

基于证据权重逻辑回归的涉恐人员身份预测

高洪美, 李保东

(上海计算机软件技术开发中心 大数据与区块链技术研究所, 上海 200443)

摘要: 精准预测涉恐人员身份对公安机关反恐工作意义重大。为此, 针对涉恐情报挖掘时的数据处理和预测效率, 综合考虑公安信息五要素, 建立涉恐人员标签体系, 利用标签精确描述涉恐人员的数据特征。同时, 引入证据权重和信息价值筛选强相关指标变量, 建立证据权重逻辑回归模型预测涉恐人员身份, 以解决虚拟变量陷阱和过拟合问题。实验表明, 所提模型对涉恐人员的预测效果优于逻辑回归等预测模型。

关键词: 涉恐人员; 身份预测; 标签体系; 逻辑回归; 证据权重; 信息价值

DOI: 10.11907/rjdk.231384

开放科学(资源服务)标识码(OSID):



中图分类号: TP393

文献标识码: A

文章编号: 1672-7800(2023)008-0017-07

Identification Prediction of Terrorism Involved Persons Based on Weight of Evidence Logistic Regression

GAO Hongmei, LI Baodong

(Big Data and Blockchain Technology Research Institute, Shanghai Development Center of Computer Software Technology, Shanghai 200443, China)

Abstract: Accurately predicting the identity of individuals involved in terrorism is of great significance for the anti-terrorism work of public security organs. To this end, in order to improve the efficiency of data processing and prediction in the mining of terrorism related intelligence, we comprehensively consider the five elements of public security information and establish a label system for terrorism related personnel, using labels to accurately describe the data characteristics of terrorism related personnel. At the same time, the evidence weight and information value are introduced to screen strongly related indicator variables, and the evidence weight logistic regression model is established to predict the identity of terrorist related persons, so as to solve the trap and over fitting problems of dummy variables. The experiment shows that the prediction effect of the proposed model on terrorist related personnel is better than that of logistic regression and other prediction models.

Key Words: terrorist-related personnel; identity prediction; labelling system; logistic regression; weight of evidence; information value

0 引言

在数字化时代, 打击恐怖活动的前提是具备情报分析和预测能力。实际上, 情报挖掘工作面临的数据类型复杂多样, 涵盖人员背景、社会活动等维度, 增加了数据处理难度^[1], 不同维度数据还可能出现缺失, 容易造成情报工作线索中断或判断失误等情况^[2]。

传统特征分析和预测主要从涉恐人员的特征角度出发, 基于犯罪学、社会学和心理学等理论, 研究涉恐人员的

人格、心理、社会等特征因素, 建立异常行为人的异常背景标签、生活标签等组成的标签预警体系, 开展涉恐人员的行为分析工作^[3-7]。同时, 利用涉恐情报挖掘方法提取数据特征, 对数据进行清洗降维, 协助反恐预警和决策^[8-9]。

随着数据复杂度和数据量提升, 深度学习方法逐渐被使用, 在个别场景预测上已具有一定成效^[10-11]。例如, 基于旋转森林集成学习^[12]、Probit模型^[13]等算法在特定场景, 能在一定程度上进行信息挖掘, 但在涉恐数据的粗糙处理及预测模型应用中仍然存在问题尚未解决, 导致其在涉恐人员的分析、预测方面发展相对滞后。

收稿日期: 2023-04-12

基金项目: 上海市软件技术创新服务平台项目(20DZ2291700)

作者简介: 高洪美(1986-), 女, 上海计算机软件技术开发中心大数据与区块链技术研究所工程师, 研究方向为大数据、数据治理、数据资产; 李保东(1992-), 男, 上海计算机软件技术开发中心大数据与区块链技术研究所工程师, 研究方向为大数据、数据治理、数据资产。本文通讯作者: 高洪美。

目前,常用的决策树、随机森林等算法对噪声数据较为敏感,容易出现过拟合现象,数据处理时易忽略数据集中属性间的相关性,缺乏数据缺失处理方法。逻辑回归算法作为一种预测算法能较为容易地解决上述问题,且对数据噪声的鲁棒性更优,不易受轻微的多重共线性影响。因此,针对涉恐人员的身份预测,在数据处理和分析预测方面,逻辑回归算法相较于其他预测算法适用性更强。

综上,本文通过构建涉恐人员标签体系来精确描述涉恐人员的数据特征,提出一种基于证据权重逻辑回归的涉恐人员身份预测模型。首先根据证据权重解决逻辑回归模型的虚拟变量陷阱问题,以提升预测模型的准确性。同时,引入信息价值计算每个指标变量的信息量程度,以度量指标变量的预测能力,有效避免模型发生过拟合现象。

