, 恐怖袭击空间分布存在逐渐增大的趋势。

1998~2018 年全球恐怖袭击事件空间格局可概括为先向北偏西转移, 后向北偏东转移, 整体呈现向北偏东移动的格局, 且恐怖袭击事件发生的空间分布呈现先缩小后逐渐扩大趋势, 中东和北非以及南亚的大部分国家一直处在椭圆内部。

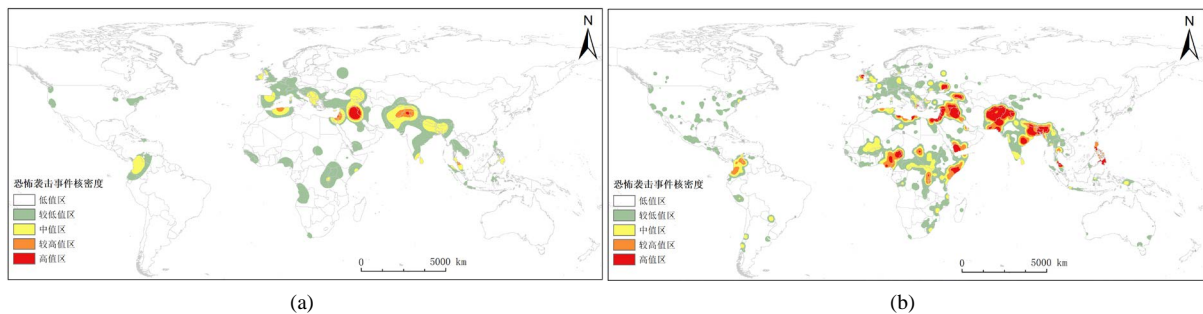


Figure 2. Nuclear density distribution of terrorist attacks from 1998 to 2018. (a) 1998~2007 year; (b) 2008~2018 year
图 2. 1998~2018 年恐怖袭击事件的核密度分布。(a) 1998~2007 年; (b) 2008~2018 年

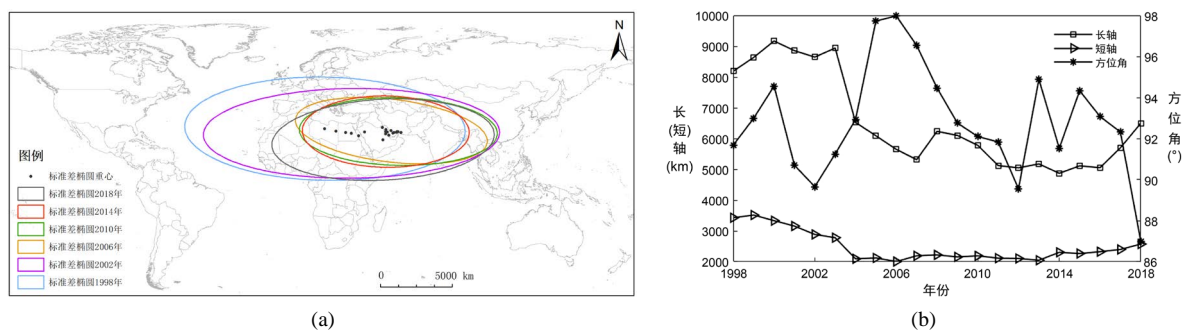


Figure 3. Standard deviation ellipse spatial distribution of terrorist attacks from 1998 to 2018. (a) Standard deviation ellipse and center of gravity shift; (b) Long and short axis and azimuth change
图 3. 1998~2018 年恐怖袭击事件的标准差椭圆空间分布。(a) 标准差椭圆和重心转移; (b) 长短轴和方位角变化

4. 地区涉恐安全风险评估与影响因素研究

4.1. 地区涉恐安全风险评估

地区涉恐安全风险水平的确定主要依赖对恐怖袭击事件破坏后果的统计, 在参考先前的研究[1] [9] [19] [20] [21] [22] [23], 考虑人员死亡和经济损失等与恐怖袭击危害性紧密相关的主要变量后, 构建地区涉恐安全风险评估指标体系。体系分为二层, 目标层为地区涉恐安全风险评估, 指标层包含恐怖袭击事件发生数、恐怖袭击死亡人数、恐怖袭击受伤人数、财产损害程度²、凶手数量、凶手死亡人数、凶手受伤人数、人质/绑架受害者人数和亡人事件百分比³。

