

Research on Risk Identification and Early Warning of Border Minority Crimes Based on RSM Analysis

Shaobing Wu¹, Changmei Wang², Haoshun Cao¹

¹College of Information Network Security, Yunnan Police College, Kunming Yunnan

²College of Energy and Environmental Sciences, Yunnan Normal University, Kunming Yunnan

Email: 1063093199@qq.com

Received: Feb. 9th, 2019; accepted: Feb. 21st, 2019; published: Feb. 28th, 2019

Abstract

Based on the collection of crimes committed by ethnic minorities in the border areas, this paper fills in and completes the data. Response surface (RSM) analysis method and C# are used to realize risk identification and early warning of ethnic minority crimes. Through response surface (RSM) analysis, modeling, regression equations are established, and the factors affecting minority crimes in border areas are obtained. These crime reason items are designed as key information. By inputting the relevant information of the relevant personnel, the system automatically calculates the probability of the person's crime risk, and makes an early warning judgment on the person under a predetermined threshold.

Keywords

RSM, C# Implementation, Ethnic Minority Crime, Risk Identification

基于RSM分析的边疆少数民族犯罪风险识别及预警研究

吴绍兵¹, 王昌梅², 曹好顺¹

¹云南警官学院, 信息网络安全学院, 云南 昆明

²云南师范大学, 能源与环境科学学院, 云南 昆明

Email: 1063093199@qq.com

收稿日期: 2019年2月9日; 录用日期: 2019年2月21日; 发布日期: 2019年2月28日

摘要

本文在收集边疆少数民族犯罪案例的基础上,通过数据填补和完备。利用响应曲面(RSM)分析方法和C#实现了少数民族犯罪风险识别及预警研究。通过响应曲面(RSM)分析,进行建模,建立回归方程,并得出影响边疆少数民族犯罪的因素。将这些犯罪原因项设计成关键信息,利用这些信息,系统自动计算出该人的犯罪风险概率,在预先给定的阈值情况下,对该人进行预警判断。

关键词

RSM, C#实现, 少数民族犯罪, 风险识别

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

2019年1月21日,习近平总书记在省部级主要领导干部坚持底线思维着力防范化解重大风险专题研讨班开班式上发表重要讲话,强调要坚持底线思维,增强忧患意识,提高防控能力,着力防范化解重大风险,保持经济持续健康发展和社会大局稳定。维护国家安全和社会大局稳定,是时代赋予我们的重要任务,如何保持边疆少数民族地区的繁荣稳定,是我们应该关注的重要问题。

人工智能时代,人们对安全和犯罪也更加关注,孙笛[1]提出了人工智能时代的犯罪风险分析,该文主要是基于技术、监管与融合的视角,指出应从限制人工智能研发应用的领域、提高人工智能系统数据和算法的客观性、完善行业自律准则与法律政策以及通过一系列具体措施实现犯罪防控与人工智能的深度融合四个方面采取有效的防控对策。给出了在人工智能时代,如何进行犯罪风险分析的一个框架式的思路。

在进行犯罪风险预测的过程中,构建模型所使用的训练数据的质量直接决定了最终预测结果的准确性,如何选择与预测结果相关的有效指标信息成为建立模型的关键。王雨晨[2]等介绍了一种对于数据噪声鲁棒、预测结果准确且稳定的组合分类算法——随机森林(Random Forest),应用于犯罪信息指标集合的选取与犯罪风险的预测中,给犯罪风险识别研究一个很好的借鉴和参考。

研究犯罪风险,找寻和识别犯罪风险源,西南政法大学的师索[3]在犯罪与风险研究论纲一文中给出了犯罪风险源包括人造风险、发展性风险、人性风险以及激发性风险,并指出了犯罪和风险发生各种关系。为进一步研究犯罪风险,预测犯罪,提供了一个新思路。

在大数据时代,进行数据分析时,常常需要研究响应变量 y 究竟如何依赖于自变量 x ,从而找到自变量的设置,使得响应变量得到最佳值。响应曲面方法就是其中最好的方法之一[4]。

邓聪颖[5]等针对加工空间机床动态特性的准确预测问题,提出一种基于响应曲面理论的数控机床空间动态特性研究方法。该方法以数控机床固有频率这一关键动力学性能指标为例,分析机床整机动态特性与机床位置姿态之间的数学关联关系;基于正交试验设计和响应曲面方法理论,构建预测广义加工空间数控机床动态性能变化的响应曲面模型。表明了响应曲面方法在数控机床领域得到了很好的应用。

以上作者对于犯罪风险的研究和分析,要么是基于政策的角度的去分析研究,给出宏观对策和建议,

要么是建立犯罪风险模型, 给出犯罪预测的思路。响应曲面方法在数控机床领域得到了很好的应用, 但未见 RSM 方法在少数民族犯罪中的应用报道。给我们研究边疆少数民族犯罪提供了很好的思路和参考。

但是这些对于基层政法干警来说, 不够实用, 可操作性不强。为此, 本文通过收集的边疆少数民族犯罪案例数据, 在结合实际田野调研经验和办案民警的经验基础上, 利用 Minitab 和响应曲面(以下简称 RSM)方法建立相关数学模型, 采用 Visual studio 之 C#和 Minitab 开发了边疆少数民族犯罪风险识别及预警系统软件。

该软件采用 C#作为前端的界面和模块的设计, 后台采用 Access 实现。在收集的 11,326 个边疆少数民族犯罪案例的基础上, 通过数据预处理、数据完备性填补、删减, 形成完备的数据, 便于进一步的分析。通过响应曲面(RSM)分析, 建立回归方程, 并得出影响边疆少数民族犯罪的因素。将这些犯罪原因项设计成关键信息, 通过输入相关人员的这些信息, 系统自动计算出该人的犯罪风险概率, 在预先给定的阈值情况下, 对该人进行预警判断。该软件界面操作简单、可视化和易使用, 为相关人员进行犯罪预测提供重要的参考。

2. MINITAB 简介

MINITAB 软件是 MINITAB 公司开发的一款专用数据处理软件, MINITAB 以其简洁的操作界面和强大的统计功能已经成为质量统计领域与六西格玛方法的常用软件, 它的实验设计操作简单、方便易用[4]。它包含了丰富的统计数据信息, 既可以进行数据分析处理, 又可以进行图形分析对比, 同时对未来加以预测。MINITAB 的核心功能之一就是进行数据分析处理, 其功能强大、快速便捷。涵盖了统计技术的基本统计、回归分析、方差分析、实验设计分析、控制图、质量工具和可靠性工具等[6]。

3. 数据收集与预处理

3.1. 数据收集

本实验所收集的数据来自中国裁判文书网, 收集了 2014~2017 年, 云南省部分县市少数民族犯罪案例。对收集的非结构化数据, 我们通过 python 软件, 从非结构化的信息中提取除结构化的数据。在数据预处理时, 首先删除民族为空和民族为汉族的数据, 其次删除, 犯罪类型为空的数据。接着对不完备的数据根据相似性原则, 进行补全。最后, 所获数据为 11,326 条。

3.2. 数据编码

通过收集的 11,326 条数据, 先对数据进行编码, 按照民族人口和案例中的民族数量, 分别将彝族、傣族、白族、壮族等依次编码为 25、24、23、...、直到 2; 将性别男编码为 3, 女编码为 2; 职业中无业编码为 6、农民编码为 5、工人(含个体户、驾驶员等)编码为 4、公职人员编码为 3、学生编码为 2; 文化程度: 小学编码为 6、文盲编码为 5、初中编码为 4、高中或中专编码为 3、大专及以上编码为 2; 年龄采用整数; 根据犯罪类型数量的多少, 依次将盗窃罪编码为 130、与毒品一个的犯罪类型编码为 120; 危险驾驶罪编码为 110。详见表 1 到表 5。