此外,证据权重逻辑回归模型能直接呈现涉恐人员不同标签特征对其预测结果的相对重要性,模型自身具有较

强的稳定性和较高的准确性,能帮助公安机关提升涉恐人员身份预测的能力和办案效率。

1 涉恐人员标签体系设计

公安数据包含人、事、物、地和组织等丰富多样的信息,从犯罪案件构成角度涉及人及其关系、行为、事、时间、空间和主观意图等数据。

合理、准确的数据标签是对人、事、物、地、组织全方位信息的深入理解和认知。通过数据标签,公安机关可从采集的海量数据中对犯罪信息进行提取、分析、研判和预测。为此,本文基于公安五要素信息模型,对采集清洗后的涉恐人员数据进行统计分类^[14],从人员、地点、物品、案(事)件、组织关系五大维度构建涉恐人员标签体系,如图1所示。

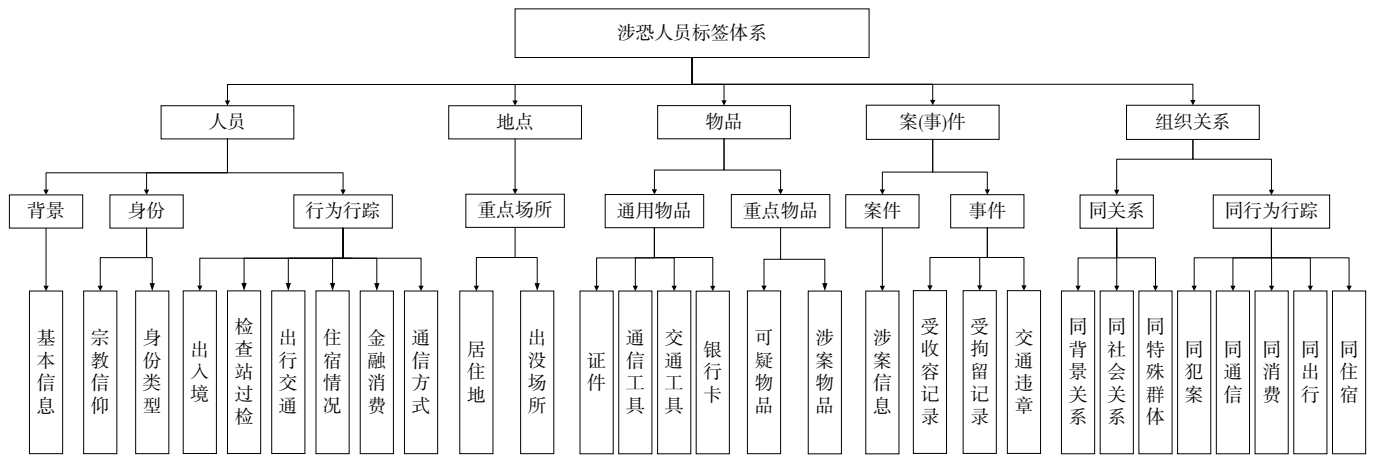


Fig. 1 Terrorist personnel labeling system

图1 涉恐人员标签体系

1.1 人员

人员类标签是公安反恐业务对象信息的高度抽象和归类,背景和身份作为描述涉恐人员的基本属性,能真实、直接的识别人物信息。例如,身份类型包含正常人、嫌疑人、涉案人员、公安部7类重点人员等标签。行为行踪则从出入境、检查站过检、出行、住宿、金融、通信等方面覆盖人员的社会活动和生活轨迹,为分析涉恐关注对象的异常行为提供参数支持。

1.2 地点

地点类标签是反恐工作重点关注的区域和场所类信息的抽象概括。一般而言,恐怖袭击较多发生在人员聚集场所,例如金融消费场所、校园、交通枢纽、办公场所、住宿等场所。因此,在进行涉恐人员身份预测时需要重点关注该地区的异常情况。

1.3 物品

物品类标签是与人员相关物品信息的抽象概括,一般包括通用物品和重点物品。通用物品能直接关联关注对象,或关注到异常状态而被记录下的各类物品,例如身份证件、交通工具、电话、银行卡等,均具有与人员行为密切

相关的使用痕迹;重点物品指受国家法规限制的可疑物品,例如违禁物品、管制物品和其他涉案物品,将直接反映涉恐人员的危险行为。

1.4 案(事)件

案(事)件类标签主要是对案件、特殊事件信息进行刻画,都是经过真实事件发生后采集获取的数据,因此该类标签具有详细、准确的定义,能作为直接依据判定涉恐对象的异常行为特征,例如涉案信息、受收容记录、受拘留记录、交通违章等。

