

Big Data Application Study in Public Security Area

Jiaping Sun, Shinian Wang, Dan Liu

Nanjing Research Institute of Electronic Technology, Nanjing Jiangsu
Email: sjpwh@sina.com

Received: Nov. 21st, 2018; accepted: Dec. 6th, 2018; published: Dec. 13th, 2018

Abstract

The article gives out the requirement and status of the big data application in public security area in current period. Firstly, the article introduces the technical architecture of the big data system, describes the process procedure including data collection, data store, data process and application service. Secondly, the article analyses the typical application including police full text retrieval, vehicle big data, face identification and etc. