

大数据背景下数据挖掘技术在公安安保维稳和案件侦查中的思考与实践

冉彤绯¹ 田海²

1. 天津警察学院 天津 300381

2. 天津市公安局和平分局 天津 300021

摘要: 在大数据时代, 数据挖掘技术通过整合多源异构数据, 构建公安大数据平台, 实现犯罪预测、嫌疑人识别、交通管理优化及反恐预警。例如, 利用聚类分析锁定高风险群体, 关联规则挖掘犯罪网络, 时间序列预测盗窃高发时段。技术实践涵盖数据清洗、机器学习建模及可视化决策支持, 显著提升公安部门主动防控与精准打击能力, 推动警务模式向智能化、预防性转型。

关键词: 大数据背景; 数据挖掘技术; 公安安保维稳; 案件侦查; 实践

引言: 在数字化浪潮席卷的当下, 大数据已深度融入社会各领域, 公安工作亦面临全新挑战与机遇。海量数据蕴含着犯罪规律、社会动态等关键信息, 数据挖掘技术作为解锁这些信息的钥匙, 能从中提炼有价值知识。其在公安安保维稳中可预测风险、提前防范, 于案件侦查里能精准锁定嫌疑人、还原犯罪轨迹, 对提升公安工作效能、维护社会稳定意义重大。

1 大数据与数据挖掘技术基础

1.1 大数据技术架构

(1) 数据采集: 作为数据链路的起点, 全面整合多元数据源, 涵盖物联网设备实时采集的现场环境、人员轨迹等感知数据, 社交媒体平台的舆情信息、社交动态等非结构化数据, 以及公安业务系统积累的案件记录、户籍信息、涉案档案等结构化数据, 实现数据的全维度汇聚。(2) 数据存储: 采用分布式存储架构, 依托Hadoop、HBase等技术实现海量数据的安全存储与高效管理; 同时搭配Flink等实时计算框架, 满足公安工作中即时数据处理的需求, 实现批量数据与实时数据的协同处理^[1]。(3) 数据治理: 聚焦数据质量提升, 通过数据清洗剔除冗余、错误数据, 开展数据标准化处理实现数据格式统一, 建立全流程质量管控机制, 保障数据的准确性、完整性和可用性, 为后续分析应用奠定基础。

1.2 核心数据挖掘技术

(1) 关联分析: 运用Apriori等算法挖掘犯罪行为间的隐藏关联, 梳理犯罪模式与规律, 为案件串并提供支撑。(2) 聚类分析: 通过聚类算法对人员、事件等数据进行分组, 实现高危人群精准识别与群体事件提前预警, 辅助防控部署。(3) 分类与预测: 基于SVM、随机森林

等模型, 结合历史案件数据构建分类预测模型, 实现案件类型预判, 提升侦查针对性。(4) 时空分析: 融合GIS技术与热力图分析, 绘制犯罪热点地图, 直观呈现犯罪空间分布与时间规律, 优化警力投放^[2]。(5) 图挖掘: 开展社交网络分析, 构建犯罪团伙关系图谱, 清晰呈现团伙成员关联, 助力打掉犯罪网络核心。

1.3 技术选型原则

(1) 实时性需求: 针对突发事件预警、实时轨迹追踪等场景, 优先选择高并发、低延迟的技术方案, 确保信息处理与响应的及时性。(2) 准确性需求: 在嫌疑人身份匹配、证据链验证等关键环节, 选用精度高、稳定性强的技术模型, 保障分析结果的可靠性。(3) 可解释性需求: 围绕侦查决策依据提供、案件结论论证等需求, 选择逻辑清晰、结果可追溯的技术, 确保分析过程与结论可解释、可验证, 契合司法规范要求。

2 大数据背景下数据挖掘技术在公安安保维稳中的应用实践

2.1 社会面安全态势感知

(1) 多源数据融合: 整合舆情、交通、气象、重点人员轨迹等多维度数据, 构建全域数据资源池。通过数据清洗、格式标准化等预处理流程, 消除多源数据异构性, 实现跨领域数据关联分析。例如, 将社交平台舆情数据与交通流量数据联动, 捕捉人员聚集相关舆论苗头与实地流动趋势的关联性, 为安全态势研判提供全面数据支撑^[3]。(2) 风险评估模型: 采用基于加权评分法的维稳等级划分模型, 结合历史安保数据与当前多源数据特征, 确定各风险指标权重。从人员聚集密度、舆情风险等级、交通拥堵程度、气象灾害隐患等维度设定评分标准,

通过数据挖掘算法计算综合风险分值，划分高、中、低三个维稳等级，为安保资源调配提供量化依据。(3) 案例：在重大活动安保中，依托数据挖掘技术开展实时人流预测与疏导。通过融合活动场馆周边交通卡口数据、公共交通出行数据、手机信令数据，构建人流预测模型，实时推演不同时段人流聚集规模与流动方向。针对预测的高风险聚集区域，提前部署警力疏导，动态调整交通管制措施，有效防范拥挤踩踏等安全隐患，保障活动顺利开展。