1.5 组织关系

组织关系类标签是通过归纳得出的两个及以上人与人、人与物所存在的信息关联描述,可称为同一性描述。例如,若关注对象的标签与涉恐人员或重点人员存在同社会关系、同行行为行踪等,则能构建起标签强关联关系,这也是涉恐人员预测模型的重要参数依据。

综上,涉恐人员标签体系充分考虑了重点性、准确性、延伸性、扩展性等原则,对人员、地点、物品、案(事)件、组织关系五大维度进行深入细分。通过对涉恐对象数据进行归纳、关联、挖掘,结合业务工作具体需求,根据业务特

征变化,对标签体系进行调整和扩展。

2 涉恐人员身份预测模型构建

2.1 逻辑回归模型

逻辑回归(Logistic Regression)是一种用于解决二分类问题的机器学习方法,也被称为广义的线性回归分析模型,通过Sigmoid函数引入非线性因素,因此可处理0/1分类问题,常用于预测、判别某个类别或事件的可能性^[15-17]。

逻辑回归相较于其他算法,对分类问题中的变量概率分析、逻辑回归分类结果、影响力较大的变量因素具备良好的解释性。假设给定样本集合 $L = \{u_i, v_i\}$,其中 u_i 表示涉恐人员的指标变量, v_i 表示一个二分类变量。 $v_i = 1$ 表示第 i 个涉恐人员为恐怖分子, $v_i = 0$ 表示第 i 个涉恐人员为非恐怖分子。构建涉恐人员身份预测模型,通过式(1)预测涉恐人员为恐怖分子的概率。

$$P(v = 1|U) = \frac{e^{\beta_0 + \beta^T U}}{1 + e^{\beta_0 + \beta^T U}} \quad (1)$$

式中: U 的维度等于涉恐人员指标变量个数的向量; β 为待估计参数,也是各个指标变量的权重。

逻辑回归作为一种非线性回归模型,通常采用极大似然估计进行求解。假设 N 个总体样本为 V_1, V_2, \dots, V_N ,随机抽取 n 个样本记为 $v_1, v_2, \dots, v_n, p_i = p(v_i = 1|u_i)$ 为在给定条件 u_i 下, p_i 为 $v_i = 1$ 的条件概率。由于 $v_i = 0, v_i = 1$ 为全概率事件,在给定相同条件 u_i 下, $p(v_i = 0|u_i) = 1 - p_i$ 为 $v_i = 0$ 的条件概率。因此,任意一个样本值的概率为:

$$P(v_i) = p_i^{v_i} (1 - p_i)^{1 - v_i} \quad (2)$$

式(2)的似然函数为:

$$R(\beta_0, \beta_1, \dots, \beta_j) = \prod_{i=1}^n p_i^{v_i} (1 - p_i)^{1 - v_i} \quad (3)$$

为便于操作通常对式(2)取对数求解,如式(4)所示。

$$\ln[R(\beta_0, \beta_1, \dots, \beta_j)] = \ln\left[\prod_{i=1}^n p_i^{v_i} (1 - p_i)^{1 - v_i}\right] = \sum_{i=1}^n v_i (\beta_0 + \beta_1 x_{i1} + \dots + \beta_j u_{ij}) - \ln(1 + e^{\beta_0 + \beta_1 u_{i1} + \dots + \beta_j u_{ij}}) \quad (4)$$

在样本随机分布条件下,通过逻辑回归模型分析发现,极大似然估计是有效、一致且渐进分布的,对式(1)推导出:

$$\ln \frac{p(v = 1|u)}{1 - p(v = 1|u)} = \beta_0 + \beta^T U \quad (5)$$

式中:等式左边指某一指标变量下,涉恐人员为恐怖分子的概率与涉恐人员为非恐怖分子概率比值的对数,数值越大表示涉恐人员为恐怖分子的概率越大。

2.2 证据权重逻辑回归模型

在实际预测中,逻辑回归模型仍存以下不足:①如何选择指标变量是预测模型建立前期较为棘手的问题;②逻辑回归模型在处理分类变量时一般需要设置虚拟变量,否

则容易出现虚拟变量陷阱,但在采集的数据集中通常存在许多分类变量,如果所有分类变量均转换为虚拟变量,将导致分析结果十分复杂且会降低预测模型的准确度。

为解决上述问题,本文在逻辑回归模型中引入证据权重(Weight Of Evidence, WOE)^[18]和信息价值(Information Value, IV)计算每个指标变量的信息量程度,以度量指标变量的预测能力,有效控制模型发生过拟合现象^[19]。

