



刑事侦查专业研究系列丛书

大数据在刑事侦查中的应用研究

李建利 李宇尘 著



吉林大学出版社

大数据在刑事侦查中的 应用研究

李建利 李宇尘 著

 吉林大学出版社

图书在版编目(CIP)数据

大数据在刑事侦查中的应用研究/李建利,李宇尘
著.——长春:吉林大学出版社,2017.3
ISBN 978-7-5677-9248-7

I. ①大… II. ①李… ②李… III. ①数据处理-应用-刑事侦查-研究 IV. ①D918-39

中国版本图书馆 CIP 数据核字(2017)第 067766 号

书 名 大数据在刑事侦查中的应用研究
DASHUJU ZAI XINGSHI ZHENCHA ZHONG DE YINGYONG YANJIU

作 者 李建利 李宇尘 著
策划编辑 刘 佳
责任编辑 陈颂琴
责任校对 云 宇
装帧设计 卷墨堂
出版发行 吉林大学出版社
社 址 长春市朝阳区明德路 501 号
邮政编码 130021
发行电话 0431-89580028/29/21
网 址 <http://www.jlup.com.cn>
电子邮箱 jlup@mail.jlu.edu.cn
印 刷 北京市媛明印刷厂
开 本 787 mm × 1 092 mm 1/16
印 张 10.5
字 数 250 千字
版 次 2017 年 3 月第 1 版
印 次 2017 年 3 月第 1 次
书 号 ISBN 978-7-5677-9248-7
定 价 40.00 元

版权所有 翻印必究

前 言

大数据是指海量且类型复杂的数据,具有种类多、流量大、容量大、价值高的特点。大数据被称为信息化时代的“石油”。未来国家的核心竞争力将在很大程度上依赖数据转化为信息和知识的速度与能力,而这种转化的速度和能力,实际上取决于数据方面的技术和管理应用到公安刑事侦查决策方面文化的形成。

大数据时代的来临,对公安刑事侦查的管理和发展提出了新的挑战。当下,犯罪呈现出更加严峻和复杂的态势。首先是犯罪总量大,犯罪率逐年上升。据统计,仅2012年公安机关刑事案件立案的案件数为6551440起,检察机关批捕、决定逮捕犯罪嫌疑人的案件数为680539,人数为986056。近二十多年来,中国犯罪率呈逐年上升趋势,刑事案件立案数平均每年增长22%以上,超过了全国GDP的增长。其次是犯罪智能化。犯罪是一种社会存在,科学的发展渗透到犯罪的方方面面,提升了其能力和危害程度。这表现为两个方面:一是运用科学思维实施的犯罪,主要表现为犯罪思维严密,犯罪前经过周密部署和策划,犯罪过程渗透着科学思维和谋略。二是利用科学技术实施的犯罪,突出表现为数字化犯罪。以利用网络犯罪为例,2012年,全国公安机关累计破获涉网违法犯罪案件11.8万余起,抓获犯罪嫌疑人21.6万余人。据赛门铁克公司2012年9月发布的诺顿安全报告估算,2011年7月至2012年7月,中国有超过2.57亿人成为网络违法犯罪的受害者。网络违法犯罪所造成的直接经济损失达2890亿元人民币,受害者人均蒙受的直接经济损失约1200元人民币。再次,犯罪时空的复杂性。现代科技的发展,使得犯罪时间非线性,犯罪空间缺席性,时空组合的多维、多样化和任意性。最后,案件因果联系复杂。相对于传统的静态、单一社会来说,现代社会是一个动态、复杂社会。在动态、复杂社会中,因果联系具有非线性、偶合性、多因性、断裂性,犯罪的因果联系往往难以确定。但是现代社会几乎被数据化了,一旦你与现代社会系统进行交换,就很有可能被数据捕捉和记录。因此,在大数据时代,不要说数字化犯罪,即使传统手段的犯罪,都可以说落入了一种“天网恢恢,疏而不漏”的网络记录和存储体系,数据化就是当下犯罪的现实生态。

本书追溯了公安刑事侦查理论基础,阐释了大数据在刑事侦查应用的前景,解析了大数据时代和刑事侦查紧密结合的方法和方案,深刻剖析了典型案例,有针对性地提出了大数据在刑事侦查发展中的展望,希望对我国公安刑事侦查在大数据应用上提供具有借鉴价值的意见和建议。

在本书的写作过程中,得到了太原科技大学电子信息工程学院赵志诚教授的细心指导,在此一并表示感谢,敬请读者批评指正。

作者:山西警察学院 李建利

2016年12月15日

目 录

第一章 引言	1
1.1 研究背景及意义	1
1.2 研究现状	12
1.3 研究内容及思路	20
第二章 传统刑事侦查方法	22
2.1 刑事侦查学概述	22
2.2 刑事侦查鉴定技术	28
2.3 刑事侦查实验技术	57
2.4 刑事侦查辨认、询问与讯问	60
2.5 本章小结	70
第三章 大数据基本理论及其应用前景	72
3.1 大数据时代背景	72
3.2 大数据的基本概念及术语	78
3.3 大数据的研究方法	84
3.4 大数据的应用展望	102
3.5 本章小结	105
第四章 数据挖掘在刑事侦查鉴定中的应用	106
4.1 引言	106
4.2 痕迹检验的数据化研究	110
4.3 数据挖掘在指纹识别中的应用	117
4.4 数据挖掘在枪弹痕迹识别中的应用	121
4.5 本章小结	126
第五章 刑事侦查数据仓库的建立和应用	127
5.1 引言	127
5.2 刑事侦查数据仓库的设计	129
5.3 数据挖掘在刑事侦查数据仓库的应用	141
5.4 本章小结	147
第六章 大数据时代刑事侦查的创新和思考	149
6.1 大数据与侦查思维创新	149
6.2 数据认知与案件事实认定	154
6.3 数据伦理与公民法益维护	158
6.4 结语	163