2.2 群体性事件预警与处置

(1) 异常行为检测：通过文本挖掘技术分析网络言论情绪倾向，构建舆情关键词库与情绪识别模型。对社交平台、论坛、短视频等平台的言论数据进行实时抓取与挖掘，识别极端情绪、煽动性言论等异常信息，精准定位潜在群体性事件苗头，为早期预警提供线索。同时，结合线下人员聚集数据，实现线上线下异常行为的协同检测。(2) 传播路径模拟：基于图模型构建谣言扩散路径预测模型，将用户作为节点、用户间互动关系作为边，通过数据挖掘算法分析谣言传播的关键节点与传播规律。预测谣言在不同时段的传播范围与影响程度，识别谣言传播的核心载体，为精准辟谣、切断传播链条提供技术支撑^[4]。(3) 案例：在某地劳资纠纷事件中，运用数据挖掘技术实施早期干预策略。通过文本挖掘监测到企业员工群体在网络上的负面情绪聚集与维权言论扩散，结合企业劳资历史纠纷数据与员工流动数据，预判可能发生的群体性上访事件。公安部门联合相关部门提前介入，搭建沟通平台，化解劳资矛盾，成功避免了事件升级，实现了从“被动处置”向“主动预防”的转变。

2.3 重点人员动态管控

(1) 行为画像构建：结合消费、出行、通信等多维度数据，通过数据挖掘技术为重点人员构建全面行为画像。提取消费习惯、出行轨迹、通信联系人、活动时段与场所等特征，生成多维度标签，如“高频出入敏感场所”“跨区域频繁流动”等，精准刻画重点人员行为模式，为管控提供精准靶向。(2) 风险评分系统：基于机器学习算法构建再犯概率预测模型，整合重点人员历史违法犯罪数据、行为画像数据、社会关系数据等，通过分类算法训练模型，预测重点人员再犯风险概率。根据风险概率划分高、中、低风险等级，实施差异化管控措施，提升管控效率与精准度。(3) 案例：在涉恐人员管控中，依托数据挖掘技术开展轨迹追踪与行动预判。通过融合交通卡口数据、高铁民航购票数据、住宿登记数据、手机信令数据，构建涉恐人员轨迹追踪模型，实时

掌握其活动轨迹与同行人员。结合涉恐人员行为画像与历史活动规律，预判其可能的行动意图与目标，提前部署防控力量，成功防范了潜在恐怖活动风险，保障了公共安全。

3 大数据背景下数据挖掘技术在公安案件侦查中的创新应用

3.1 案件线索关联分析

(1) 跨案件串并：基于时空、作案手法、物证特征开展关联挖掘，构建多维度案件特征体系。通过数据挖掘算法提取不同案件的案发时间窗、地点坐标、侵入方式、遗留物证等核心特征，计算特征相似度阈值，实现零散案件的自动串并。此举打破个案侦查壁垒，助力发现系列案件作案规律，提升侦查效率。(2) 隐匿关系发现：依托图数据库 (Neo4j) 挖掘嫌疑人关联网络，将嫌疑人、受害者、涉案场所、资金账户等要素作为节点，以通信、资金往来、同行轨迹等为边构建关系网络。通过社团检测、最短路径分析等算法，精准识别隐藏的同伙、中间人等关联人员，破解传统侦查中“看不见的关系”难题。(3) 案例：在系列盗窃案侦查中，通过关联挖掘实现嫌疑人身份突破性锁定。对多起入室盗窃案的案发时段、作案目标选择、撬锁手法等特征串并后，利用图数据库分析涉案区域监控中频繁出现的人员轨迹，关联其通信记录与前科人员数据，最终锁定跨区域流窜作案团伙，成功抓获3名嫌疑人，破获系列案件12起。

3.2 犯罪模式识别与预测

(1) 犯罪热点迁移预测：基于时间序列分析构建动态热力图，整合历史案件数据、人口流动、节假日、天气等多维度数据，通过ARIMA、LSTM等算法预测不同时段、区域的犯罪发生概率。生成动态犯罪热力图，直观呈现热点区域迁移趋势，为前瞻性防控提供精准指引。(2) 作案手法演化分析：通过文本挖掘提取案件卷宗、报案记录中的关键信息，生成作案手法特征词云。追踪不同时期犯罪手法的关键词变化，分析诈骗话术、盗窃工具、洗钱方式等演化规律，提前预判新型犯罪形态，为侦查人员提供针对性识别策略^[5]。(3) 案例：在电信诈骗案件侦查中，通过话术模式识别精准打击犯罪。对海量报案记录、涉案通话录音进行文本挖掘，提取高频话术关键词与句式结构，总结出“冒充公检法”“虚假投资”等6类主流诈骗话术模式。据此构建话术识别模型，联动运营商精准拦截涉案通话，协助锁定诈骗窝点15处。