2.2.1 计算WOE和IV

设 $U = (u_1, u_2, \dots, u_n)$ 为涉恐人员的 n 个指标变量, V_i 表示涉恐人员实际身份, $V_i = 1$ 表示涉恐人员为恐怖分子, $V_i = 0$ 表示涉恐人员非恐怖分子。为合理分箱指标变量,首先将其按照升序排列,记为 $U' = (u_{(1)}, u_{(2)}, \dots, u_{(n)})$;然后根据等概率法选取 $m-1$ 个分位点,将 U 分为 m 个箱($2 < m < 5$)^[20];接下来设 D_i, Q_i 分别为第 i 个区间的恐怖分子和非恐怖分子的数量, D, Q 分别为恐怖分子和非恐怖分子的总数,以此计算各分箱内的 WOE_i, IV 值。最后,比较每个分箱方式的IV值,选择IV值最大的分箱方式作为指标变量的合理分箱^[21]。

$$WOE_i = \ln \frac{Q_i/Q}{D_i/D} \quad (i = 1, 2, \dots, m) \quad (6)$$

$$IV = \sum_i \left(\frac{Q_i}{Q} - \frac{D_i}{D} \right) WOE_i \quad (7)$$

2.2.2 根据IV值筛选指标变量

由式(6)、式(7)可知,IV值由WOE值加权求和得出,表示自变量对因变量的贡献程度,IV值越大表示IV值与该指标变量的相关性越强,如表1所示。因此,在模型构建过程中选取IV值较大的指标变量($IV \geq 0.3$),舍弃IV值较小的指标变量。

Table 1 Correlation between IV value and indicator variables

表1 IV值与指标变量的相关性

IV值	相关性
$IV < 0.02$	不相关
$0.02 \leq IV < 0.1$	弱相关
$0.1 \leq IV < 0.3$	中等相关
$IV \geq 0.3$	强相关

2.2.3 计算涉恐人员身份预测概率

将选取指标变量对应的WOE值作为新设计矩阵代入逻辑回归模型,并利用模型中的方法解决选取变量间可能存在的多重共线性问题^[22],以计算涉恐人员身份预测概率。

3 案例应用

3.1 数据来源

本文样本数据来自某公安大数据项目,剔除采集样本数据的重复项后,共保留了1 250条涉恐人员的数据作为模型回归拟合的总样本。其中,1 000条作为训练集,250条作为测试集。研究表明,当非涉恐人员数量为涉恐人员

数量的3~5倍时评估模型更有效,因此在1000条训练样本设置750个非恐怖分子、250个恐怖分子,测试集样本则为随机样本。

3.2 指标变量选取

本文从人员、地点、物品、案(事)件和组织关系5个方面,选取标签体系中具有代表性的变量指标,作为影响涉恐人员身份的变量,用于构建涉恐人员身份预测评估模型。指标变量的选取原则如下:

3.2.1 操作性

公安五大要素包含的指标变量众多,过多的指标变量可能会导致混淆和解释性问题,过少的指标变量可能导致研究结果不准确,因此需要从每个因素中选取具有代表性的指标变量作为计算依据,以符合实际的可操作性。