²财产损害程度是分类变量, 本文中地区财产损害程度的计算方法为: 首先对财产损害程度变量取负号, 作正向化处理, 然后对每个损害程度类别, 计算恐怖袭击事件发生数占地区恐怖袭击事件总数的百分比, 最后在此基础上, 以比值为权重计算地区财产损害程度。

³地区亡人事件百分比是以“存在死亡人员的恐怖袭击事件数/全部恐怖袭击事件数”的比值度量, 是衡量恐怖袭击活动危害性的一个重要指标。

本文聚类采用 GTD 中 1998~2018 年 12 个地区的恐怖袭击事件样本数据, 剔除变量数据不完善的事件后, 共有样本事件 118,084 条。通过对数据进行标准化处理后, 获得统一形式的数据表示, 随后对数据采用 DRF-KPCA 算法进行降维处理, 获得关键特征。

为了比较 DRF-KPCA 与其他主成分算法[24] [25]提取数据特征的能力, 给出 2018 年相同数目的主成分所能解释的累计方差百分比, 如表 2 所示。从表 2 中可以看出数目相同的非线性主成分比线性主成分能够获取更多的累计方差百分比, 即能获取更多数据信息的同时, 抑制离群点所带来的干扰。

Table 2. The cumulative percentage variance explained by the same number of principal components (2018)

表 2. 相同数目的主成分所能解释的累计方差百分比(2018 年)⁴

主成分数目	累计方差百分比(%)			
	PCA	KPCA	RKF-KPCA	DRF-KPCA
1	49.56	75.68	83.30	95.20
2	78.73	86.64	96.97	98.77
3	94.69	94.38	98.77	99.99
4	98.44	98.65	99.79	100

表 3 为 1998~2018 年 12 个地区涉恐安全风险水平的排序结果, 地区排序号越大, 即地区涉恐安全风险水平越高, 意味着恐怖主义对地区安全的影响越大, 地区安全态势越动荡; 反之地区排序号越小, 恐怖主义对地区安全的影响较小, 则地区处于相对安全的态势水平。1998~2018 年澳大利亚和大洋洲、西欧、北美以及中美洲和加勒比海区排序位置靠前, 南亚、中东和北非以及撒哈拉以南的非洲排序位置在后, 东南亚、南亚、中东和北非以及撒哈拉以南的非洲排序结果总体上比较稳定, 东亚、东欧以及中美洲和加勒比海区排序结果的变化相对明显。

Table 3. Regional ranking results of terrorism-related safety risk level

表 3. 涉恐安全风险水平地区排序结果

序号	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1	地区 2	地区 12	地区 12	地区 12	地区 2	地区 1	地区 4	地区 12	地区 12	地区 4	地区 12
2	地区 12	地区 1	地区 1	地区 4	地区 4	地区 7	地区 7	地区 1	地区 4	地区 12	地区 2
3	地区 4	地区 2	地区 4	地区 2	地区 1	地区 2	地区 1	地区 2	地区 8	地区 7	地区 8
4	地区 1	地区 4	地区 8	地区 8	地区 12	地区 4	地区 2	地区 7	地区 1	地区 8	地区 7
5	地区 8	地区 7	地区 3	地区 7	地区 8	地区 8	地区 12	地区 8	地区 7	地区 1	地区 1
6	地区 9	地区 8	地区 2	地区 1	地区 7	地区 12	地区 3	地区 4	地区 9	地区 2	地区 3
7	地区 5	地区 5	地区 7	地区 9	地区 3	地区 3	地区 8	地区 9	地区 2	地区 9	地区 9
8	地区 3	地区 3	地区 9	地区 5	地区 5	地区 9	地区 5	地区 5	地区 3	地区 3	地区 4
9	地区 7	地区 9	地区 5	地区 11	地区 9	地区 11	地区 11	地区 3	地区 5	地区 5	地区 5
10	地区 6	地区 10	地区 11	地区 3	地区 11	地区 5	地区 9	地区 11	地区 11	地区 11	地区 11
11	地区 10	地区 11	地区 10	地区 10	地区 6	地区 6	地区 10	地区 6	地区 6	地区 10	地区 10
12	地区 11	地区 6	地区 6	地区 6	地区 10	地区 10	地区 6	地区 10	地区 10	地区 6	地区 6