第一章 引言

1.1 研究背景及意义

詹姆斯·邦德在失败数次后,终于攻入到敌人大本营,却惊讶地发现自己的生活起居、政府特工总部的密室等全都置于敌人的监控之中……

这是电影《007 幽灵党》中的一个镜头,电影中的“幽灵党”利用大数据技术,实时监控着特工组织的一举一动。电影中的情节看上去似乎离现实生活很远,但在现实生活中,绝大多数人都或多或少接触过大数据技术,如“猜你喜欢”、定位导航、实时路况、自动驾驶等。随着大数据的快速发展,如何保护安全隐私,也成为人们日益关注的话题。

“安全隐私问题,几乎成为现在阻碍大数据发展的唯一制约因素。”2016年12月8日,在中国大数据技术大会(BDTC)上,主办方中国计算机学会(CCF)的大数据专家委员会副秘书长潘柱廷指出,在如今的弱人工智能时代,大数据的应用与发展依旧缺乏法律的规范。

据了解,我国已经将大数据技术运用到刑侦领域。“大数据的重要作用就在于犯罪预测,通过对潜在的可能引发犯罪风险的因素进行采集和分析,来预测潜在的犯罪行为,预测潜在的违法者行为,并预测潜在的犯罪被害人。”北京航空航天大学副教授裴炜称,我国有些公安系统已经开始采用这项技术,比如,苏州公安犯罪预测系统,主要针对一些盗窃案件的实时警情监测,监督潜在的犯罪分子。

随着大数据技术的不断发展,犯罪预测技术必将不断完善,如今的弱人工智能化必然会向未来理想的人工智能化推进,未来,强人工智能有可能制造出真正能推理和解决问题的智能机器,它有自己的直觉和自我意识,通过数据的收集和自主分析便能得到相应的推理结果。我国刑事诉讼法第十二条规定,未经人民法院依法判决,对任何人都不得确定有罪。届时,这项预测技术与传统原则或将产生冲突。

同时,也许将出现这样的局面:当某个人被列为潜在的犯罪分子或潜在的犯罪被害人,其相关权利的保护、监控约束的尺度以及在现有司法程序上犯罪预测技术的介入先后,都将成为大数据技术在司法活动中发展的法律挑战。

“随着大数据技术的发展,未来案件审理的数据采集也会广泛运用大数据技术,这也许会导致控辩双方的地位更加悬殊。”裴炜说道。

裴炜指出,政府和司法机关相对于辩方,在数据收集的能力上是不一样的。政府和司法机关有权介入一系列相关数据库,而现有的程序法显然没有给予辩方相应的权利,所以辩方显然不具备政府与司法机关的数据收集能力。

早在2014年7月,苹果手机违法收集用户定位信息就被央视曝出。报道指出,苹果手机升级到7.1.1版本后就会有常去地点的功能,该功能详细记录了用户的行动轨迹、常去地点和滞留时长等信息,这些信息经过专家的简单分析,便能猜测出用户的生活习惯甚至职业。

中国信息通信研究院互联网法律研究中心主任李海英指出,在如今的大数据时代初期,由于政策与法规不健全,数据收集的权属也缺乏明确的法律定义,以致数据的收集呈现出无时无刻、无处不在的局面:电脑上的网页浏览、邮件收发、聊天记录和搜索关键词,以及手机的位置信息、通话信息、收发短信、APP 下载信息等诸多环节,都不知不觉地收集着用户的信息。

李海英认为,这些信息收集渠道没有标准化规定和法律的支持,只会进一步加大数据信息的滥用。

此前,阿里巴巴董事局主席马云在世界互联网大会开幕式上曾指出:“未来,数据是生产资料,计算是生产力。”

“既然数据信息被视为生产资料,那么一定具有资产性或财富性。”李海英说道。

“对大数据或者任何一种数据来说,我们把它放入机器的一刹那,就意味着终究有一天会把这个数据从机器当中取出来。”裴炜认为,数据的采集目的无疑是为了应用,如果在数据采集时没有进行方法和标准的规定,就会导致数据的滥用,届时便会威胁到与该数据相关的领域,如金融、交通、医疗、司法甚至国家安全。因此,数据收集需要方法和标准的规定,同时,数据应用也需要标准化和相应的法律规程。

国家保密科技测评中心副主任谭辉也表示:“总体上看,我们国家面向信息化大数据的政策法规还是缺位的。”他认为,由于政策与法规的缺失、数据收集权缺乏界定,致使一些商家借此误用现有法规,为了利益故意绕开、避开法律规范,由此出现了“系统和数据裸奔”的现象。

而大数据如今已经成为国家的战略资源,同时也是国家之间竞争的焦点,如果不对这些数据信息的采集进行规范,甚至使其成为开源数据源,那么这种情况势必会威胁到信息安全。

“例如,现下常见的地图应用,方便了人们的出行,但同时这些地图应用在很大程度上也暴露了我国的地理信息。”谭辉说。

谭辉建议,应当将数据从源头分类定密,以我国现有保密法规为标准,在数据被收集时,将各类信息采取分类、分级、分层的方式进行管理,使得泄密行为有据追查,有责任可追。

1.1.1 “大数据”引领智慧刑侦

1. 内蒙古赤峰市敖汉旗公安局案例

2013年12月16日,内蒙古赤峰市敖汉旗公安局图侦中队民警正在审看视频监控资料。随着视频监控设备日益广泛应用,今年敖汉旗公安局正式组建了图像侦查中队,运用云检索、大数据、智能图像识别等新技术,深入开展视频巡查、图侦作战。截至目前,图侦中队已破获各类刑事案件53起,抓获犯罪嫌疑人21人,打掉犯罪团伙4个,成为维护社会稳定、提升群众安全感的一支重要力量。

2. 平湖的成功案例

2015年4月份,平湖新埭辖区实现连续十五天刑事警情零,出现这个数据绝不是偶然。2015年以来,整个平湖境内刑事警情只有个位数日子越来越多,人民群众的安全感满意度也是持续保持高位,这在几年前是难以想象的。

出现如此良好的局面虽然原因是多方面的,但其中一点,运用“大数据”进行预警分析,

进而合理地安排警力压降刑事功不可没。

2015年以来,平湖公安以大数据为支撑,投入使用可视化智慧平台系统,每个派出所的综合指挥室每天、每周、每月都会根据辖区的警情的发生时间、地点自动生成图表进行分析研判,然后发布预警,合理安排警力对案件进行防控。每天晚上7点半的每日点评会商,各派出所又会将当天的有关情况向局里汇报。随后,再局层面,再根据当天各派出所的警情情况进行分析研判,然后布置警力。通过以大数据为支撑,构建了一个立体化的防控体系。

