

监所在押人员多维行为数据 AI 风险建模与预警机制

符浩源 陈建晖 郭宇锋 冼均源

广东警官学院 侦查学院

DOI:10.32629/ej.v9i6.3567

[摘要] 针对监所在押人员的安全隐患,传统的人工监督方式存在着依靠经验判断、风险识别滞后、覆盖面窄等问题,已不能满足目前的精细化监督需求。针对上述问题,本研究以多维的行为数据为基础,通过对在押人员数据采集,运用AI技术对可能存在的安全风险进行评价,从而实现其风险的早期识别与精确预警。通过对监所安全监控的智能化支撑,有效地提升了监所监管的主动性与预防性管控水平,协助维持监所秩序和社会治安。

[关键词] 监所在押人员; 多维行为数据; AI风险建模; 预警机制

中图分类号: X820.4 **文献标识码:** A

AI Risk Modeling and Early Warning Mechanism Based on Multi-dimensional Behavioral Data of Detainees in Prisons and Detention Centers

Haoyuan Fu Jianhui Chen Yufeng Guo Junyuan Xian

School of Investigation, Guangdong Police College

[Abstract] Traditional manual supervision for potential safety hazards of detainees relies heavily on empirical judgment, which is plagued by delayed risk identification and limited coverage, failing to meet the current requirements of refined supervision. Based on multi-dimensional behavioral data of detainees, this study collects relevant data and adopts artificial intelligence technology to assess potential safety risks, so as to realize early identification and accurate early warning of risks. As intelligent support for prison safety monitoring, the proposed mechanism effectively improves the initiative and preventive management capacity of prison supervision, and helps maintain the order of detention facilities and social security.

[Key words] detainees in prisons and detention centers; multi-dimensional behavioral data; AI risk modeling; early warning mechanism

引言

监所安全作为社会治安的一项重要内容,随着监管领域标准化和管理要求的不断提高,以往的经验评价都是建立在干警人工观测和风险防控模式之上,这种方法不但耗费人力物力,还会因为个人认知差异、注意力疲劳等原因造成错误判断,很难在短时间内察觉到自伤、自残、暴力冲突、逃逸等风险,风险管理多为被动应对。近年来,随着人工智能、物联网等传感技术的迅速发展,监所智能化监控成为必然趋势。融合多源异质行为数据和人工智能技术,建立基于数据驱动的风险预警体系,是目前监所安全管理研究的一个重要发展趋势。

1 机制建设背景与意义

在监所羁押量居高不下、在押人员结构日趋复杂的情况下,我国监所管制模式在近几年出现了新的变化,并面临新的挑战。在押人员危险行为具有突发性、隐秘性和复杂性等特征,传统识

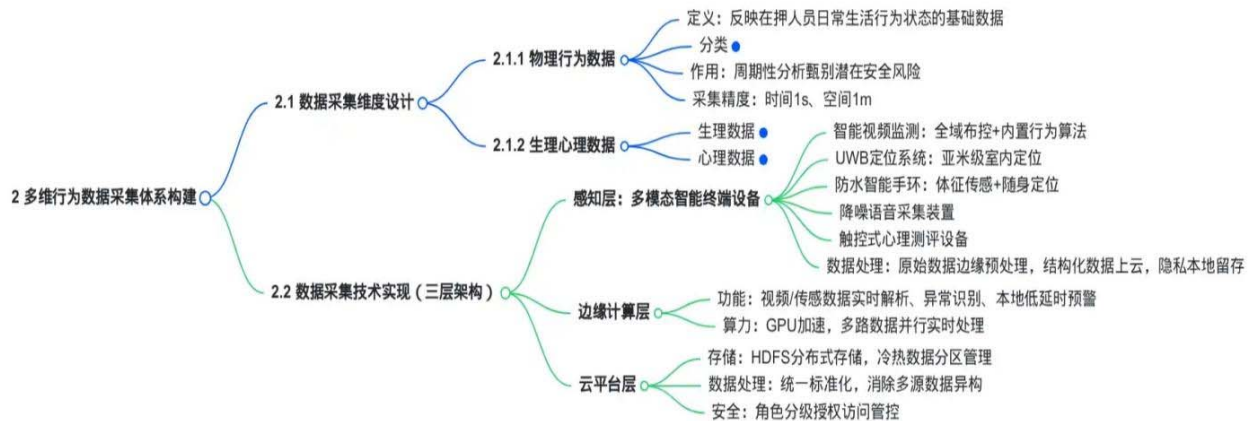
别方法主要依靠人工观察和经验判断,存在反应滞后、检测偏差大等问题,已不能适应现代化监所的需要。

