

警务大数据应用途径研究

——基于文献计量分析

■ 刘 铭

摘 要 在我国大数据应用蓬勃发展的背景下，聚焦警务大数据应用现状，以中国知网（CNKI）数据库收录的大数据相关文献为研究对象，对比分析现阶段警务大数据应用成果，提出面临的挑战 and 对策，为深入发展警务大数据应用提供参考。

关键词 文献计量分析 警务大数据应用 信息化 智慧化

大数据应用在提升公安工作的智能化水平、增强防控能力、提高决策效率、促进信息共享和协作、提高服务群众水平等方面为公安工作的发展提供有力支持。2023年2月中共中央、国务院印发《数字中国建设整体布局规划》，数字中国建设有了较为完整的顶层设计和整体谋划。在这样的时代背景下，深化警务大数据应用，做好公安信息化的“算”“数”题，是推进公安工作数字化转型现代化改革的关键环节。

一、警务大数据应用发展的现状

（一）数据基础深厚

公安机关信息化建设起步早，针对具体业务，聚焦实战孜孜不倦、专研创新，经过一系列里程碑式的建设与实战应用，从在全国率先开展的公安 Intranet 专网建设开始，到目前以大数据应用为核心的发展进程中，网络结构配置合理，数据传输可靠，业务模型持续优化、系统架构稳定，应用涉及民生领域广泛，数据资源积累丰富。同时，随着不同来源、不同类型的多模态数据，如文本、图像、声音、视频等多种形式数据，与传统结构化数据的交叉融合，共同构成了深厚的数据基础。

（二）应用效果显著

警务大数据应用对犯罪预测和打击的效

作者：北京警察学院教授

基金项目：本文系北京市教育委员会科技计划项目“警务大数据应用技术研究”（项目编号：KM201914019002）阶段性成果。

果有显著提高。通过对历史犯罪数据等相关信息深入分析,能够更精准地定位高风险地区,采取有针对性的预防和打击措施,从而有效降低犯罪率。警务大数据应用加速案件调查和解决的过程。通过自动化的数据分析和关联,更快速地识别犯罪嫌疑人、查明案发经过,提高破案效率,确保及时的司法程序。警务大数据应用更有效地分配资源。通过分析犯罪热点、警力需求等数据,可以优化警务部署,确保在需要的地方和时候有足够的警力介入,提高应对突发事件的能力。警务大数据应用改善社会的整体安全感。公众能够感受到公安部门更为智能、高效的工作,这不仅减少了犯罪对社区的负面影响,也提升了市民对公安部门的信任感。警务大数据应用为决策者提供更科学的信息支持。基于数据的决策可以减少主观判断的偏见,确保决策更加客观、全面,从而更好地应对复杂的治安和犯罪挑战。

(三) 面临两个矛盾

数据规模和数据处理方式相对不足的矛盾。数据为警务工作提供了丰富的信息资源,但同时也带来了数据分析和处理的挑战。传统的数据分析方法难以应对如此庞大的数据量,导致数据分析和利用的效率不高。深化警务大数据应用的过程中,探索更加高效和智能的数据分析处理方式,突出应用创新,对充分挖掘数据中有价值的信息的需求日益增加。

数据规模发展和提升数据质量的矛盾。目前警务大数据应用系统大多数是围绕业务,通过数据驱动做出来的应用系统,这样的系统保证结果正确性的前提是数据的正确性,对数据质量要求非常高,为减少性能受人为因素影响,提高系统鲁棒性,避免发生重大的偏差,需要进行大量的训练数据样本,

对数据规模要求出现几何级数的增长。然而,数据规模的增加并不意味着数据分析结果一定正确、数据蕴含价值一定增加,相反随着数据规模变大,数据噪音也会随之增加,其中的“脏”数据会越多,对分析结果会产生偏差。甚至如果数据本身在说“假话”,那么系统分析的结果一定是不正确的。