3.2.2 实现性

实际上,并非每个指标具有完整的数据,数据的缺失将降低模型的可信性,因此需要从众多指标中选取具有代表性且有完整数据的指标作为计算依据,以符合可实现的需求。

3.2.3 现实性

本文指标的选取参考了公安实际工作中关注的指标变量,模型研究联系实际应用,具备一定的现实性。

综上所述,通过操作性、实现性、现实性3个维度对指标变量进行筛选,得到涉恐人员身份预测变量类型如图2所示。

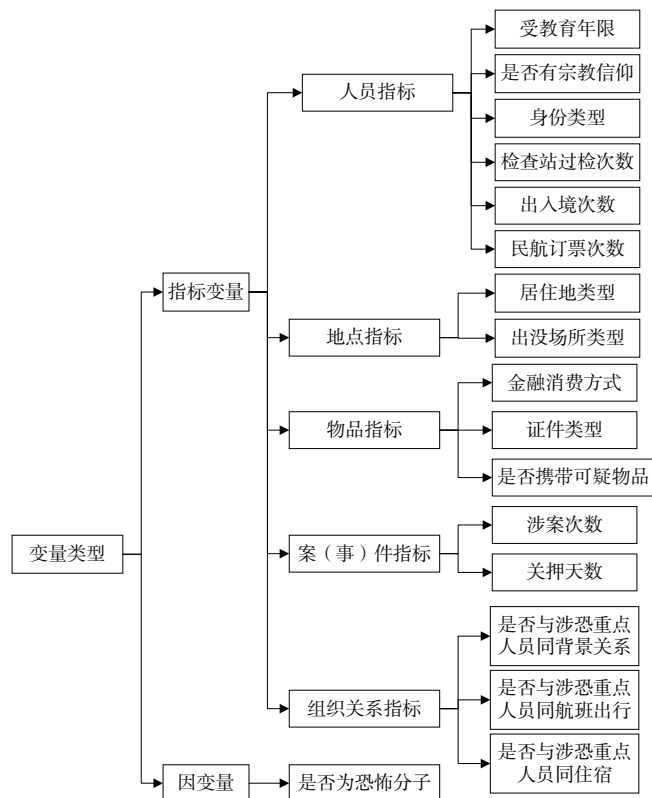


Fig. 2 Types of variables predicting the identity of persons involved in terrorism

图2 涉恐人员身份预测变量类型

3.3 指标变量分组与筛选

为了减少计算复杂度、增强模型稳定性,依据m-1分位法和IV值对指标变量进行合理分组。具体为,首先计算各组分类变量的WOE、IV值;然后按升序排列连续变量,按照等概率法选取m-1分位点,计算出各组的WOE、IV值;最后选择IV值最大的一种分组作为该变量的分组。本文以涉案次数为例分组计算WOE、IV值,通过比较不同涉案次数的分组发现,将涉案次数分组列为4组时IV值最大,因此该指标变量的分组为4组。表2为各组WOE、IV值,表3为通过该方法计算出所有指标变量的IV值。

Table 2 Calculation results of indicator variables

涉案次数	Qi	Di	Qi/Q	Di/D	WOEi	IV值
涉案次数≤1	650	20	0.866 7	0.080 0	2.382 6	1.874 3
1<涉案次数≤3	80	30	0.106 7	0.120 0	-0.117 8	0.001 6
3<涉案次数≤7	18	80	0.024 0	0.320 0	-2.590 3	0.766 7
7<涉案次数	2	120	0.002 7	0.480 0	-5.193 0	2.478 8
合计	750	250	1.000 0	1.000 0	-5.518 4	5.121 4

Table 3 Statistical results of indicator variable IV value

指标变量	IV值
受教育年限	0.087 6
是否有宗教信仰	0.908 0
身份类型	3.832 1
出入境次数	0.213 2
检查站过检次数	2.165 8
民航订票次数	0.209 8
金融消费方式	0.020 5
居住地类型	1.232 7
出没场所类型	0.221 0
证件类型	0.059 8
是否携带可疑物品	2.341 2
涉案次数	5.121 4
关押天数	2.848 3
是否与涉恐重点人员同背景关系	3.468 9
是否与涉恐重点人员同航班出行	0.010 6
是否与涉恐重点人员同住宿	1.776 1

依据表1的IV值大小与指标变量相关性规定,剔除IV值<0.3的指标变量,筛选出9个强相关性的指标变量,如表4所示。由此可知,IV值越大表明该指标变量对判定涉恐人员身份的影响程度越大,图3展示了各指标变量的IV值排序结果。

3.4 回归结果分析与验证

首先,经过上述步骤处理,采用SPSS工具通过二元逻辑回归计算上述9个指标变量,结果如表5所示。由此可知,9个指标均对涉恐人员身份预测产生了显著的正向影响。然后,将训练样本带回检验得到二元逻辑回归模型的预测准确性,如表6所示。由此可知,模型对训练样本的

Table 4 Screening results of indicator variable IV value

表 4 指标变量 IV 值筛选结果

指标变量	IV 值
是否有宗教信仰	0.908 0
身份类型	3.832 1
检查站过检次数	2.165 8
居住地类型	1.232 7
是否携带可疑物品	2.341 2
涉案次数	5.121 4
关押天数	2.848 3
是否与涉恐重点人员同背景关系	3.468 9
是否与涉恐重点人员同住宿	1.776 1