⁴文中给出 2018 年相同数目的主成分所能解释的累计方差百分比, 实际上, 对于其他年份也可以进行类似比较和解释。

Continued

序号	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	地区 12	地区 2	地区 12	地区 1	地区 12	地区 12	地区 12	地区 9	地区 4	地区 12
2	地区 7	地区 12	地区 1	地区 8	地区 2	地区 8	地区 8	地区 3	地区 8	地区 8
3	地区 8	地区 1	地区 7	地区 2	地区 8	地区 2	地区 1	地区 8	地区 9	地区 2
4	地区 2	地区 8	地区 4	地区 7	地区 7	地区 3	地区 3	地区 12	地区 1	地区 9
5	地区 4	地区 7	地区 8	地区 12	地区 1	地区 7	地区 9	地区 1	地区 12	地区 1
6	地区 3	地区 3	地区 2	地区 9	地区 4	地区 1	地区 4	地区 4	地区 7	地区 3
7	地区 9	地区 4	地区 3	地区 3	地区 9	地区 4	地区 2	地区 7	地区 3	地区 4
8	地区 1	地区 9	地区 9	地区 4	地区 3	地区 5	地区 7	地区 2	地区 2	地区 7
9	地区 5	地区 5	地区 5	地区 5	地区 5	地区 9	地区 5	地区 5	地区 5	地区 5
10	地区 11	地区 11	地区 11	地区 11	地区 11	地区 11	地区 11	地区 11	地区 11	地区 10
11	地区 10	地区 10	地区 10	地区 10	地区 10	地区 6	地区 10	地区 6	地区 6	地区 11
12	地区 6	地区 6	地区 6	地区 6	地区 6	地区 10	地区 6	地区 10	地区 10	地区 6

为验证本文构建的“DRF-KPCA + 层次聚类”集成方法对于处理地区涉恐安全风险数据的聚类问题的有效性和合理性, 本文将直接使用层次聚类和先使用传统 PCA 降维再进行聚类(记为“PCA + 层次聚类”)两种方法作为对照, 分析聚类结果⁵。为了更好地表示分级结果, 本文将数据集分为 1998-2004 年、2005~2011 年和 2012~2018 年三段, 得到地区涉恐安全风险的五个等级, 分别为低风险、较低风险、中风险、较高风险和高风险, 具体见图 4、图 5 和图 6。

从这些图可以看出, 相较于层次聚类和“PCA + 层次聚类”方法, 使用“DRF-KPCA + 层次聚类”方法得到的聚类结果具有明显的层次性和差异性, 且与实际情况比较符合, 一定程度上说明本文构建的