(四) 机关与基层发展不均衡

机关数据集中、处理能力强和基层数据相对分散、处理能力弱是伴生警务大数据应用发展的现象。由于数据规范和分析处理在机关、数据采集在基层,机关积累了大量数据,分析处理手段丰富,这些数据也较为完整、规范,同时数据集中导致数据处理压力大。而基层数据相对分散则是指在基层工作中,面对不同数据采集要求,数据标准规范不统一,数据处理手段单一,存在重复性,数据冗余大,但难以直接应用于基层警务工作,导致对基层警务工作的信息支持不足。

二、深化警务大数据研究的挑战

(一) 代表性成果形成相对较晚

我国大数据时代的开启在 2010 年,公认的中国大数据元年是 2013 年。^[1]为反映国内高质量、有代表性的大数据领域研究成果,以主题为“大数据”,出版年度为“2013 到现在”,期刊来源为“SCI 来源期刊、EI 来源期刊、北大核心、CSSCI、CSCD”等作为条件进行检索,在中国知网共检索到 42078 篇文献,发文量总体趋势如图 1 所示。

这表明,我国在过去的 10 年间,大数据领域得到普遍关注,逐年持续取得丰硕的研究成果,高质量的大数据研究学术成果趋于稳定,大数据领域的研究规模已经形成。^[2]

为反映出警务大数据的研究成果,在

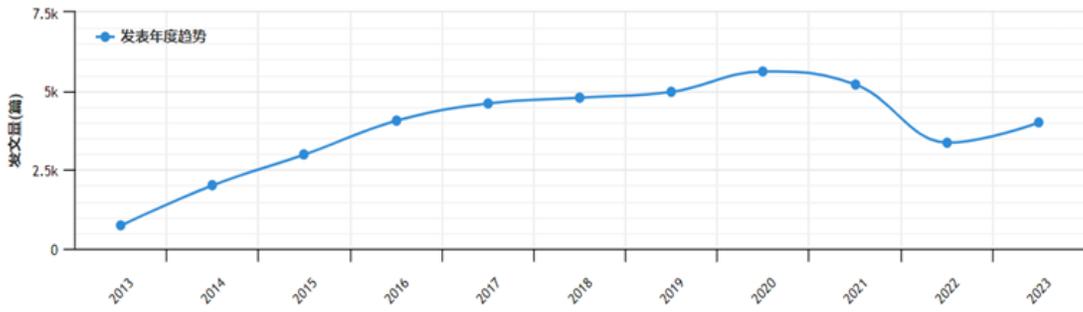


图 1 发文量总体趋势

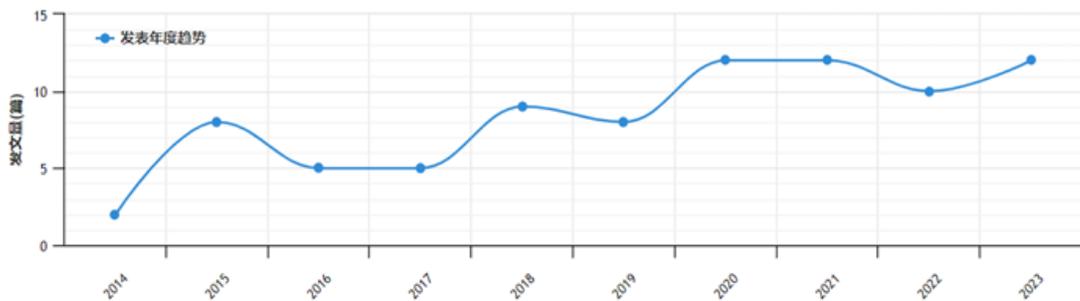


图 2 警务大数据研究发文量总体趋势

表 1 被引用次数最高的 100 篇期刊文章指标

| 指标 | 数量 |
|-------|----------|
| 总参考数 | 2414 |
| 总被引数 | 65379 |
| 总下载数 | 2313141 |
| 篇均参考数 | 24.14 |
| 篇均被引数 | 653.79 |
| 篇均下载数 | 23121.41 |
| 下载被引比 | 0.03 |