2015年8月初,在独山港镇区域内连续发生多起盗窃农宅的案件,平湖公安在抓紧破案的同时,迅速启动预警系统,晚上增加乡村巡逻的警力,并加强视频侦查的力度,一发现可疑车辆,立即指令附近巡逻力量进行拦截排查。

后根据数据分析,办案人员发现了嫌疑人的活动轨迹的规律,并增派警力在相应的区域内进行蹲守。2015年9月3日,在沉寂了半个多月后,嫌疑人带着工具准备再次作案,被守候的民警逮了个正着。此后,独山港区域内再无发生类似的案件。

平湖作为一个经济较为发达的县级市,外来人口约40%,复杂多变的社会治安形势,对依赖“人海战术”的传统警务模式提出了巨大的挑战。

平湖市在警务智能化建设再次“提档升级”,展现出了“大数据”实战运用的巨大威力,也对大力推进公安体系和能力现代化具有重大示范意义。

2013年,在“大数据”引领之下,平湖公安组建了合成作战中心,通过两年来的不断磨合和创新,合成作战产生了巨大的威力。

五年以前,凡是遇上通信(网络)诈骗案件,办案人员总是束手无策。然而,自合成作战中心成立之后,平湖公安运用“大数据”引领破案,实现了打击通信网络诈骗破案零的突破。经过两年的磨合,如今平湖刑侦在打击通讯(网络)诈骗案件方面是越打越顺手。

“打击通讯(网络)诈骗案件,对合成作战的能力要求非常高,是检验合成作战运用能力的磨刀石。通过这两年的发展壮大,如今我们合成作战中心的作战能力已经有了质的飞跃。只要在案件前期查到犯罪嫌疑人的相关线索,通过相应的合成作战,破案率可达90%以上。”

“我们现在对这类案件的打击已经越来越顺畅了,只要前期能查到犯罪嫌疑人的一些线索,基本上都能将犯罪嫌疑人带回。”刑侦大队副大队长徐挺烽自豪地说道。

2015年7月10日中午,犯罪嫌疑人假冒消防大队长的名义想让在平湖做生意的老李代购一批帐篷,在高昂回报的诱惑下,老李轻信了对方,被骗了20万元。

接到报案后,平湖刑侦迅速启动破案程序,对此案展开侦查,通过嫌疑人的电话号码、资金流向为线索,运用“大数据”进行分析研判,最终锁定了嫌疑人的落脚点——河南驻马店。随后,刑侦部门派出网警、刑侦、经侦等警力做进一步侦查,最终成功抓获了三名犯罪嫌疑人,并成功追回了老李被骗的20万元。

2014年1月,苏、浙、沪15市(区、县)第61次警务合作会议在平湖召开,平湖作为会议的主办方首次在会议上提出了“共享资源、共担风险”建立一个信息共享平台,并交换了大量的信息资源,这为区域协作破案带来了极大的方便。

平湖位于长三角经济圈的中心地带,区位优势越来越便利,在为百姓出行提供便捷的同时,也为犯罪分子流窜作案提供了方便。但由于行政区域的不同,跨市、跨省的案件在侦查办案时往往会遇到很多的障碍,尤其是在信息资源共享这一方面尤为突出。

在这一背景之下,平湖公安顺应“大数据”发展的趋势,提出了这一方法,得到了与会各

市(县、区)兄弟单位的大力支持,首次数据交换,其中的数据就有数十亿条。

有了大量的基础数据作为支撑,让数据引路,让数据说话,源源不断地为刑侦破案提供支持。

“几年以前我们破案,还基本要靠走访、询问,这样的传统方式费时又费力,但这几年随着视频监控网络的建成,DNA 实验室投入使用,各种新型的职能平台层出不穷,大大提高了刑侦破案的效率。”

近年来,发生在平湖的“两枪案件”破案率几乎达到了 100%,破案最快的一起案件,只用了 10 分钟就把犯罪嫌疑人拦截抓获了。如此高效的刑侦破案能力,“大数据”的运用功不可没。

刑事警情数连续三年下降的基础上实现再下降,打击破案数却连年上升,如此骄人的战绩背后的刑侦民警却没有一味地埋头苦干,通过积极顺应大数据时代的发展需要,充分发挥隐藏在数据中的强大能量,向科技要警力,向科技要成效,不断优化内部配置,平湖刑侦的战斗力的迈上了一个新台阶。

3. 西安“80 后”刑警用“大数据”破案

西安市公安局刑侦局,26 岁的刑警王悦,目不转睛地盯着眼前的电脑屏幕。

这是他每天最主要的工作,行话是“刑侦信息研判”,通俗点讲,是从浩如烟海的犯罪信息中,寻找不同案件的关系,串并案,获取破案线索。

近日,王悦作为西北五省唯一入围决赛的选手,参加了全国公安安全警种科技比武大赛,从 800 名精英中突围,获得了全国公安科技比武冠军,赢得了公安“科技之星”称号,这是公安部对西安公安科技人才储备和技术进步的高度赞誉。

“压力大,难度大,很紧张。”这个面容清秀,略带腼腆的民警告诉记者,决赛时,是两道主观性试题,题目的背景分别是一个全国跨区域团伙系列案件以及一个公安部正在督办的案件。

参赛选手要在一个小时内,通过电脑、人脑的高度结合,分析案件,找出真凶,同时逻辑和理由需要一一印证。

“两道考题的关键是,如何结合实际破案经验和大数据科技,在仅有的线索中抽丝剥茧,最终形成完整线索链条。”

他举了个例子,在西安市公安局全息作战平台上,集合了刑侦、治安、交管、禁毒、经侦等多警种优势,并整合了全国公安数据库以及陕西西安公安网络平台等 27 个平台的信息资源。

作为研判组成员,王悦每天要对西安市现发的 3 大类 18 种案件进行关注,利用海量数据进行分析、比对、串并,获取案件线索,最终将上述线索形成指令要求一线侦查单位开展打击破案工作。