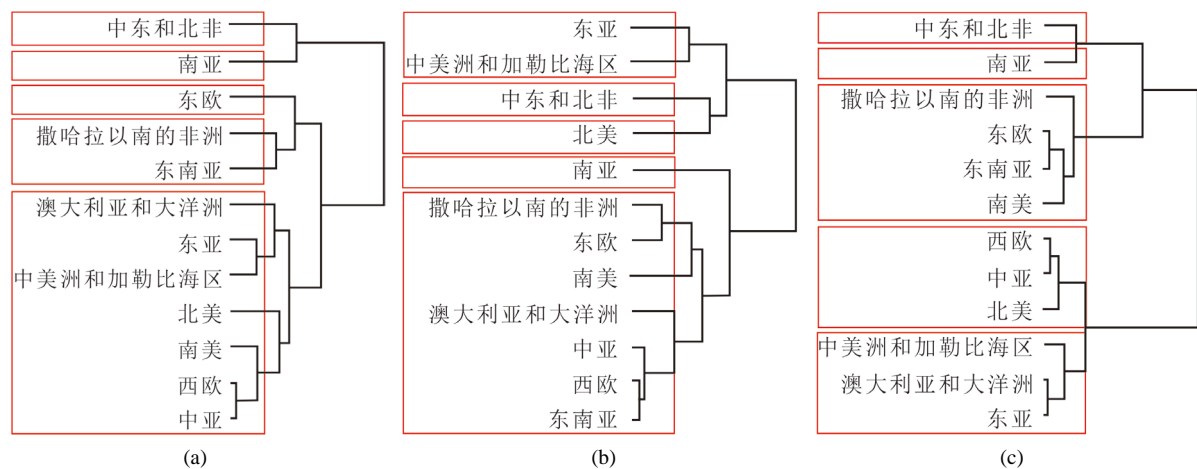


Figure 4. Comparison of three clustering structures from 1998 to 2004. (a) Hierarchical clustering; (b) PCA + hierarchical clustering; (c) DRF-KPCA + hierarchical clustering

图 4. 1998~2004 年三种聚类结构对比。(a) 层次聚类; (b) PCA + 层次聚类; (c) DRF-KPCA + 层次聚类

⁵ “KPCA+层次聚类”和“RKF-PCA + 层次聚类”有与“DRF-KPCA + 层次聚类”类似的结果, 考虑到算法的运行效率和篇幅所限, 本文不列出“KPCA + 层次聚类”和“RKF-PCA + 层次聚类”的结构图。

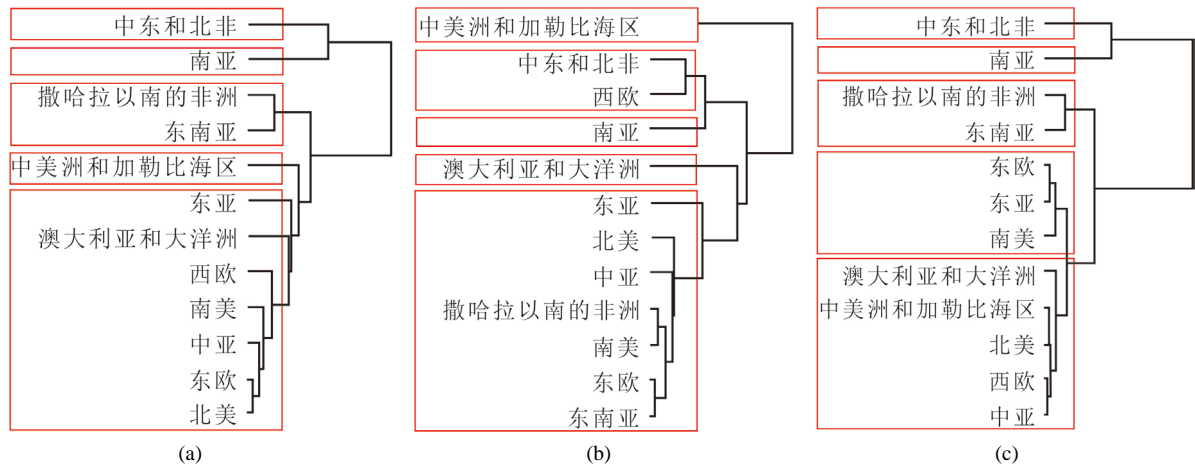


Figure 5. Comparison of three clustering structures from 2005 to 2011. (a) Hierarchical clustering; (b) PCA + hierarchical clustering; (c) DRF-KPCA + hierarchical clustering