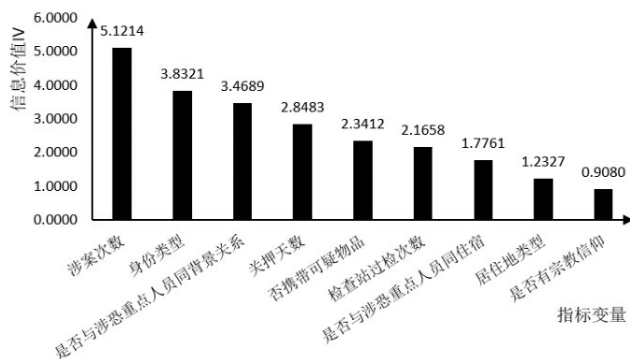


Fig. 3 Index variable information value ranking

图 3 指标变量信息价值排序

Table 5 Results of binary logistic regression analysis

表 5 二元逻辑回归分析结果

项	B	标准误差	瓦尔德	自由度	显著性	EXP (B)
是否有宗教信仰	1.272	0.281	20.504	1	0.000	3.569
身份类型	0.186	0.139	1.790	1	0.181	0.831
检查站过检次数	0.131	0.023	30.997	1	0.000	1.140
居住地类型	0.680	0.249	7.458	1	0.006	1.975
是否携带可疑物品	0.434	0.253	2.953	1	0.086	1.544
涉案次数	0.057	0.053	1.158	1	0.282	1.058
关押天数	0.004	0.001	12.014	1	0.001	1.004
是否与涉恐重点人员同背景关系	0.946	0.279	11.469	1	0.001	2.576
是否与涉恐重点人员同住宿	0.644	0.248	6.739	1	0.009	1.904
常量	-4.230	0.324	170.828	1	0.000	0.015

总体预测准确率为 84.4%,拟合情况较好。最后,将上述 9 个指标变量的证据权重值 WOE 代替原值,再次进行二元逻辑回归拟合,结果如表 7 所示。同时,将训练样本带回

Table 6 Prediction accuracy of brought-back test for training samples

表 6 训练样本带回检测预测准确性

涉恐人员	预测非恐怖分子	预测恐怖分子	准确率/%
样本非恐怖分子	651	99	86.8
样本恐怖分子	57	193	77.2
总准确率			84.4

检验得到证据权重逻辑回归模型预测准确性,如表 8 所示。由此可知,模型对训练样本的总体预测准确率为 88.9%,拟合情况相较于原模型更优。

Table 7 Evidence weight binary logistic regression analysis results

表 7 证据权重二元逻辑回归分析结果

项	B	标准误差	瓦尔德	自由度	显著性	EXP (B)
是否有宗教信仰-WOE	1.341	0.280	19.619	1	0.000	3.460
身份类型-WOE	0.272	0.139	1.528	1	0.216	0.842
检查站过检次数-WOE	0.132	0.023	27.835	1	0.000	1.131
居住地类型-WOE	0.722	0.250	6.203	1	0.013	1.862
是否携带可疑物品-WOE	0.436	0.252	2.920	1	0.087	1.539
涉案次数-WOE	0.058	0.053	1.131	1	0.287	1.058
关押天数-WOE	0.004	0.001	13.353	1	0.000	1.004
是否与涉恐重点人员同背景关系-WOE	0.955	0.279	11.757	1	0.001	2.599
是否与涉恐重点人员同住宿-WOE	0.674	0.247	6.839	1	0.009	1.909
常量	-4.152	0.321	167.794	1	0.000	0.016

Table 8 Evidence weight training samples brought-back model detection accuracy

表 8 证据权重训练样本带回模型检测准确性

涉恐人员	预测非恐怖分子	预测恐怖分子	准确率/%
样本非恐怖分子	675	75	90.0
样本恐怖分子	36	214	85.6
总准确率			88.9

本文为了更好地判断两种模型在涉恐人员身份预测中对恐怖分子的区分能力,选用 ROC 曲线进行检验,如图 4 所示。由此可见,证据权重逻辑回归模型对涉恐人员身份预测的效果相较于逻辑回归模型更优。

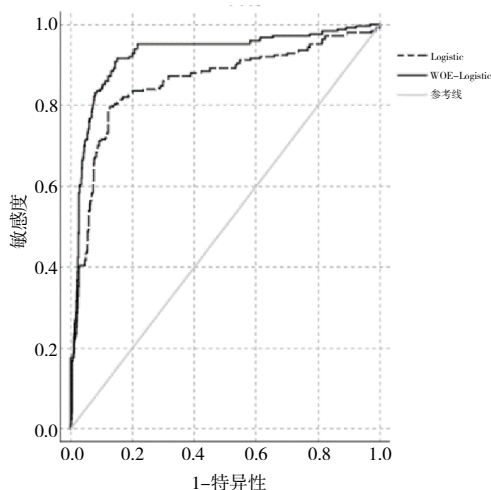


Fig. 4 Receiver operating characteristic curve of training sample

图 4 训练样本 ROC 曲线

为了进一步证明证据权重逻辑回归模型对涉恐人员身份预测的能力,选取 250 个测试样本再次带入两个模型进行验证,模型准确性如表 9、表 10 所示。测试样本的 ROC 曲线如图 5 所示。由此可见,证据权重逻辑回归对测