表 2 大数据警务相关文献指标统计

| 指标 | 数量 |
|-------|-------|
| 总参考数 | 948 |
| 总被引数 | 970 |
| 总下载数 | 62601 |
| 篇均参考数 | 13.54 |
| 篇均被引数 | 13.86 |
| 篇均下载数 | 894.3 |
| 下载被引比 | 0.02 |

中国知网，以主题为“大数据”和“警务”，以期刊来源为“SCI 来源期刊、EI 来源期刊、北大核心、CSSCI、CSCD”等作为条件进行检索，检索到文献总数为 80 篇，发文量总体趋势如图 2 所示。

研究发现，警务大数据研究起步相对较晚，在一段时间发展趋势未能与全国大数据研究持续发展同步。在 2016 年，公安部提出了“以大数据应用为核心”的发展方向后，在大数据总体战略布局的推动下，呈显出规模化发展的态势。^[3]

(二) 大数据领域影响力不足

尽管论文被引用次数并不完全能够体现学术文章重要性，但确实可以作为评价一个研究成果是否具有代表性、得到学术界认可的一个重要指标。通过选取全国范围内大数据研究被引用次数排名最高的 100 篇期刊文章分析，得出各项分析指标如表 1 所示。

与警务大数据相关的 80 篇文献指标如表 2 所示。单从文献来评价影响力，基本反映在参考数、被引述、下载数等指标上。不难看出，警务大数据研究成果与国内大数据代表性研究成果的指标差距，产生这些差距的客观原因主要是公安应用领域、覆盖范围

和行业特点所决定的。通过总参考数、篇均参考数等指标对比,警务大数据研究成果明显低于国内大数据代表性研究成果,在一定程度上反映出警务大数据研究成果形成过程中对已有成果借鉴还不够充分,研究已有大数据理论成果对警务大数据研究支撑作用广度不够;通过篇均被引数和篇均下载数指标对比,反映出警务大数据研究成果对整个大数据研究领域影响力较小,还没有产生规模效应。

(三)应用研究的着力点向新技术领域迁移不足

文献共引情况统计如表 3 所示。通过共引文献总体分析,可以反映出文献间的相关性程度,反映出研究成果共同关注的主题。观察共引文献情况不难发现,在公安信息化迅速发展环境下,业务工作对数据需求发生本质变化背景下,开展了警务大数据研究,但是对全国大数据领域中代表性理论成果借鉴较少,研究的理论依据、研究方向存在一

表 3 共引文献情况表

| |
|--|
| 吕雪梅.美国预测警务中基于大数据的犯罪情报分析.情报杂志.2015 |
| G. O. Mohler;;M. B. Short;;Sean Malinowski;;Mark Johnson;;G. E. Tita;;Andrea L. Bertozzi;;P. J. Brantingham.Randomized Controlled Field Trials of Predictive Policing.Journal of the American Statistical Association.2015 |
| (英)莱特克里菲(Ratcliffe,J.),著.情报主导警务.中国人民公安大学出版社.2010. |
| 张兆端.关于公安大数据建设的战略思考.中国人民公安大学学报(社会科学版).2014. |
| 查云飞.人工智能时代全自动具体行政行为研究.比较法研究.2018 |
| 程雷.大数据侦查的法律控制.中国社会科学.2018 |
| 杨婷.论大数据时代我国刑事侦查模式的转型.法商研究.2018 |
| 高文英.警务数据采集与运用中的法律问题探究.中国人民公安大学学报(社会科学版).2019 |
| 张兆端.智慧公安.中国人民公安大学出版社.2015 |
| 迈耶-舍恩伯格.大数据时代.浙江人民出版社.2012 |
| 赵大伟,主编.互联网思维独孤九剑.机械工业出版社.2014 |
| 彭知辉.大数据:让情报主导警务成为现实.情报杂志.2015 |
| 陈凤;张治军.警务信息综合应用管理平台建设的研究.吉林公安高等专科学校学报.2009 |
| 曾明星;周清平;蔡国民;陈生萍;黄云;董坚峰.基于校园私有云的软件工程实验实训综合平台研究与实践.实验技术与管理.2015 |
| 陶国林;曾晓松;肖夏.实验教学信息化平台建设的关键问题探讨.实验技术与管理.2015 |
| 彭知辉.公安信息化建设对当前公安情报工作的影响.中国人民公安大学学报(社会科学版).2012 |
| 卢建军.浅谈公安信息化与情报主导警务的协同发展.公安研究.2014 |
| 蔡海南;张浩凯;张林.情报如何主导警务——情报导警的内涵及发展层次.湖北警官学院学报.2011 |
| 马德辉;苏英杰.“Intelligence Studies”视域下的中国公安情报学若干基本问题研究.情报理论与实践.2013 |
| 桑托什.犯罪分析与犯罪制图.人民出版社.2014 |
| 彭知辉.论公安情报学研究范式及其整合.情报学报.2013 |
| 张远煌,主编.犯罪学.中国人民大学出版社.2007 |
| (德)乌尔里希·贝克(UlrichBeck)著;何博闻译.风险社会.译林出版社.2004 |
| 彭知辉.大数据:开启公安情报工作新时代.公安研究.2014 |