3.3 侦查资源智能调度

(1) 警力分配优化：基于遗传算法规划巡逻路线，结合犯罪热点预测数据、巡逻区域交通状况、警力配置

基数等约束条件,构建巡逻效率最优目标函数。通过遗传算法迭代优化巡逻路线与警力部署方案,实现重点区域警力全覆盖,非热点区域高效巡查,提升警力使用效能。(2)多部门协同机制:依托数据共享平台实现联合侦查,打破公安内部刑侦、网安、经侦等部门及跨区域公安、银行、通信运营商的数据壁垒。通过数据挖掘技术整合各部门涉案数据,实现线索实时共享、同步分析,构建“数据驱动”的多部门协同侦查链条。(3)案例:在跨区域贩毒案件协同打击中,数据共享平台发挥关键作用。通过平台整合沿线多省市公安的卡口数据、涉毒人员轨迹、资金流向等数据,利用数据挖掘锁定贩毒网络关键节点与运输路线。多部门同步部署警力,在贩毒交易关键环节实施精准打击,成功摧毁跨5省的贩毒网络,抓获涉案人员28名。

4 公安大数据应用中的挑战与优化对策

4.1 技术挑战

(1)数据隐私与安全:公安数据包含大量个人敏感信息,隐私泄露风险较高。当前虽采用匿名化处理遮蔽身份标识,但仍存在去匿名化破解隐患;区块链存证技术虽能保障数据溯源与不可篡改,却面临与现有数据管理体系的融合适配难题,如何平衡数据利用与隐私保护成为核心技术瓶颈。(2)数据质量:公安多源数据存在明显的缺失值与噪声数据问题,如部分案件卷宗关键信息缺失、交通卡口数据受环境干扰出现误差。现有缺失值填补与噪声过滤方法在处理异构数据时效果有限,易导致数据挖掘结果偏差,影响后续决策的准确性。(3)算法可解释性:深度学习等黑箱模型在公安大数据分析中应用广泛,但模型决策过程难以可视化解。公安执法决策需具备明确的逻辑依据,黑箱模型的不可解释性使其与公安决策的兼容性不足,限制了高风险场景下的技术应用。

4.2 管理挑战

(1)跨部门数据壁垒:公安内部情指、政保、刑侦、治安、信支、网安、交管等部门数据标准和管理尚未完全融合统一,外部与政法、通信、教育、医疗、金融等部门缺乏常态化共享机制。数据分散存储、格式异构导致协同分析效率低下,难以形成数据应用合力。(2)人员技

能缺口:公安队伍中兼具大数据技术与警务业务能力的复合型人才稀缺。现有警务人员多缺乏数据挖掘、算法应用等专业技能,技术人员又对警务工作流程尚未全面熟知,导致技术与业务融合脱节。

4.3 优化策略

(1)技术层面:引入联邦学习技术,实现多主体在隐私保护下的协同挖掘。通过模型参数共享而非原始数据传输,规避数据跨境流动风险,同时整合多部门数据资源提升分析精度,破解隐私保护与协同分析的矛盾。(2)管理层面:构建“数据-情报-行动”一体化作战平台。统一数据标准规范,打通跨部门数据链路,实现数据采集、分析、情报生成、警务行动的全流程闭环管理,推动数据价值向实战效能转化。(3)制度层面:完善公安大数据应用的法律法规体系。明确数据采集、使用、存储的权限边界,界定数据挖掘结果在执法中的证据效力,建立数据应用伦理审查机制,为公安大数据应用提供合法合规的制度保障。

结束语

大数据背景下,数据挖掘技术为公安安保维稳与案件侦查工作带来了深刻变革与创新发展。通过有效运用该技术,公安部门在风险预警、线索发现、犯罪打击等方面取得了显著成效。然而,数据安全、技术融合等问题仍待解决。未来,需持续优化技术手段,加强跨部门协作,培养专业人才,让数据挖掘技术在公安实战中发挥更大效能,为维护社会稳定、保障人民安全构筑更坚实的防线。

参考文献

- [1]韩丽莎.浅谈大数据与数据挖掘技术[J].计算机产品与流通,2020,(07):114+133.
- [2]吴燕珍.浅谈大数据中数据挖掘技术的应用[J].电脑知识与技术,2022,(03):20-21.
- [3]王好姝.基于大数据的数据挖掘技术与应用[J].数字技术与应用,2021,(06):125-127.
- [4]刘铭,吕丹,安永灿.大数据时代下数据挖掘技术的应用[J].科技导报,2020,36(09):78-81.
- [5]王红英.数据分析在大数据时代下的发展趋势及价值体现[J].科技创新与应用,2023,(04):165-167.