Table 1. Gender coding table
表 1. 性别编码表

名称	编码
男	3
女	2

Table 2. Educational degree code table**表 2.** 文化程度编码表

名称	编码
文盲	6
小学	7
初中	5
高中	4
大专以上	3

Table 3. Occupational coding table**表 3.** 职业编码表

名称	编码
无业	6
农民	5
工人(含个体、驾驶员等)	4
公职人员	3
学生	2

Table 4. Ethnic code table**表 4.** 民族编码表

名称	编码
彝族	25
傣族	24
哈尼族	23
...	...
怒族	2

Table 5. Crime type code table**表 5.** 犯罪类型编码表

名称	编码
盗窃罪	130
贩卖(含运输、组织、走私)毒品罪	120
危险驾驶罪	110
故意伤害罪	100
非法持有枪支罪	90
交通肇事罪	80
抢劫罪	70
滥伐林木罪	60

4. 模型建立

4.1. 响应曲面方法设计实验

响应曲面方法进行实验时, 常见的有两种方法: 中心复合设计(Central composite design, CCD)与 Box-Behnken 设计。

中心复合设计整个实验由三部分: 立方体点、中心点和星号点等构成。根据星号点的设置和位置的

不同中心复合设计(CCD)又有三种形式: 中心复合序贯设计(Central composite circumscribed design, CCC)、中心复合有界设计(Central composite inscribed design, CCI)和中心复合表面设计(Central composite face-centered design, CCF)。

Box-Behnken 设计是将因子各实验点取在立方体的棱的中点上, 和中心复合设计类似, 它也具有近似旋转性, 但没有序贯性。它与中心复合设计的特点比较如下表 6 所示:

Table 6. Comparison of features between central composite design (CCD) and Box-Behnken design
表 6. 中心复合设计(CCD)与 Box-Behnken 设计的特点比较

CCD			Box-Behnken
CCC	CCI	CCF	
具有旋转性	具有旋转性	不具有旋转性	具有或近似具有旋转性
预测精度较高	与 CCC 相比, 只在较小设计区间内提供同样精度的预测	预测精度高	与 CCI 类似, 只在较小设计区间内提供同样精度的预测
每个因子需要 5 个水平	每个因子需要 5 个水平	每个因子需要 3 个水平	每个因子需要 3 个水平 不易进行序贯和区组设计

响应曲面分析方法主要有以下几步[4]:

第一步: 浏览数据

浏览一下数据的基本情况, 观察是否有异常的数据存在, 是否有手工输入错误的的数据。

第二步: 拟合模型

在 minitab 软件中采用“统计 → DOE → 响应曲面 → 分析响应曲面设计”(Stat → DOE → Response Surfac → Analyze Response Surface Design)窗口中选定安全模型后, 在运行窗口中可得到计算结果。

第三步: 简化模型

操作同上, 在 minitab 软件中采用“统计 → DOE → 响应曲面 → 分析响应曲面设计”(Stat → DOE → Response Surface → Analyze Response Surface Design), 把响应项移入响应栏内进行操作即可。

第四步: 残差诊断

在 minitab 软件中采用“统计 → DOE → 响应曲面 → 分析响应曲面设计”(Stat → DOE → Response Surface → Analyze Response Surface Design), 点击项(Terms), 选择相关项, 点击确定(OK), 回到分析响应曲面设计窗口。选择图形(Graphs), 在残差图, 单击相关选项, 按确定(OK), 即可进行残差分析。

第五步: 判断模型是否合适, 需要改进吗?

如果合适, 输入相关项和对应的系数。

根据这些项和对应的系数, 带入下面的二次多元回归拟合方程中, 即可得出回归方程:

$$y = \beta_0 + \sum_{i=1}^7 \beta_i x_i + \sum_{i=1}^7 \beta_{ii} x_i^2 + \sum_{i=1}^2 \sum_{j=1}^7 \beta_{ij} x_i x_j \tag{1}$$

式中: y 代表响应值; β_0 、 β_i 、 β_{ii} 表示偏移项、线性偏移和二阶偏移系数; β_{ij} 是交互效应系数; x_i 各因素的编码值。以 $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ 为自变量, y 为响应值。