试样本数据的精准度相较于逻辑回归更优,拟合效果更好。

Table 9 Test accuracy of brought-back logistic regression model for test samples

表9 测试样本带回逻辑回归模型检测准确性			
涉恐人员	预测非恐怖分子	预测恐怖分子	准确率/%
样本非恐怖分子	142	21	87.1
样本恐怖分子	14	73	83.9
总准确率			86.0

Table 10 Test accuracy of brought-back evidence weight logistic regression model for test samples

表10 测试样本带回证据权重逻辑回归模型检测准确性			
涉恐人员	预测非恐怖分子	预测恐怖分子	准确率/%
样本非恐怖分子	151	12	92.6
样本恐怖分子	9	78	89.7
总准确率			91.6

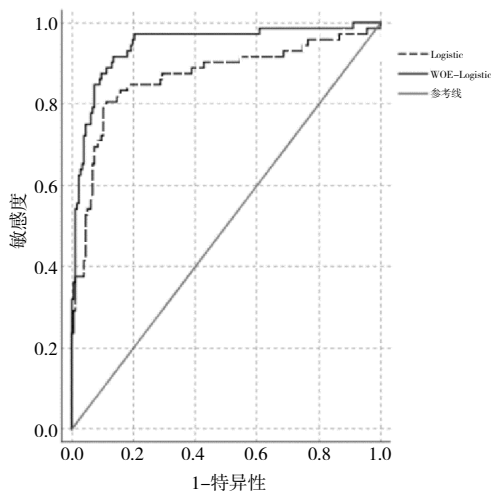


Fig. 5 ROC curve of test sample
图5 测试样本的ROC曲线

3.5 模型比较

在相同训练集和测试集下,比较分析Probit、决策树、随机森林等模型对涉恐人员预测的准确性,如表11所示。由此可知,证据权重逻辑回归模型的预测准确率优于其他预测算法,进一步证明了本文模型在涉恐人员身份预测方面的适用性更优。

4 结语

本文针对涉恐人员身份预测问题,从人员、地点、物品、案(事)件、组织关系五大维度,全面设计、构建了涉恐人员标签体系,利用标签体系精确、全面地描述涉恐人员的数据特征,以提升模型数据处理和预测效率。

同时,本文提出基于证据权重逻辑回归的身份预测模

Table 11 Comparison of prediction accuracy of different models

表11 不同模型预测准确率比较 (%)	
模型名称	预测准确率
Probit	83.8
决策树	84.1
随机森林	84.9
逻辑回归	86.0
证据权重逻辑回归	91.6

型,通过计算WOE、IV值以有效解决逻辑回归中的指标变量选择和虚拟变量陷阱问题,避免模型发生过度拟合情况,进而提升模型预测的准确度。

实验表明,该模型的总体预测准确性相较于其他预测模型更优,预测方法切实可行、有效,在反恐领域具有较好的应用前景。

参考文献:

- [1] HU X. Application of data mining method in anti-terrorism intelligence analysis[J]. Information Research, 2017(11):1-6.
扈翔. 数据挖掘方法在反恐情报分析中的应用研究[J]. 情报探索, 2017(11):1-6.
- [2] WANG B, WANG Y J. Study on the cause model of safety & security intelligence failure[J]. Information Studies: Theory & Application, 2021, 44(1):36-41.
王秉,王渊洁. 安全情报失误致因模型研究[J]. 情报理论与实践, 2021,44(1):36-41.
- [3] HUANG B. Extraterritorial practice and enlightenment of risk assessment of terrorists and extremists[J]. Journal of People's Public Security University of China(Social Sciences Edition), 2020,36(6):19-28.
黄彬. 恐怖分子、极端分子危险性评估的域外实践及启示[J]. 中国人民公安大学学报(社会科学版),2020,36(6):19-28.
- [4] LYU M C. Psychological research on the correction of terrorist offenders [J]. Chinese Criminology Review, 2015(6):27-34.
吕美琛. 恐怖主义罪犯矫正心理学研究[J]. 犯罪研究,2015(6): 27-34.
- [5] LI Y, WU S Z. Vehicle-ramming attack labeling early warning system based on behavioristic psychology[J]. Journal of Henan Police College, 2019,28(5):12-22.
李扬,吴绍忠. 基于行为主义心理学的驾车冲撞恐怖袭击标签预警体系[J]. 河南警察学院学报,2019,28(5):12-22.
- [6] ZHU Z H, YU X Y, LI C. Application and analysis of personnel background tag based on big data of public security[J]. Computer Knowledge and Technology, 2018,14(21):28-30.
朱振华,于晓昀,李超. 基于公安大数据的人员背景标签应用分析与研究[J]. 电脑知识与技术,2018,14(21):28-30.
- [7] ZHANG D W. Research on the construction and application of police data tagging[J]. Police Technology, 2017(6):38-41.
张大伟. 公安数据标签化建设应用研究[J]. 警察技术,2017(6): 38-41.
- [8] LI Y N. Association analysis of hierarchical attributes in the field of counter-terrorism intelligence based on multi-level association rule mining[J]. Information Science, 2021,39(11):127-132.