在人工智能飞速发展的今天,以大数据、机器学习为基础,建立智能化风险预警体系已成为国际监所管理现代化的重要趋势。在此基础上,融合人体行为数据,构建基于多维行为数据的人工智能风险模型与预警机制。本项目采用人工智能算法,实现监所在押人员身体、心理等行为的提前发现与精确预警,积极高效地提升监所管理水平。该机制的构建,既可以减少安全事故的发生,保障监管体系的稳定性,又可以使警力配置由“被动处置”向“主动预防”过渡,推动监所管理向智能化、精细化方向发展。^[1]

2 多维行为数据采集体系构建

2.1 数据采集维度设计

2.1.1 物理行为数据



如图1 多维行为数据采集体系构建

物理行为数据是一种基本的、能够反映在押人员日常生活行为状态的基础数据,主要分为三种类型:空间移动轨迹、行为特征和活动规律模式。空间移动轨迹数据利用卫星定位技术获取在押人员在羁押区域的实时位置、移动速度、移动范围及逗留时间等信息。行为特征数据采用影像学方法获取行为特性,包括肢体行为(如姿势不正常、暴力行为、自残行为)、面部表情(情绪变化、异常表情)以及异常行为(如长期卧姿、频繁踱步、夜间异常活动等)。活动规律模式数据包含每日的生活行为数据,如对日常生活的遵守、饮食与睡眠习惯、对物件使用等。通过对这些指标的周期性变化进行分析,可以有效地识别出潜在的风险。物理行为数据采集需要实现1秒的时间精度和1米的空间精度,保证行为特征的精确捕获。

2.1.2 生理心理数据

生理、心理数据是评估在押人员身体健康状况的重要基础。生理指标数据采用可穿戴式智能终端与生物特征相结合的方式,采集人体生理指标,如心率、血压、皮电、体温、呼吸频率等,并结合睡眠结构(睡眠时间、睡眠时长、睡眠质量)与运动强度等。心理状态数据采用结构化心理测评量表、对话记录分析及行为观测记录等方法,考查心理状态(焦虑、抑郁、愤怒等)、认知特征(思考方式、注意程度)、个性特征及自杀危险度。按照各项指标特征,设定生理数据的采集频次^[1]。每隔一小时进行一次静态指标的收集,对动态指标进行实时监测,对精神状态进行每月的全面评价,在特定环境中,可以进行实时采集。

2.2 数据采集技术实现

多维行为数据采集体系采用“智能感知终端+边缘计算+云平台”三层结构的多维行为数据获取系统。在感知层面,利用多模态智能终端,包含智能监测(包含行为分析算法,覆盖监控室、公共区域及热点区域),定位系统(基于超宽带室内定位技术,可实现亚米级定位精度),智能手环(内置心率、体温、动作等传感器,可防水可移动定位),语音采集装置(降噪语音识别收发装置),心智测量装置(触控式心理量表填写设备)等。所有感知终

端采集到的原始数据会先传输至边缘计算节点,在本地完成数据初步清洗、格式标准化和特征预处理,仅将处理后的结构化数据上传至云平台存储,既降低了数据传输带宽压力,也能实现敏感数据的本地留存,保障在押人员隐私数据安全。

在边缘计算层次上,通过构建边缘计算节点,对视频流、传感器数据进行实时分析,对关键行为、异常情况进行识别,进行低延时局部预警。在此基础上,利用GPU加速能力,实现对多路视频流及传感器数据的并行处理,达到了实时性的要求。

在云平台层次上,建立统一的数据管理中心,采用分布式存储结构(HDFS)对海量数据进行存储,并进行全寿命管理,从而达到对冷、热数据的隔离。通过对采集到的数据进行规范化处理,并解决数据的异构性问题。设定数据访问控制机制,根据使用者角色分配数据访问权限,以保证数据的安全性。

3 AI风险建模核心技术

3.1 行为特征工程

3.1.1 静态特征提取

静态特征提取是指从多个维度的行为数据中抽取出具有较强稳定性的特征,从而建立起在押人员的基准行为模式。在此基础上,通过对不同类型行为数据的分析行为特征(活动范围偏好、空间使用效率等)、时间规律(如日常活动、活动高峰时段等)以及行为模式(如行为习惯、物品使用偏好等)。通过对数据处理,得到行为指标的均值、方差和分位数,建立行为基准的参数。生理心理数据的静态特征包括正常心率区间、平均血压等、心境特征包括个性特征、情绪稳定性及心理危险程度。在此基础上,通过长期跟踪研究,构建基于个体生理、心理特点的标杆模式^[2]。