定的局限性，主要集中在公安情报领域，并没有把大数据作为一个整体的、系统的、复杂的、技术与管理相融合的生态系统进行综合深入研究，缺少大数据系统的组成要素、本质特征、关键技术、系统架构、管理和安全等方面的研究，严格上讲并没有从多个维度把握到警务大数据应用研究的本质。

三、深化警务大数据应用有效途径

（一）充分借鉴已有成果深入关键技术应用研究

警务大数据研究的重点面向实战，关注应用实践研究，是在国内大数据研究学术成果相对成熟后发展起来的，前期研究正好为警务大数据奠定了研究基础，使其在理论体系、架构模型、关键技术及可行性等方面得到保证。警务大数据应用研究应进一步深入借鉴大数据领域代表性理论成果，结合公安业务工作实际，应用研究重点放在探索基于已有理论成果基础之上实现警务大数据应用最佳路径，探索在海量数据规模及数据流动速度超过了传统认知能力的挑战下，提升大数据驾驭能力，加大数据采集、数据清洗、数据存储、数据可视化、大数据系统架构等方面关键性技术的可行性研究。重新审视一个大数据解决方案需要面对三个关键问题，即大数据的存储、大数据的分析和大数据的管理。借鉴已有国内大数据研究成果，融入大数据研究大环境，结合警务业务工作流程实际情况，做好顶层设计，进行关键技术的适用性研究，如采用 Hadoop 生态系统开展警务大数据应用建设是如何解决的上述三个关键问题^{[4][5][6]}，用 HDFS、HBase 和 OpenTSDB 解决大数据警务数据存储问题是否可行^[7]，用 Hive 和 Spark 解决数据分析问

题是否可靠^[8]，用 Sqoop 和 Ganglia 等解决大数据管理问题是否有效。

（二）正确处理数据规模和数据质量的关系

警务大数据应用研究要处理好发展数据规模和保证数据质量的关系。数据规模发展为大数据应用提供基础，保证数据质量是大数据应用发挥作用的关键。警务大数据中，完全绝对纯洁的数据是不存在的，“脏”数据的存在是伴随警务大数据应用的固有属性，因此在不仅要提高数据的使用价值，更要使数据严密的关联，这就要求一方面通过数据清洗技术满足数据质量的提高^[9]，另一方面容忍非清洁数据和实现知识服务效率与水平的同步提高^[10]。上述问题触发警务大数据应用研究对 CAP 原则的重新思考。CAP 原则是分布式系统设计中的一个重要理论，它阐述了任何分布式系统都无法同时满足一致性（Consistency）、可用性（Availability）和分区容忍性（Partition Tolerance）这三个属性，而只能在其中选择两者。^[11]在警务大数据应用中，平衡一致性、可用性和分区容忍性，从而在保证数据规模增长的同时，保证数据质量。

一致性（Consistency）与数据质量：一致性保证了数据的正确性和准确性。在大数据应用中，数据通常需要在多个节点之间进行复制和同步，这就可能影响到数据质量。通过选择适当的一致性策略，如强一致性、弱一致性或最终一致性，可以保证在系统规模增长的同时，数据质量不会受到影响。

可用性（Availability）与系统规模：可用性保证了系统的正常运行和服务。在大数据应用中，随着数据处理规模的增加，系统的可用性可能会受到影响。通过设计高可用的系统架构和选择适当的分区容忍性策略，

可以保证系统在面临网络分区等问题时，仍然能够正常运行。

分区容忍性（Partition Tolerance）与数据质量：分区容忍性是系统在面临网络分区时，仍然能够继续工作的能力。在大数据应用中，分区容忍性可以帮助在系统面临网络分区等问题时，仍然能够保证数据质量。例如，通过设计数据副本和同步策略，可以保证在分区发生时，数据仍然能够被正确地复制和同步。

（三）传统业务应用系统与警务大数据应用要共同生长

正视传统业务系统为警务大数据应用发展起到的基础数据支撑作用。公安信息化建设起步早，大多数系统建立在关系型数据库基础之上，业务系统成熟，数据资源积累厚，同时随着音视频等类型的数据不断积累，为警务大数据系统建设提供数据基础。尽管大数据技术对公安信息化产生巨大变革和影响，但是不会影响关系型数据库发挥基础数据支撑作用，在警务大数据应用建设中，重点是打破传统数据“壁垒”，实现数据共享融合，关系数据管理技术和非关系数据管理技术必将在公安信息化建设过程中不断竞争，互相取长补短，在新的大数据分析生态体系内找到自己的各自适用位置。

技术融合与创新：传统业务应用系统与大数据应用的共同生长，首先体现在技术层面的融合与创新。大数据技术可以为传统业务应用系统提供更智能、高效的数据处理和分析能力。同时，传统业务应用系统也可以为警务大数据应用提供丰富的业务场景和实际应用价值。这种技术融合与创新有助于提高公安信息化运营效率、降低成本、优化决策，从而提升实战竞争力。

业务模式变革：传统业务应用系统与

大数据应用的共同生长，还体现在业务模式的变革。大数据技术可以帮助业务部门更好地理解实战需求、把握公安业务趋势，从而推动精准警务新模式。^[12]同时，传统业务应用系统在引入大数据技术后，可以实现业务流程的优化、自动化和智能化，提高效率，开拓公安工作新领域。

数据驱动决策：大数据技术在传统业务应用系统中的融合应用，推动公安机关更加依赖数据进行决策。数据驱动的决策方法可以为公安机关提供更精准、实时的洞察力，帮助各部门快速响应变化，提高决策质量和效率，降低决策风险。

（四）警务大数据应用要协同发展

在数据规模呈指数级发展的驱动下，跨业务、跨警种的大数据融合应用需求必将不断涌现，突破了原有对公安业务领域的界定认识。因此，在大数据环境下，要适应新的科技革命和产业变革的需要，从新的社会和经济发展趋势来重新理解和认识警务大数据应用及其范畴，赋予其跨越现有业务边界的新内涵。随着社会经济的发展及其产业变革，不会因为对业务管理职能的界定而局限在某个单一的部门内，这必然突破原有的界限和划分。^[13]在各类业务应用系统需要采用新技术，系统应用效果呈现周期缩短等现实需求推动下，警务大数据应用建设局限在单位内部的壁垒必将被打破，警务大数据应用发展一定要走上协同发展、融合应用的道路。

警务大数据应用协同发展需要研究协调警务大数据应用建设中各主体关系。形成公安部门主导、行业指导、企业、高校和科研机构广泛参与的协同发展模式。在重视各警种部门间传统数据共享模式的同时，把社会数据资源作为大数据警务构成的重要组成部分，梳理细化各类信息资源对接需求，制定

数据标准规范, 指导广泛采集、吸纳社会数据资源, 必要时可以购买企业提供的数据服务和安全服务, 构建起公众、企业、政府部门共同参与的数字基础设施平台, 使得决策科学化、执行高效化、监督立体化。