图 5. 2005~2011 年三种聚类结构对比。(a) 层次聚类; (b) PCA + 层次聚类; (c) DRF-KPCA + 层次聚类

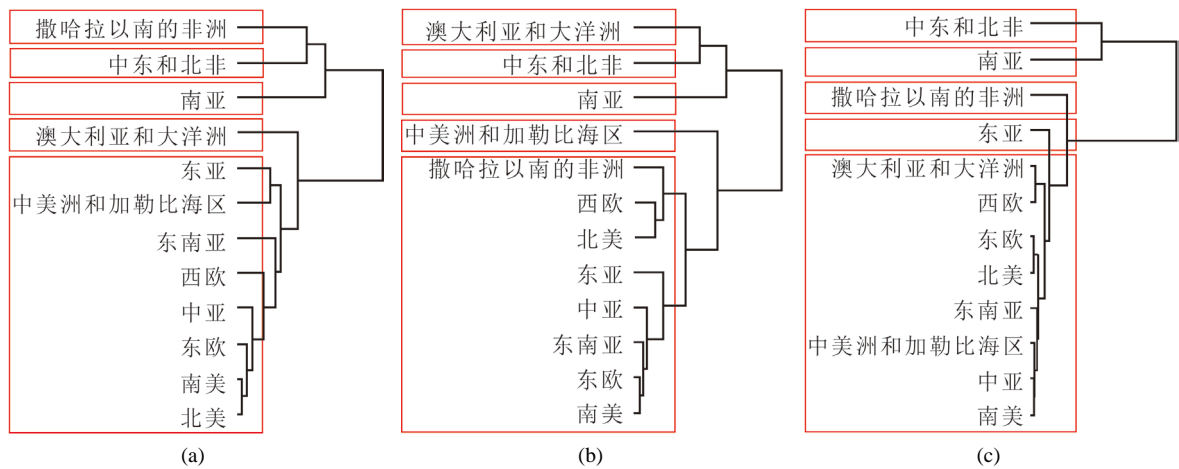


Figure 6. Comparison of three clustering structures from 2012 to 2018. (a) Hierarchical clustering; (b) PCA + hierarchical clustering; (c) DRF-KPCA + hierarchical clustering

图 6. 2012~2018 年三种聚类结构对比。(a) 层次聚类; (b) PCA + 层次聚类; (c) DRF-KPCA + 层次聚类

Table 4. Regional terrorism-related safety risk classification results under “DRF-KPCA + hierarchical clustering”

表 4. “DRF-KPCA + 层次聚类” 下地区涉恐安全风险分级结果

地区	1998~2004 年	2005~2011 年	2012~2018 年
北美	较低风险	低风险	低风险
中美洲和加勒比海区	低风险	低风险	低风险
南美	中风险	较低风险	低风险
东亚	低风险	较低风险	较低风险
东南亚	中风险	中风险	低风险
南亚	高风险	较高风险	较高风险
中亚	较低风险	低风险	低风险
西欧	较低风险	低风险	低风险

Continued

东欧	中风险	较低风险	低风险
中东和北非	较高风险	高风险	高风险
撒哈拉以南的非洲	中风险	中风险	中风险
澳大利亚和大洋洲	低风险	低风险	低风险

“DRF-KPCA + 层次聚类”方法在处理地区涉恐安全风险多维面板数据的聚类分级上显示出更准确合理的性能, 表 4 报告了这三个时间段的地区风险分级结果。1998~2018 年涉恐安全风险等级没有变化的地区是中美洲和加勒比海区(低风险)、澳大利亚和大洋洲(低风险)以及撒哈拉以南的非洲(中风险)。在这一阶段, 南美、东欧和东南亚安全风险等级显著下降, 从中风险变为低风险, 北美、中亚和西欧从较低风险变为低风险, 南亚从高风险变为较高风险, 这说明安全新秩序的建构和反恐的团结合作使得一些地区的涉恐安全风险程度下降。结合表 4 可知, 1998~2018 年不存在涉恐安全风险表现为显著增长的地区, 东亚从低风险变为较低风险, 中东和北非从较高风险变为高风险, 这说明对地区安全构成威胁的因素依然可能是不稳定的政情、资源的争夺和局部武装冲突的不断。