4.2. 响应曲面分析

利用 Minitab 软件, 按照“Minitab - 统计 - DOE - 响应曲面 - 分析响应曲面设计”的顺序, 对中心复合设计实验结果进行响应曲面分析, 分析结果见表 7 所示。从表 6 中可以看出, x_4 即职业对少数民族犯罪有显著的影响, 而且曲面效应 $x_1 * x_2, x_1 * x_3, x_1 * x_4$ 和 $x_1 * x_4$ 都很显著。表明在性别与年龄、民族和职业, 以及年龄和职业对少数民族犯罪的影响显著。

Table 7. Regression coefficient and significance of quadratic equation model
表 7. 二次方程模型回归系数和显著性

项	效应	系数	系数标准误	T 值	P 值	方差膨胀因子
常量		65.56	6.81	9.62	0.000	
x_1	-7.58	-3.79	3.86	-0.98	0.326	
x_2	10.9	5.4	12.4	0.44	0.660	
x_3	7.32	3.66	5.60	0.65	0.513	
x_4	104.76	52.38	6.74	7.77	0.000	
x_5	-2.53	-1.26	5.20	-0.24	0.808	
$x_2 * x_2$	54.23	27.12	6.90	3.93	0.000	+
$x_3 * x_3$	-19.00	-9.50	2.01	-4.73	0.000	+
$x_4 * x_4$	-34.53	-17.26	3.74	-4.61	0.000	+
$x_5 * x_5$	-5.07	-2.54	2.29	-1.11	0.268	
$x_1 * x_2$	-25.87	-12.94	4.74	-2.73	0.006	+
$x_1 * x_3$	16.29	8.15	2.47	3.29	0.001	+
$x_1 * x_4$	-19.82	-9.91	2.83	-3.51	0.000	+
$x_1 * x_5$	0.39	0.19	2.17	0.09	0.929	
$x_2 * x_3$	-4.57	-2.28	6.53	-0.35	0.727	
$x_2 * x_4$	41.65	20.83	8.30	2.51	0.012	+
$x_2 * x_5$	-24.17	-12.09	6.52	-1.86	0.064	
$x_3 * x_4$	-9.34	-4.67	3.76	-1.24	0.215	
$x_3 * x_5$	1.21	0.61	2.82	0.21	0.830	
$x_4 * x_5$	-0.99	-0.50	3.16	-0.16	0.875	

注：“+”表示对应项影响显著。

4.3. 响应曲面方程的建立

根据实验分析结果得到的各因素效应系数，建立少数民族犯罪对无个显著影响因素的多元二次方程。

分别将性别(x_1)、年龄(x_2)、民族(x_3)、职业(x_4)和文化程度(x_5)作为自变量，犯罪类型(y)作为因变量，在上述编码的基础上，采用 minitab 中的回归分析及响应曲面分析，得出其回归方程模型如下：

$$y = 65.56 - 3.79x_1 + 5.4x_2 + 3.66x_3 + 52.38x_4 - 1.26x_5 + 27.12x_2^2 - 9.50x_3^2 - 17.26x_4^2 - 2.54x_5^2 - 12.94x_1x_2 + 8.15x_1x_3 - 9.91x_1x_4 + 0.19x_1x_5 - 2.28x_2x_3 + 20.83x_2x_4 - 12.09x_2x_5 - 4.67x_3x_4 + 0.61x_3x_5 - 0.50x_4x_5$$

籍贯(本地户口、外来人口和境外人员)根据数据样本在总样本中出现的概率情况，分别赋予一定的权值，将一个人是否具有人格障碍、反社会性、法律意识，以及是否知晓法制宣传教育制度和办案民警的经验等，分别通过专家打分的形式，赋予不同的权值，作为模型的一部分，从而使得建立的模型，更加准确。

在建立模型的基础上，系统自动计算出犯罪风险识别概率和积分，在预先给定的阈值情况下，给出是否关注的预警提示。

5. 边疆少数民族犯罪风险识别及预警的编程实现

根据上面响应曲面设计方程，结合上面的算法思想，利用 C#设计边疆少数民族犯罪风险识别及预警系统界面，在该界面中，依次选择民族、性别、职业、学历和籍贯(本地人口或外地人口)，输入年龄，选择“人格障碍”、“法律意识”、“反社会性”、“法律宣传教育制度”和输入专家经验的分数。单击