- 李勇男. 基于多层次关联规则挖掘的反恐情报跨层特征关联分析[J]. 情报科学, 2021, 39(11): 127-132.
- [9] OLABANJO O A, ARIBISALA B S, MAZZARA M. An ensemble machine learning model for the prediction of danger zones: towards a global counterterrorism[J]. *Soft Computing Letters*, 2021, 3: 100020.
- [10] SAIDI F, TRABELSI Z. A hybrid deep learning-based framework for future terrorist activities modeling and prediction[J]. *Egyptian Informatics Journal*, 2022, 23(3): 437-446.
- [11] GUO X, WU W H, XIAO Z T, et al. Anti-terrorism intelligence mining based on deep learning and open source information [J]. *Information Studies: Theory & Application*, 2017, 40(9): 135-139.
- 郭璇, 吴文辉, 肖治庭, 等. 基于深度学习和公开来源信息的反恐情报挖掘[J]. 情报理论与实践, 2017, 40(9): 135-139.
- [12] SUN F F, LIN P, CAO Z. Research of an rotation forest based on ensemble classifier in terrorists entity mining[J]. *Intelligence magazine*, 2015, 34(5): 190-195.
- 孙菲菲, 林平, 曹卓. 基于旋转森林集成学习的涉恐实体挖掘研究[J]. 情报杂志, 2015, 34(5): 190-195.
- [13] LUO S L, LIU Z, GUO L, et al. Research on suspected culprit recognition based on probit[J]. *Transactions of Beijing Institute of Technology*, 2011, 31(11): 1337-1341.
- 罗森林, 刘峥, 郭亮, 等. 基于Probit的犯罪嫌疑人判定方法研究[J]. 北京理工大学学报, 2011, 31(11): 1337-1341.
- [14] YANG Y C, QIU B L. Design method of public security management information system based on "five elements" model[J]. *Journal of People's Public Security University of China(Science and Technology)*, 2001(6): 22-25.
- 杨永川, 仇保利. 基于“五要素”模型的治安管理系统设计方法[J]. 中国人民公安大学学报(自然科学版), 2001(6): 22-25.
- [15] LEI J B. Research on equipment state degradation assessment and trend prediction based on logistic regression and support vector machine [D]. Shanghai: Shanghai Jiao Tong University, 2008.
- 雷金波. 基于逻辑回归和支持向量机的设备状态退化评估与趋势预测研究[D]. 上海: 上海交通大学, 2008.
- [16] XIANG Y T, WANG M, CAO L. Financial risk prediction method based on multi-objective deep learning model[J]. *Statistics & Decision*, 2022, 38(10): 184-188.
- 向有涛, 王明, 曹琳. 基于多目标深度学习模型的财务风险预测方法[J]. 统计与决策, 2022, 38(10): 184-188.
- [17] WAN L Y. Educational forecasting in the age of algorithms and the transformation of its research paradigm [J]. *Journal of Distance Education*, 2022, 40(3): 35-44.
- 万力勇. 算法时代的教育预测及其研究范式转变[J]. 远程教育杂志, 2022, 40(3): 35-44.
- [18] GOOD I J. Weight of evidence: a brief survey[J]. *Bayesian Statistics*, 1983, 2: 249-269.
- [19] WANG J Z. Based on the weight of evidence logistic regression model to assess P2P company's credit risk [J]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2016.
- 王金珠. 基于证据权重逻辑回归模型的P2P公司信用风险评估[D]. 南京: 南京航空航天大学, 2016.
- [20] SUN G Y. Risk scoring model of illegal fund-raising based on weight of evidence logistic regression [D]. Chengdu: Southwestern University of Finance and Economics, 2018.
- 孙光焱. 基于证据权重逻辑回归的企业非法集资风险评分模型[D]. 成都: 西南财经大学, 2018.
- [21] GREIFF W R. Maximum entropy, weight of evidence and information retrieval [D]. Amherst: University of Massachusetts Amherst, 1999.
- [22] LI H. Statistical learning method [M]. Beijing: Tsinghua University Press, 2012.
- 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.

(责任编辑: 刘嘉文)