4.2. 地区涉恐安全风险的影响因素分析

本文选取“DRF-KPCA + 层次聚类”分级结果作为被解释变量, 关于地区涉恐安全风险水平的主要影响因素, 结合《国际安全态势感知指数 2017》报告和数据的可得性, 确定人均国内生产总值、资产形成总额占比、军费支出占比、政府消费最终支出占比、航空运输货运量、人口增长率以及货物和服务出口占比作为地区涉恐安全风险的影响指标。这些变量的数据来源于世界银行的官方网站, 回归模型样本时间区间是 2009~2018 年。

基于 DRF-KPCA 的多维面板数据分级模型, 我们得到地区的涉恐安全风险水平(0 = 低风险、1 = 中风险、2 = 高风险)⁶。Mundlak [26]指出, 一般情况下都应该将个体效应视为随机的, 试图区分随机效应、固定效应本身就是存在问题的。由于 LR 检验结果显示在 5%显著性水平下拒绝原假设, 且固定效应顺序 Logit^{Logit} 模型的回归结果中解释变量都不显著, 我们最终以随机效应顺序 Logit 模型对考察期内地区面板数据进行回归, 如表 5 所示。根据模型数值积分稳健性检验结果显示变量前的系数估计值相对差距均没有超过 10^{-4} , 说明数值积分结果是稳健的。

由虚拟响应变量理论可以得出, Logit 模型并非表示解释变量与响应变量之间的直接关系, 而是间接地表明解释变量对地区涉恐安全风险发生概率的影响。人均国内生产总值、军费支出占比、政府消费最终支出占比和人口增长率四项变量指标通过统计检验, 表明这些指标对地区涉恐安全风险的发生有着明显影响。军费支出占比和人口增长率与地区涉恐安全风险的发生呈现显著的正相关关系, 即军费支出占比和人口增长率的增加对地区涉恐安全风险的上升起到明显的推动作用, 模型中军费支出占比对风险的影响大于其它指标。人均国内生产总值和政府消费最终支出占比与地区涉恐安全风险的发生呈现显著的负相关关系, 即它们对地区涉恐安全风险的上升起到明显的抑制作用。资产形成总额占比、航空运输货运量与货物和服务出口占比则与地区涉恐安全风险之间的关系不明显。

军费支出占比的提高, 一定程度上为地区扩大军购提供资金来源, 落伍军备的更新和防务能力的提升成为地区间军备竞赛的重要驱动力, 差别的存在可能导致社会的失序, 损害地区的公共安全和社会稳定。人口增长速度的加快, 对于一些政策失误或状况混乱的地区, 很可能加剧既有的生存空间之争和社

⁶ 本文在非线性面板模型构建中保留地区分析框架的同时, 剔除地区中数据缺失的国家, 这种情况使得基于 DRF-KPCA 的地区涉恐安全风险分为三级, 分别用代码 0、1 和 2 表示。

会矛盾, 对社会治理带来潜在的负面影响。恐怖袭击事件频发区多数为经济发展落后的地区, 人民生存物质条件的保证和生活条件的改善可以减少部分袭击事件的极端化。

Table 5. Estimated results of panel order Logit model
表 5. 面板顺序 Logit 模型估计结果

变量	系数	标准误差	z 值	显著性	95% 置信区间	
					上限	下限
人均国内生产总值	-0.00005	0.00002	-1.66	0.096*	-0.00010	0.00001
资产形成总额占比	0.05700	0.03859	1.48	0.140	-0.01864	0.13264
军费支出占比	1.57293	0.31435	5.00	0.000***	0.95681	2.18905
政府消费最终支出占比	-0.17444	0.09826	-1.78	0.076*	-0.36703	0.01814
航空运输货运量	-0.00005	0.00004	-1.07	0.286	-0.00014	0.00004
人口增长率	0.73022	0.38351	1.90	0.057*	-0.02145	1.48190
货物和服务出口占比	-0.01933	0.02674	-0.72	0.470	-0.07173	0.03307