3.1.2 动态特征提取

动态特征提取主要研究行为数据的时间序列变化及异常规律,以发现潜在的危险前兆。针对物理行为数据,利用时序分析技术,从行为时序规律、异常事件频次(比如违规次数的变化)和行为突变性(比如活动量的骤升或骤降)等。利用滑动窗口技

术,将时间序列数据分割成若干段,并在此基础上计算出各时间窗口的变动率及各时段的波动性。生理心理动态特征其主要表现为:生理指标(心率变异、血压波动幅度等)的变化规律;情感状态的时间演化(情绪稳定性的改变,负性情绪持续时间)和应激响应的特点。在此基础上,利用趋势分析及突变检测等方法,辨识出各生理、心理指标的变异点。

3.2 风险预测模型构建

3.2.1 基础模型架构

风险预测模型采用多模态融合深度学习架构,由数据输入层、特征融合层、风险预测层和结果输出层组成。数据输入层接受多个维度的特征,其中包含静态特征矢量和一系列动态特征。在特征融合层次上,利用注意力机制对多个特征进行有效的融合。在此基础上,通过自注意力模型对各特征向量进行重要性加权,并利用多个模态间的关联关系,形成一个融合特征矢量^[3]。

风险预测层设计一个多任务学习框架,通过多任务学习方法,实现对自我伤害风险、暴力行为风险、逃跑风险、心理危机风险等多类风险的同步预测。每个风险都有一个独立的预测子网络,该子网络采用二维LSTM网络对预测子网络进行建模,并将其与完全连通层相融合,得到相应的风险概率。在模型框架中加入领域知识约束,用规则化术语把管理经验转换成模型的约束,可以有效地提高模型的可信度。该基础模型以模块化方式实现了对不同类型风险的柔性扩充,以及对模型参数的自主优化。

3.2.2 模型训练优化

模型训练采用两阶段训练策略:第一阶段使用标记数据进行监督训练,第二阶段结合未标记数据进行半监督学习,以提高模型的泛化能力。训练数据采取分层抽样的方式,保证各种危险样本的均衡分布,并利用数据强化技术(包括时序数据再抽样、特征扰动等)扩大训练样本。优化算法利用AdamW优化器建立动态的学习速率调度机制,使其具有0.001的初始学习率,并利用余弦退火的方法来调节学习速率。损失函数被设计为加权交叉熵损失,根据不同风险类型的重要性设置权重系数,为高风险样本赋予更高的权重。如图1,风险预测模型。

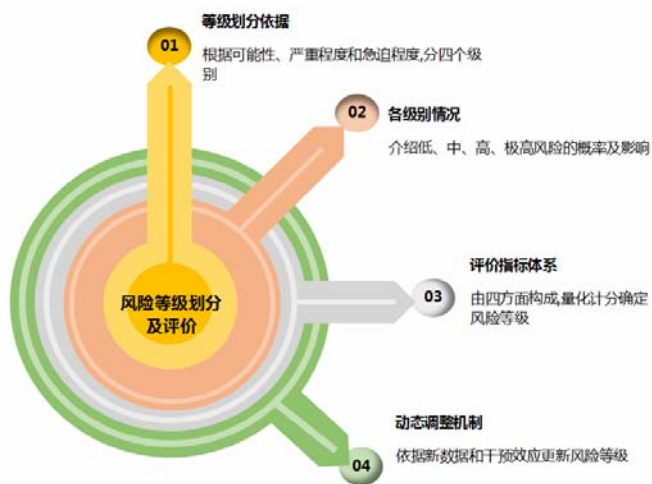


如图2 风险预测模型

4 风险预警与处置机制

4.1 风险等级划分

风险等级划分根据风险的可能性、后果的严重程度和紧迫程度,运用多维评估方法对风险水平进行分级,并将其划分为低、中、高、极高四个级别。低风险(蓝色警告)表示在押人员行为表现正常,有轻度反常但无实际危险,发生概率小于5%,影响范围只局限于个体。中度风险(黄色警告)表明有显著的不正常行为,风险概率介于5%—30%,可能会对个人或小范围造成影响,应引起高度重视。高风险(橙色预警)是指风险较大(30%—70%),可能造成人员伤亡或造成当地社会秩序混乱,急需紧急干预。极高风险(红色预警)是指发生的可能性较大(70%以上),造成后果非常严重,有可能造成人员伤亡或群体性事件,因此有必要启动应急预案。风险等级评价指标体系由行为异常、生理心理状态、历史风险记录、环境影响因素四个方面构成。每一个维度均设有量化计分准则,并以加权求和计算综合风险得分,根据得分区间确定风险等级^[4]。在此基础上,构建风险等级动态调整机制,依据最新采集的数据及干预效应,实时更新风险等级,对风险进行动态追踪。如图2,风险等级划分。