在这样的一个用大数据信息指导实战破案的过程中,一整套规范信息流转、加工、研判、预警流程成为王悦的“武林秘籍”。短短的 3 年,王悦仅仅坐在办公室的电脑前,就破获了 121 起案件、抓获 97 名犯罪嫌疑人。

西安市公安局刑侦局三处副处长马志云将此概括为“一次建制性改革”,“以信息研判为主导的侦查顶层设计,已在近几年指导实战初见成效。”

4. 东莞清溪警方大数据提升刑侦工作效能

2014年底,公安部召开全国刑侦工作会议,要求全面实行“科学指挥、合成作战、科技支撑、情报导侦”的打击犯罪新机制,真正实现对刑事犯罪活动的精确打击、深度打击。

借着“深化公安改革”的东风,广东省公安厅部署深化全省刑侦工作改革,要求公安刑侦部门用3~5年时间,在全省初步构建起以打击犯罪新机制为中心,以合成作战的“大侦查”布局为依托的刑侦工作新格局,实现全省侦查破案综合实力明显增强、刑事打击效能明显提升。

2015年初开始,为进一步提升刑侦工作效能,增强打击违法犯罪工作力度,大力推动“3+2”专项打击整治行动的开展,东莞清溪警方提前谋划,主动实施刑侦工作改革,取得了良好成效,发案量大幅下降。“2016年以来,我们大力推动‘刑侦改革’,做大做强情报中队,以大数据为指引,实现了对犯罪的打击数量与打击质量并重。”清溪公安分局局长黄淦洪说。

东莞清溪镇,毗邻中国港澳地区,与深圳市、惠州市接壤,被誉为“珠三角的香格里拉”。3月是清溪禾雀花开放的季节,每年来此赏花的人络绎不绝,近些年已经突破数百万人次,对人流的疏导、对治安的控制,都给清溪警方带来巨大的压力。

2015年3月3日,清溪市民钟某在赏花时,其停放在清溪镇大王山森林公园停车场的一辆佳美小汽车被人用解码器盗走车内存放的11600元人民币。由于珠三角大量的游客进入清溪赏花,“若此案不破,该类案件极有可能再发生。”警方表示。

清溪公安分局副局长张景传介绍,2014年初,清溪公安分局指挥中心将视频应用岗完全从110接警员中脱离出来,挑选了视频工作经验丰富的人员成立专门的视频应用岗,同时,将视频监控系统与情报系统、指挥调度系统关联对接,实现在一个平台上研判指挥,一个屏幕上联动展示。

2015年以来,通过改革,清溪公安分局更进一步着力构建以信息科技为支撑的刑侦工作模式,推进信息导侦工作,进一步提升侦查破案水平。

“我们在原来的指挥中心视频应用岗、刑警中队情报组的基础上,抽调专业人员,组建刑侦大队情报中队,最大限度地发挥情报信息应用效益,使情报信息研判成果在指导侦查破案、控制发案、安全防范工作中发挥更大作用。”张景传说,“情报中队相当于我们刑侦大队的‘中枢大脑’。”

据了解,情报中队坚持每日召开刑事情报工作例会制度,研究和分析每周全镇刑事案件发案情况,从而科学、及时、全面地掌握犯罪动态和发案走势,分析串并辖区内发生的各类刑事案件,提出打防控措施并共同提出分析串并意见。

上述案件发生后,情报中队民警立即联合刑侦大队、派出所等开展了案件分析研判,发现该案与2015年2月2日发生在清溪森林公园黄茅田停车场的一起盗窃车内财物案在作案方式上极为相似。“我们一方面通过视频与治安防控相结合,加强了清溪森林公园和大王山森林公园的巡逻防控,一方面通过治安视频卡口与便衣伏击相结合,加强打击现行力度,终于在13天后成功破获该案。”清溪公安分局刑侦大队负责人吴杰伟说。

该案破获后,赏花期间,清溪再无发生过车内财物被盗案件。2016年1—10月,清溪公安分局通过串并案破案32起,同比去年上升183%。

如果说情报中队是清溪刑警的“大脑”,那么,刑事技术中队就是清溪刑警的“心脏”。黄淦洪表示,清溪公安充分认识到队伍保障对于推进刑侦改革、落实“一长四必”(即县、市、

区公安局长对刑事案件现场勘查工作负总责,现场勘验工作要做到“必勘、必采、必录、必比”)建设的重要作用,积极向清溪镇委、镇政府汇报公安开展刑侦改革、“一长四必”现场勘查新机制建设及创建一级技术室等工作情况。为推进相关工作,镇委镇政府召开联席会议,决定拨出200多万元用于支持公安分局刑侦场所改建和技术装备升级。

此外,清溪警方先期通过调研、统筹有关刑事人才,做好人才库建档,组建和完善专业刑事勘查队伍,同时,按照“一长四必”新机制的要求,由刑侦大队牵头,组织轮训派出所等的刑事勘查力量的现场勘查技能,培育出了一批侦查、破案的“多面手”,全面提升刑事勘查队伍整体的专业技能。

在提升刑事技术力量的同时,清溪警方通过健全案件协作机制,加强了各个中队间的沟通协作,深入梳理串并案信息和线索,不断加强串并案工作。“我们尤其加强了对系列性、流窜性侵财型案件的分析研判以及新类型案件的作案手法,分析研究发案规律、作案手段和犯罪特征,强化串并侦查。”吴杰伟说。

这样做的目的,是为了提高“从案到人”的多元化打击模式,提升侦查、打击效能。据介绍,清溪警方依托公安情报信息综合应用平台、刑侦网上作战平台、视频监控系统,联合刑事技术、刑事侦查等部门,对每天发生的案件进行分析研判,积极开展情报信息采集、信息研判、串并案件、情报导侦等“网上作战”各项工作,实现刑侦部门与情报系统的有效对接,为侦查破案提供有价值的线索。

如2015年以来,为摸清清溪镇内生产、销售有毒有害食品犯罪的情况,刑侦大队民警全方位、多渠道排查辖区内食品违法犯罪活动的线索。2015年10月20日上午,警方成功打掉位于清溪镇银湖市场的一家肠粉店,抓获涉嫌销售有毒有害食品的犯罪嫌疑人欧某(男,47岁,河源人),并现场查扣有毒有害添加剂硼砂一罐。