注: ***表示在 1% 水平显著, *表示在 10% 水平显著。

本文以分类精确率(*PR*)、召回率(*RE*)以及 F_1 值为评价指标, 计算公式定义为[27]

$$PR = \sum_{i=1}^k p_i \frac{a_i}{a_i + b_i}, RE = \sum_{i=1}^k p_i \frac{a_i}{a_i + c_i}, F_1 = \sum_{i=1}^k p_i \left(\frac{2 \frac{a_i}{a_i + b_i} \frac{a_i}{a_i + c_i}}{\frac{a_i}{a_i + b_i} + \frac{a_i}{a_i + c_i}} \right) = \sum_{i=1}^k p_i \left(\frac{2a_i}{2a_i + b_i + c_i} \right)$$

式中, k 为聚类个数, a_i 为正确分到第 i 类的样本数, b_i 为误分到第 i 类的样本数, c_i 为应该分到第 i 类却没有分到的样本数, p_i 为第 i 类样本的比例。

对构建的 Logit 回归分类模型进行预测效果对比检验, 结果见表 6 所示。可以发现, 先用 DRF-KPCA 进行降维再层次聚类, 然后用 Logit 回归分类的效果优于其他两个组合得到的分类效果。本文所采用的组合在样本内有较高的预测准确性, 精确率、召回率和 F_1 值均达到 75% 以上。

Table 6. Classification and comparison results of order Logit model
表 6. 顺序 Logit 模型分类对比结果

组合	<i>PR</i>	<i>RE</i>	F_1
层次聚类 + Logit 回归	0.550	0.742	0.632
PCA + 层次聚类 + Logit 回归	0.580	0.575	0.544
DRF-KPCA + 层次聚类 + Logit 回归	0.855	0.758	0.791

5. 结语

如何展开地区涉恐安全风险评估与影响因素研究已经成为在研判地区未来的反恐态势、配置反恐资源时考虑的重要问题。本文基于 DRF-KPCA 和非线性面板模型对地区涉恐安全风险进行量化评估和影响因素研究, 具有较强的信息挖掘和确定风险控制措施的作用。

本文首先借助空间统计方法, 分时段明确地区恐怖袭击事件发生态势, 以可视化方式实现安全预警。接着提出一种适用于地区涉恐安全风险量化评估的多维面板数据统计算法, 通过计算每个地区截面的主

成分因子实现了对各时间节点上的量化评估, 获得的特征向量集为后续的分类提供了基础。与其他的主成分聚类算法相比较, 文中所提算法的合理性和准确度都有提升。结合主成分方法提取出来的安全风险级别因子, 运用最大似然估计方法建立随机效应面板顺序 Logit 模型, 考察宏观经济和社会发展因素对地区涉恐安全风险水平的影响, 结果显示模型辨识了大部分的地区涉恐安全风险等级, 同时给出因素在地区涉恐安全风险水平确定中的作用机理, 可以适应反恐风险监测及预警技术的需要。