“计算”按钮，系统会自动计算出输入相关特征信息后，该人的犯罪风险概率、当概率大于一个预先给定的阈值后，系统会给出预警提示。选择不同的参数，输入不同的值后，得出的犯罪风险概率值不同。如图 1 和图 2 所示。



Figure 1. Probability and warning of the crime risk of ethnic minorities in border areas
图 1. 边疆少数民族犯罪风险的概率及预警提示

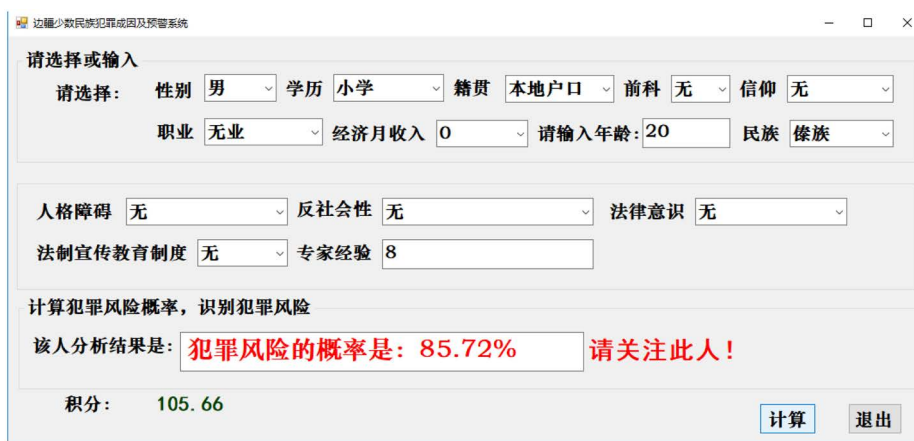


Figure 2. Probability and warning of the crime risk of ethnic minorities in border areas
图 2. 边疆少数民族犯罪风险的概率及预警提示

6. 结语

本文提出的采用响应曲面(RSM)技术及 C#来开发的少数民族犯罪风险识别及预警系统，在大数据分析的基础上，建立数学模型，编程实验求解，同时在系统中，引入了“人格障碍”、“反社会性”、“法律意识”和“法律宣传教育制度”以及有经验民警和专家打分的“专家经验”值。使得该系统的预警更加科学合理，利于指导实践，使得民警在打击犯罪进行犯罪风险识别与预警中进行科学预判。

与其他方法不同的是，未见利用 RSM 和 MINITAB 对少数民族犯罪进行分析研究，同时未见少数民族犯罪风险识别和预警的报道。本文的研究，作为一个尝试，收到了非常好的效果。

基金项目

国家社会科学基金项目(13CFX038); 云南省教育厅科学研究基金项目(2013C188); 云南警官学院教育教学改革项目(2018YJJGB01)。

参考文献

- [1] 孙笛. 人工智能时代的犯罪风险分析[J]. 中国人民公安大学学报(社会科学版), 2018, 164(4): 11-16.
- [2] 王雨晨, 过仲阳, 王媛媛. 基于随机森林的犯罪风险预测模型研究[J]. 华东师范大学学报(自然科学版), 2017(4): 89-96.
- [3] 师索. 犯罪与风险研究论纲——风险社会视野下的犯罪治理[J]. 中南大学学报(社会科学版), 2011, 17(6): 83-91.
- [4] 闵亚能. 实验设计(DOE)应用指南[M]. 北京: 机械工业出版社, 2011.
- [5] 邓聪颖, 刘蕴, 殷国富. 基于响应曲面方法的数控机床空间动态特性研究[J]. 工程科学与技术, 2017, 49(4): 211-218.
- [6] 王会良, 朱德荣, 桂斌, 杨进濮. MINITAB 软件在工序质量控制中的应用[J]. 煤矿机械, 2006, 27(10): 114-116.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org