2015年以来,清溪警方通过DNA、指纹比对,对200余名犯罪嫌疑人逐一进行情报筛查核实、串并、比对、研判。截至10月26日,清溪公安分局通过指纹、DNA比中到案26人,直接破获刑事案件119起,同比去年上升295%。

通过视频侦查,已经成为警方破案的一个重要手段。2016年以来,清溪警方已经完成了近300个高清视频监控点及高清治安卡口的全部建设任务。此外,2016年以来,东莞警方开通了涉案视频库系统,它会将东莞35个公安分局指挥中心视频研判员每天查看视频时发现的作案视频,标记上时间、地点、作案人员等信息,全部下载永久保存到服务器中,实现视频信息标准采集入库、串并案侦查和深度研判应用,为破大案管小案服务。

但由于涉案视频库功能尚未完善,以及操作人员理解能力的不同,清溪警方在实际录入工作中遇到了大量的疑难问题。吴杰伟介绍,为保障涉案视频库正常运行,清溪警方通过开展培训指导,基层民警、视频侦查员的平台操作水平得到了不同程度的提高。

“这是一个非常有用的系统,在一宗案件发生后,警方可以调取案件发生前后的视频监控,对案件的情况进行分析,同时这些视频会被存储下来,直到案件破获,可以在警方抓获犯罪嫌疑人后作为证据使用。”吴杰伟介绍。

在建设高清视频监控的同时,清溪警方亦将辖区每个村、各个网吧、酒店、学校等重点地方的视频监控都纳入到警方的视频监控网中。同时,通过对辅警、治安员的培训,警方建立了一支“以分局监控中心为中心,各社区村监控室为子单位”的专职视频侦查队伍。

2015年6月12日零时许,清溪镇三中村某网吧发生一起抢劫案。被害人罗某与其朋友胡某在某网吧内上网时,发现两名陌生男子想偷走其手机。两名陌生男子见扒窃的行为

败露,遂将扒窃得来的手机交还给罗某后准备离开。胡某阻止两名男子离开时,被其中一名男子拿出随身携带的刀具刺伤。

案发后,清溪警方立即成立专案组,全力跟进侦破案件。专案组通过对网吧视频跟踪调查,发现该两名陌生男子涉嫌与发生在清溪镇内的多起网吧扒窃案有关。专案组通过案件串并研判,摸查出该两名陌生男子的真实身份分别为韦某(男,22岁)、潘某(男,22岁,均为广西人),并掌握到该两名男子的活动轨迹。2015年7月21日,专案组了解到韦某两人正在厚街镇活动,马上组织警力采取抓捕行动。经过专案组周密部署,警方在厚街镇一出租屋内一举将韦某两人抓捕归案。

综上所述,大数据将使犯罪侦查模式发生根本性变革。大数据驱动的侦查模式是时代的必然选择,这不仅在于复杂的犯罪态势及其数据化生态,更在于大数据技术使得这种选择成为现实。在大数据时代,侦查要确立在线开放的理念、数据主导侦查理念、相关性理念、线上破案与线下证明相结合的理念。大数据驱动的侦查是一体性侦查、全景侦查、预测侦查和算法侦查。其机制主要有犯罪监控机制、犯罪侦破机制、犯罪预测机制。

1.1.2 复杂的犯罪态势与大数据时代犯罪的数据化生态

当下,犯罪呈现出更加严峻和复杂的态势。首先是犯罪总量大,犯罪率逐年上升。据统计,仅2012年公安机关刑事案件立案的案件数为6551440起,检察机关批捕、决定逮捕犯罪嫌疑人的案件数为680539起,人数为986056人。近二十多年来,中国犯罪率呈逐年上升趋势,刑事案件立案数平均每年增长22%以上,超过了全国GDP的增长。其次是犯罪智能化。犯罪是一种社会存在,科学的发展渗透到犯罪的方方面面,提升了其能力和危害程度。这表现为两个方面:一是运用科学思维实施的犯罪,主要表现为犯罪思维严密,犯罪前经过周密部署和策划,犯罪过程渗透着科学思维和谋略。二是利用科学技术实施的犯罪,突出表现为数字化犯罪。以利用网络犯罪为例,2012年全国公安机关累计破获涉网违法犯罪案件11.8万余起,抓获犯罪嫌疑人21.6万余人。据赛门铁克公司2012年9月发布的诺顿安全报告估算,2011年7月至2012年7月,中国有超过2.57亿人成为网络违法犯罪的受害者。网络违法犯罪所造成的直接经济损失达2890亿元人民币,受害者人均蒙受的直接经济损失约1200元人民币。第三,犯罪时空的复杂性。现代科技的发展,使得犯罪时间非线性,犯罪空间缺席性,时空组合的多维、多样化和任意性。第四,案件因果联系复杂。相对于传统的静态、单一社会来说,现代社会是一个动态、复杂社会。在动态、复杂社会中,因果联系具有非线性、偶合性、多因性、断裂性,犯罪的因果联系往往难以确定。

随着计算机及网络技术的发展,使得当下社会已经进入了大数据时代。大数据时代首先是数据记录时代。在数据记录时代,数据记录成为默认模式,人类社会处在被无所不在各种各样传感器和微处理器构成“万维触角”的数据网络记录之下,手机、网络、监控探头、射频技术等等无所不在地记录着我们的行为乃至我们的思想。早上出门,电梯的摄像头记录着我们的出行时间;开车上班,道路的摄像头记录着我们的位置和车速;工作期间,网页记录着我们的浏览习惯和搜索记录,电话记录着我们的联网对象和通话时长;下班回家,购物记录界定着我们的职业身份、家庭背景甚至性格特征,电视机顶盒记录着我们的收视习惯和价值品位,“在数字世界里,我们都会留下电子‘脚印’或电子‘指纹’。”“我们正处于一种不断变化却日趋紧密的被监视状态中。事实上,现在我们的一举一动都能在某个数据库中找到线索。”