基金项目

上海市重点学科项目(T0502); 沪江基金资助项目(B14005)。

参考文献

- [1] 张将星. 1970-2013 年全球恐怖主义活动的特征分析——兼论“9·11”事件前后时段比较[J]. 同济大学学报(社会科学版), 2016, 27(2): 48-57.
- [2] Ordu, G.E.O. (2017) Trends and Patterns of Boko Haram Terrorist and Militants' Aggression in Nigeria. *Aggression and Violent Behavior*, **37**, 35-41. <https://doi.org/10.1016/j.avb.2017.08.006>
- [3] 李丽华, 韩思宁. 暴恐事件网络舆情传播机制及预防研究-英国典型案例的实证分析[J]. 情报杂志, 2019, 38(11): 102-111, 54.
- [4] 刘乐. 论恐怖主义的社会解构[J]. 国际安全研究, 2019, 37(4): 30-50, 157-158.
- [5] 周秋君. 恐怖主义在欧洲发展的新态势及其原因分析[J]. 社会科学, 2019, 40(2): 29-37.
- [6] Oliveira, A. and Huertas, A. (2019) How Do Destinations Use Twitter to Recover Their Images after a Terrorist Attack? *Journal of Destination Marketing and Management*, **12**, 46-54. <https://doi.org/10.1016/j.jdmm.2019.03.002>
- [7] 刘德海, 周婷婷. 基于认知差异的恐怖主义袭击对策分析[J]. 系统工程理论与实践, 2015, 35(10): 2646-2656.
- [8] 付举磊, 肖进, 孙多勇, 等. 基于社会网络的恐怖活动时空特征分析[J]. 系统工程理论与实践, 2015, 35(9): 2324-2332.
- [9] 王雷, 王欣, 赵秋红. 基于和声搜索算法优化支持向量机的突发暴恐事件分级研究[J]. 管理评论, 2016, 28(8): 125-132.
- [10] Estrada, M.A.R. and Koutrouas, E. (2016) Terrorist Attack Assessment: Paris November 2015 and Brussels March 2016. *Journal of Policy Modeling*, **38**, 553-571. <https://doi.org/10.1016/j.jpolmod.2016.04.001>
- [11] 郭璇, 吴文辉, 肖治庭. 基于事件树和 PRA 的民航机场恐怖袭击风险评估模型[J]. 计算机应用研究, 2017, 34(6): 1809-1811.
- [12] Buscema, M., Sacco, P.L., Massini, G., et al. (2018) Unraveling the Space Grammar of Terrorist Attacks: A TWC Approach. *Technological Forecasting and Social Change*, **132**, 230-254. <https://doi.org/10.1016/j.techfore.2018.02.006>
- [13] Lanouar, C. and Goaid, M. (2019) Tourism, Terrorism and Political Violence in Tunis: Evidence from Markov-Switching Models. *Tourism Management*, **70**, 404-418. <https://doi.org/10.1016/j.tourman.2018.09.002>
- [14] 何晶, 李本先. 基于多局域的恐怖组织网络择优增长演化模型[J]. 自动化学报, 2019, 45(11): 2137-2147.
- [15] 宋艳, 滕辰妹. 基于现存设施布局的设施选址: 应对恐怖袭击[J]. 运筹与管理, 2019, 28(10): 5-12.
- [16] Tao, X.M., Chang, R., Li, C.X., et al. (2019) Density-sensitive robust fuzzy kernel principal component analysis technique. *Neurocomputing*, **329**, 210-226. <https://doi.org/10.1016/j.neucom.2018.10.052>
- [17] 肖泽磊, 李帮义, 刘思峰. 基于多维面板数据的聚类方法探析及实证研究[J]. 数理统计与管理, 2009, 28(5): 831-838.
- [18] 陈强. 高级计量经济学及 Stata 应用[M]. 第 2 版. 北京: 高等教育出版社, 2014.
- [19] 李兵, 颜晓晨. 中国与“一带一路”沿线国家双边贸易的新比较优势——公共安全的视角[J]. 经济研究, 2018, 53(1): 183-197.
- [20] 谢振东, 文帆, 周子雅, 等. 对恐怖袭击事件的多元统计分析[J]. 经济数学, 2019, 36(3): 89-94.
- [21] 舒洪水, 李燕飙. 基于大数据视角的恐怖袭击特点与趋势分析——以 GTD 中的 7133 次恐怖袭击为样本[J]. 情报杂志, 2019, 38(11): 87-93, 198.
- [22] 李国辉. 全球恐怖袭击时空演变及风险分析研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2014.

- [23] 许通. 社会分裂与恐怖主义动因探究[D]: [硕士学位论文]. 厦门: 厦门大学, 2019.
- [24] Scholkopf, B., Smola, A. and Muller, K.R. (1998) Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, **10**, 1299-1319. <https://doi.org/10.1162/089976698300017467>
- [25] Heo, G., Gader, P. and Frigui, H. (2009) RKF-PCA: Robust Kernel Fuzzy PCA. *Neural Networks*, **22**, 642-650. <https://doi.org/10.1016/j.neunet.2009.06.013>
- [26] Mundlak, Y. (1978) On Pooling Time Series and Cross Section Data. *Econometrica*, **46**, 69-85. <https://doi.org/10.2307/1913646>
- [27] 国显达, 那日萨, 崔少泽. 基于 CNN-BiLSTM 的消费者网络评论情感分析[J]. 系统工程理论与实践, 2020, 40(3): 653-663.