如图3 风险等级划分

4.2 预警响应流程

预警响应流程采用分级响应机制,确定应对措施以及在不同的风险等级下的时限需求。针对低风险监察组民警会作出反应,其中包括加强监控、谈话、行为控制,在2个小时之内作出反应,并建立每天观察报告。中风险预警由监区领导组织响应,对风险发出警告,并开始实施危险介入程序,其中包括加大检测次数;对监舍安排、心理辅导等措施进行调整,并在30分钟内完成响应,并建立专门的跟踪记录。高风险预警通过监所管理部门牵头,组织专家(咨询人员、医务人员)介入,实施应急管理(如重点监控、个人护理等),15分钟后,每4小时上报处置进展情况。高危险预警将在第一时间启动,并由监所主管领导,实施必要隔离,如有需要,可向外界求助(公共安全和医疗部门),响应时间5分钟,并对处置过程进行实时反馈。对预警进行闭环管理,由接收预警、确认信息、采取措施、执行跟踪、效果评价、预警解除

六个环节组成。建立一套记录反应进程的体系,对每一个过程进行详尽的记载,并编制出处置档案。

4.3 应急联动机制

应急联动机制建立“内外协同”的二级联动系统,保证对突发事件的快速处置。在团队协作层面,建立“指挥中心—监区—监控中心—医疗—心理咨询处”的协作机制,确定各部门间的职责分工与协作。指挥中心是集信息、资源分配、指导于一体的枢纽。监控中心对现场视频进行实时监控;提供医学及心理咨询室的专业技术支援,由管理层负责场地的清理工作。内部联动与应急指挥平台之间的信息交互与协同工作,最多可在3分钟内做出响应。外部机制与公安、消防、卫生、司法等部门建立联动机制,签订应急联动机制,明确联动的触发条件、响应程序及资源保障。建立对外联络表,以保证发生突发事件时,能迅速与相关部门取得联系。针对各种类型的紧急情况,如暴力攻击、自杀伤害、火灾等,建立特殊协作方案,并定期进行联合演练。应急联动机制也包含一套保障资源、必要应急装备(防护装备、医疗急救设备、通信工具等)储备、组建应急力量(应急指挥部、医学救援队等)、定期开展应急训练与演练的制度^[5]。

5 结束语

综上所述,对在押人员进行风险管理,是保证监所安全与稳定的重要一环。本项目建立了多维行为数据获取体系、人工智能风险预测模型以及全程预警处理机制,突破了以往基于人工经验进行风险辨识的局限,通过客观数据,实现对在押人员风险

的动态精确辨识。该机制将人工智能技术与监管业务深度结合,有助于监管人员对可能存在的风险进行预警、缓解工作压力,提升监所风险的智能化、精细化管理水平。未来,结合生物传感与大尺度建模等手段,实现预测准确性与解释性的持续优化,为监所安全管理提供强有力的技术支撑。

[基金项目]

本文系广东警官学院国家级大学生创新创业项目“AI驱动的监所在押人员安全风险评估及精准预警创新研究”(项目编号:202511110006)研究成果之一。

[参考文献]

- [1]韩微晓.“派驻+巡回”监所检察监督机制研究[D].河南财经政法大学,2025.
- [2]张玲.上海公安“智慧监所”建设中存在的问题及对策研究[D].华东师范大学,2024.
- [3]李善敏.“监所服务E通办”的实用性探究——以温州公安监管部门先行先试为例[J].公安研究,2024,(04):65-70.
- [4]郝庆浩.监所风险预测及安全巡检规划方法研究[D].山东大学,2023.
- [5]李强,赵尚上,李胜广.基于半监督机器学习的监所人员风险计算[J].警察技术,2022,(01):25-28.

作者简介:

符浩源(2004—),男,汉族,广东湛江人,本科在读,侦查学专业。