(五) 扁平化管理模式下警务大数据应用优化策略

推动警务大数据应用在机关和基层的扁平化管理具有重要意义。扁平化管理模式下警务大数据应用优化策略体现了一种前瞻性的管理理念, 为警务系统在面对机关与基层数据异质性的挑战时提供了创新性的解决方案。扁平管理通过减少层级, 促进信息的自由流动, 从而加强了机关与基层之间的沟通。这种高效的信息传递机制有助于避免数据断层, 使得数据在不同层级之间更为畅通, 有利于整体数据协同与整合。扁平化管理模式下, 决策权更加分散, 使得基层警务人员更容易参与到数据应用的决策和优化过程中。这种参与度的提升激发了基层工作人员的创新潜力, 有利于找到更适应实际需求的数据应用策略。扁平管理减少了决策传递的环节, 使得在面对数据问题时能够更为迅速地做出反应和调整。这有助于机关更加灵活地根据基层数据的特点制定应对策略, 从而提高数据应用的针对性和实效性。扁平管理鼓励基层更多地参与管理决策, 有利于培养和发现更多具有数据分析与处理能力的人才。通过激励机制的调整, 可以更好地激发人员的积极性, 使得他们更愿意投入到警务大数据应用的优化工作中。扁平管理的灵活性和高效性使得整体绩效水平得以提升。通过机关与基层数据的协同优化, 整个警务系统能够更加迅速、精准地应对各类犯罪和安全挑战, 实现警务大数据应用的最佳效果。

尽管警务大数据应用蓬勃发展, 但与实际应用的成功相比, 其进展仍显不足。通过基于文献计量分析, 对警务大数据应用的研究成果进行深入审视, 为警务大数据发展提出新的努力方向。实践中, 对成果的梳理和提炼显得尤为关键。将警务大数据应用研究融入整个大数据研究的格局中, 不仅要关注警务领域的特殊性, 更要深入研究大数据研究领域的普适性问题。同时强调对原创性研究的重视, 以提高警务大数据应用研究成果在学术界和实际应用中的影响力; 强调理论总结与实践创新的结合, 以期更好地推动这一领域的发展, 使警务大数据应用能够更好地服务于社会安全与治安维护的实际需求。

注释:

- [1]人民网. 人民日报. 2013大数据元年. <http://cpc.people.com.cn/n/2013/1225/c83083-23940252.html> [EB/OL]. 2013. 12. 25
- [2]刘岩、刘铭. 基于CNKI的国内大数据研究热点及趋势分析[J]. 北京警察学院学报, 2020. 1
- [3]常轩. 公安数字化转型的内涵、逻辑与进路[J]. 公安研究, 2023. 8
- [4]王志华、庞海波、李占波. 一种适用于Hadoop云平台的访问控制方案[J]. 清华大学学报(自然科学版), 2014. 1
- [5]陈吉荣、乐嘉锦. 基于Hadoop生态系统的大数据解决方案综述[J]. 计算机工程与科学, 2013. 10
- [6]崔杰、李陶深、兰红星. 基于Hadoop的海量数据存储平台设计与开发[J]. 计算机研究与发展, 2012. S1
- [7]范建永、龙明、熊伟. 基于HBase的矢量空间数据分布式存储研究[J]. 地理与地理信息科学, 2012. 5
- [8]赵龙、江荣安. 基于Hive的海量搜索日志分析系统研究[J]. 计算机应用研究, 2013. 11
- [9]蒋勋、刘喜文. 大数据环境下面向知识服务的数据清洗研究[J]. 图书与情报, 2013. 5
- [10]李建中、王宏志、高宏. 大数据可用性的研究进展[J]. 软件学报, 2016. 7
- [11]Brewer, E. A. (2000). Towards robust distributed systems. In Proceedings of the 21st ACM SIGOPS symposium on Operating Systems Principles (pp. 205-214). ACM.
- [12]曾令德. 浅谈精准警务[J]. 公安研究, 2023. 7
- [13]林健. 面向未来的中国新工科建设[J]. 清华大学教育研究, 2017. 2

责任编辑 马煜童