狡猾的犯罪者能有例外而成为“数据隐士”吗?要成为“数据隐士”,意味着你要完全脱离现代社会系统,不仅不能使用数字化产品,还要完全意义上不食“人间烟火”。因为现代社会几乎被数据化了,一旦你与现代社会系统进行交换,就很有可能被数据捕捉和记录。然而,这并不是说犯罪者的具体犯罪的任何要素或片段如犯罪时间、犯罪空间、犯罪行为、犯罪工具等等都会直接且完整无缺被数据记录和储存;而是说犯罪者隐藏的犯罪信息总是被相关的海量数据从不同的侧面记录着,即便是某些甚至是主要或关键的犯罪要素或片段缺失,也可以通过不同侧面相关海量数据联接、分析,拼接或描画出犯罪过程。因此,在大数据时代,不要说数字化犯罪,即使传统手段的犯罪,都可以说落入了一种“天网恢恢,疏而不漏”的网络记录和存储体系,数据化就是当下犯罪的现实生态。

1.1.3 大数据驱动的侦查模式是时代的必然选择

模式指经过提炼和抽象的标准样式。侦查模式反映了侦查要素的结构关系和运行逻辑。侦查模式可以按照不同的标准进行分类。学界按照侦查是否运用信息科技手段,把侦查模式分为传统的侦查模式、信息主导侦查模式。然而,如果从信息论的视角来看,传统侦查模式与信息主导侦查模式的本质区别不是是否运用信息,而是信息记录、存储、提取以及分析方式上的根本差别。按照侦查所能运用信息的记录、存储、提取以及分析方式,可以把侦查模式划分为传统侦查模式、业务信息主导的侦查模式和大数据驱动的侦查模式。学界一般将业务信息主导的侦查模式和大数据驱动的侦查模式合称为信息主导侦查模式,但两者之间不仅是发展阶段上的差异(大数据驱动的侦查模式是在业务信息主导侦查的基础上发展起来的),而且在信息类型、信息提取和研判方式上也有根本差异,最重要的是由此差异而带来侦查理念、特征和机制上的根本变革。

传统侦查模式是在信息存储、提取和分析上几乎没有什么科技含量的模式。传统社会,人类对信息的记录和存储方式主要是人的大脑和书写体系(传统社会由于信息记录的需要发展出一整套书写体系,由此而产生了许许多多按时间汇集的分门别类的书写档案库)。对于犯罪的信息记录来说,除了大脑和书写档案外,犯罪现场也以物质交换的形式记录着犯罪信息。因此,传统的侦查主要手段是调查访问(对大脑储存的信息提取)、书写档案的查询。人脑信息的存储和提取的特点是:分散在不同的人身上;信息的准确性差,受到外在环境和信息储存者自身感受能力、记忆能力等影响;信息缺乏稳定性,信息量和准确性随着时间变化而衰减;信息能否提取以及提取的质量,首先取决于能否找到储存信息的人,其次取决于侦查人员的询问技术(经验)、被询问人表达能力、情绪、配合态度等等多种因素。书写档案记录信息的优点是准确性高、稳定性强,但其有两个重大缺陷:一是提取困难。人们要找到其中一点有用信息,就得把所有的资料翻阅一遍;尽管后来建立了图书馆式的目录索引,但查找起来依然耗时费力。二是不能提供直接的犯罪信息。书写档案不可能是犯罪的实时记录,只可能是犯罪破获后一种事后登记,因此这种档案对于需要破获的犯罪来说,不能提供直接的犯罪信息。传统侦查的信息分析研判主要依靠侦查人员的经验,有经验的侦查人员往往成为是否破案的关键。总之,这种模式科技含量低、粗放型特征突出,能否破案主要取决于侦查人员的经验和投入的人力多少,不仅如此,还取决于侦查人员的运气。这对于传统静态、单一的社会及其犯罪也许能够适应,而与动态、复杂的社会及其犯罪几乎完全不匹配。

业务信息主导侦查模式是在信息技术引领下的以业务信息存储、提取和研判为基础的侦查模式。随着信息技术的发展,各种各样信息记录和存储设备被广泛使用。信息记录和存储不再完全依赖人脑和书写档案,而是电子化的记录,存储设备成为人类记录和存储信息的主要方式。这些设备代替人脑和书写档案实时记录着人类的行为,也记录了犯罪行为。所记录和存储的信息从来源和存储分布来看,形成于不同的业务经营并分布储存在不同的业务信息库中,如商家记录和存储人们的消费信息、银行记录和存储了人们的金融交易信息、医院记录病人信息等等。这些信息库缺乏整合,相互之间形成信息孤岛,信息冗余和信息孤岛成为信息存在的基本生态。就业务信息主导侦查模式来说,其主要特征是:一是侦查部门依赖于公安平台所累积的结构化的数据库主要用于人、事、物的核查、比对,实时犯罪信息仍然主要依靠人工采集。二是信息提取依然困难。不可否认,相比传统侦查模式,业务信息主导的侦查模式针对公安机关所累积结构化信息来说,确实大大提高了查询、比对效率,但是面对越来越多的被累积的不同来源、不同结构的数据,尤其是大量的半结构化和非结构化数据,既缺乏数据整合的技术和机制,也缺乏信息提取的技术手段。结构化数据是先有模型后有数据,大多具有事后登记的性质(也有少量的实时记录的数据如旅馆住宿等),很难有实时犯罪行为记录信息,其主要价值在于对人、事、物的核查;而正是不同来源的半结构化、非结构化数据中实时记录了犯罪的“蛛丝马迹”。三是信息分析、研判仍然主要依靠侦查人员的经验。业务信息系统主要用于简单的查询、比对,但是不能进行智能化的算法分析。总的来说,这种侦查模式面对当下的犯罪态势,尤其是流动性犯罪、数字化犯罪等,难有成效。

大数据驱动侦查模式是建立在大数据和云计算平台的基础上,是大数据时代的信息主导侦查模式的升级换代。在大数据时代,大数据驱动的侦查模式是一种时代的必然选择,这不仅在于复杂的犯罪态势及其数据化生态,更在于大数据技术使得这种选择成为现实。

首先,犯罪的数字化生态是大数据驱动侦查模式的现实基础。面对当下复杂的犯罪态势,人们似乎有点不知所措。犯罪的控制某种程度上是一种侦查技术对犯罪技术保持优势。然而现代性的发展使犯罪者具有更强的匿名性、流动性等,从而一度打破了公安机关曾经具有的优势,这也是如今犯罪爆发性增长的原因之一。然而犯罪作为一种社会存在,当社会成就犯罪条件时,也会给人类提供制约其的机会。犯罪的数字化生态根本改变了犯罪信息的记录和存储方式,极大地扩大了“社会记忆”,大数据技术将彻底改变侦查技术与犯罪技术之间的对比关系。因此,我们必须改换传统的侦查模式,采用大数据驱动侦查模式以控制犯罪和打击犯罪。

其次,在大数据时代,侦查所面对和所能处理的数据不再是小数据,而是大数据。如今,侦查所面对和所能处理的数据具有体量大、类型多、价值密度低的特征。“池塘”和“大海”最容易发现的区别就是规模。过去侦查,即使是业务信息主导侦查阶段,所面对或所能处理的数据量相当于“池塘”,而与此相对照,现代侦查所面对和能处理的数据量则是“大海”。不仅如此,现代侦查所面对的则是数据的多样性:从结构上看,不仅有结构化数据,还有大量半结构化和非结构化数据;从数据类型看,有业务数据、用户原创数据、传感器感知数据;从数据表现形式看,有文字、图片、音频、视频、链接等;从犯罪案件构成角度看,有人及其关系、行为、物、时间、空间和主观意图数据。数据的价值密度低。在巨量的数据中,有关犯罪数据混杂其间,仅仅是其中小小的“浪花”,但其却弥足珍贵。以视频为例,连续不间断监控过程中,可能有用的数据仅仅有一两秒。

第三,大数据技术能从海量的数据中对犯罪信息进行提取、分析研判以及预测未来。大数据是其规模或复杂程度超出了常用技术按照合理的成本和时限捕捉、处理的数据。而以云计算为依托的大数据技术可以突破常规技术成本和时限的要求。具体来说,其一,大数据技术能适时提取和分析处理多结构多源数据,尤其是半结构和非结构化的数据,能够从海量的、杂乱无章的数据中抽取大量的与犯罪相关的细节、点滴片断、不同侧面数据、信息,并且能把“数据联系起来、信息点连接起来、片断串联起来”,从而能将表面看来毫无意义、互不关联的数据碎片拼出一幅清晰完整的犯罪图画。而对于确定一个犯罪嫌疑人的身份来说,也许只需要四个信息点就足够了。其二,大数据以云计算为依托,能够在合理时间内进行信息提取和分析。以周克华案件为例,南京警方动用上百名警力花费了数天时间对视频监控数据进行人肉搜索,而运用大数据技术也许只要几个小时就足够了。其三,大数据技术,一个最为根本的突破是能够运用海量数据进行算法分析,进行信息研判,从而帮助我们认识过去,分析原因,揭示犯罪发生的规律。最后,大数据能在分析过去中寻找有意义的模式,从而预测未来,为我们优化警力资源配置、打击犯罪提供先机。

1.1.4 大数据驱动的侦查模式的理念变革

黑格尔指出,“理念是任何一门学问的理性”,并认为理念中包含着“某种预想的东西”,具有前瞻性、导向性和设计性。侦查模式转换首先是理念转换。侦查模式中的理念就是指贯穿在侦查模式中反映了侦查规律的并具有引导、支配、决定侦查活动的观点、看法、信念。大数据驱动侦查模式不仅是一种新的工作模式,更是一种新思维、新理念。在大数据时代,侦查要确立的理念有:

1. 在线、开放的理念

大数据首先是在线数据。大数据不仅是体量大,更是实时记录社会的复杂动态数据:用户原创和各种传感器感知数据,而正是这些数据混杂了犯罪的“蛛丝马迹”。对于侦查来说,公安大平台累积的结构化数据是重要的,尤其是对人、事、物的核查具有重要价值,但是很难有实时的犯罪记录。大数据驱动的侦查就是在公安大平台累积的结构化数据的基础上,对不断变动用户原创和各种传感器感知数据进行提取、分析和处理,获取信息。因此,对于大数据驱动的侦查,我们必须坚持在线和开放的数据理念,以获得我们需要的海量数据,进而分析、处理这些数据。

2. 数据主导侦查理念

大数据时代,数据是犯罪的生态,侦查过程就是数据储存、提取和分析过程,数据贯穿于侦查的各个环节,“让数据说话”成为侦查的基本思维。数据主导侦查的理念至少包括以下三个方面的内容:首先,有关犯罪的一切现象皆可数据化。凡事皆可量化,皆可数据化。不仅与犯罪相关的有形之物如时间、空间、人的特征(生物识别特征、行为习惯等)、行为、手段、物等可以量化和数据化,那些与犯罪相关的无形之物如人的价值观念、态度、情绪等等也可以量化和数据化。其次,大数据是侦查的基础资源,是侦查的工具箱。侦查就是对数据开矿式的挖掘和分析,侦查能否成功某种程度上取决于对大数据资源的提取、分析能力;运用大数据各种分析技术,可以获得我们所需要的犯罪信息。最后,在大数据时代,数据居于侦查过程的核心地位,支配着侦查的运行。犯罪现场重建、侦查决策、侦查途径的选择、侦查分析、数据摸排、侦查预测等等无不围绕数据运行。

3. 相关性理念

大数据是通过量化两个数据值之间的数理关系来确定相关关系。相关关系强,是指当一个数据值增加时,另一个数据值很有可能随之增加。传统侦查,是按照因果关系和数据结构的标准来采集数据和分析数据。到大数据时代,我们能分析、运用几乎所有相关数据,收集数据不必再拘泥于因果关系和数据结构标准,而是坚持相关性标准,不仅采集结构化数据,还要采集半结构化和非结构化数据。这种相关关系虽然不能直接揭示内在的因果关系,但是对于犯罪侦查和控制来说,其展现的相关关系仍具有较强的效用价值。

相关关系能让侦查人员全方位、多角度地思考分析案情。相关关系虽然不追求精确性,但是其追求丰富性,不拒绝任何机会,尽可能去创造和利用机会。通过相关关系,才能将看起来没有联系的信息内在地联系起来,从而更为全面地认识案件情况。这也许可以帮助我们发现破案线索,理清破案思路,划定侦破范围。

相关关系可以给我们进一步确定因果关系以指引,从而确定犯罪原因和证明犯罪。相关关系的分析是分析因果关系的基础。相关关系并不必然是因果关系,但因果关系必然是高度相关关系。通过相关关系,我们可以进一步探究其中是否存在因果关系,从而证明犯罪。

相关关系的一个重要价值是可以监控犯罪情势。如上所述,当下影响犯罪的原因是纷繁复杂的,要确定犯罪发生的原因相当不容易甚至不可能。对于侦查人员来说,重要的也许不是去理清犯罪原因,而是控制犯罪。通过相关关系,确定关联物,进而可以监控犯罪情势,从而使我们有效配置警力资源,打击犯罪。

通过相关关系,可以预测犯罪。大数据的核心价值是预测。通过收集具有相关关系的数据,建立大数据模型,我们可以从微观上预测什么时间、什么地点、什么人、什么类型等等的犯罪容易发生,也可以从宏观上预测犯罪趋势,这为我们防范和打击犯罪提供了更好的机会。

线上破案与线下证明相结合的理念。大数据使得发现和确定某一犯罪嫌疑人似乎变得相当容易。但是数据只是事实的镜像,并不等于就是事实;而且大数据的算法逻辑(强调相关关系、确定的只是一种概率,甚至由于噪音等因素会出现致命的误差)与法律证明逻辑(强调因果关系和排除合理怀疑标准)存在差异,因此,犯罪侦查尚需要进一步按照法律体系的操作要求进行证明。即使我们通过大数据可以确定犯罪嫌疑人,达到了排除合理怀疑的标准,我们也必须把大数据的算法体系转化为符合法律规范要求的证明体系,把数据确定转换为法律确定。然而,线上破案和线下证明并不是割裂的,大数据能对我们证明起引导作用,帮助我们寻找证据,确定因果关系。因此,在大数据时代我们既不能抛弃相关关系,只追求因果关系,也要必须防止用相关关系代替因果关系,防止用预测来代替事实。

1.1.5 研究意义

1. 为刑事案件信息化侦查提供理论指导

以往刑事案件的侦查大多借鉴一般经验型思路开展工作,缺乏模式化方法。本书根据刑事案件的特点,借鉴一般经验型案件的侦查思想,结合公安机关办理刑事案件流程及其他相关理论,提出了适合于刑事案件侦查的数据仓库模型和建模方法,为刑事案件的侦查提供了理论基础和工作指导。

2. 发挥现有刑侦信息资源的整体合力

根据大情报体系的总体方案和要求,研究开发具有各种情报分析功能的计算机信息系统——情报综合分析平台。系统主要架构在线索型情报系统、案件信息系统、其他刑侦业务信息系统和社会系统之间,通过搭建一个平台,利用上述系统中的相关数据,实现对人、组织、事件、物品等相关静态资料和动态信息的整合,为各级刑侦机关尤其是各级情报部门的关联查询、批量比对和综合分析服务。利用数据仓库技术关联多个刑侦业务系统和社会采集信息系统的相关数据,为查询用户提供网状结构的关联查询,即通过一个数据项目或多个数据项目进行层层关联查询,提高查询工作效率。

3. 实现刑侦工作的信息化

以协同办案综合系统、指纹核查比对系统、协同侦查分析系统和情报线索系统为主,构建了刑事侦查的应用骨干系统,为刑事侦查工作提供了有力的信息支持,创建了刑事侦查的新模式。其中协同办案综合系统,通过对工作流的管理,完整、鲜活地收录了从接警到结案的刑事、治安案件信息,可以为侦查工作提供丰富的信息分析资源。侦查分析系统,实现了传统的刑事串并分析模型的软件化操作,为侦查人员提供了分析平台。情报信息搜集系统,使得刑侦人员可以根据自己对案件的理解,对海量信息进行分析研究。刑侦信息化的初步实现,改变了传统运作模式中常常出现的信息不对称现象,使得不同岗位、不同单位、不同层面的侦查员占有了相同的信息资源,为侦查人员自身的能动性提供了极大的空间,提高了破案的能力和效率。

4. 完善刑侦工作模式精准打击

改变传统的“由案到人”的侦查模式,完善为从案到案,从物到案,从人到案的侦查模式。变各自为战为多警种、多手段互助互补。通过整合,能够在不增加情报力量的前提下,提高刑侦情报的效能。将已经建设应用的指纹、足迹、DNA等痕迹数据库和系统进行整合,建立起集网上排查、网上串并、网上控嫌、网上控赃、网上缉捕于一体的“网上作战”立体系统。可以自动或半自动网上跟踪违法犯罪嫌疑人的活动轨迹,发现违法犯罪嫌疑人的作案规律。网上作战的基本原理,就是以计算机网络为载体,以各类信息资源为实体,依托一系列信息整合分析工具软件,通过对各类有形痕迹和无形痕迹的信息数据分析,自动或半自动地在较大的空间范围与时间跨度内发现犯罪嫌疑人的踪迹,揭露和刻画犯罪嫌疑人的轨迹,特别是作案轨迹和活动动向,从而发现破案线索和锁定、快速准确缉捕犯罪嫌疑人。

1.2 研究现状

“互联网+”时代下的信息每一秒都在发生变化,在办理职务犯罪案件过程中,要想在审讯中尽快突破犯罪嫌疑人,必须事先对他进行多方面的了解,更多地与“互联网信息”打交道将成为审讯工作的新着眼点。“互联网+”“大数据”驱动的侦查模式已取得了初步的研究成果,据研究表明,初步可总结为:

审讯前期——既依靠“脚板”,更注意“指尖”。初查时不仅要盯住人,更要充分发挥“互联网+”的作用,采取话单分析、网络流量嗅探、已删除文件恢复等技术手段,搭建“立体式”的初查路径。我们在办理市建设工程质量监督站朱某受贿案中尝试采用了“电话+短信”双料话单分析手段,通过对通话次数、时间、短信频率的分析,锁定了10名有可能向朱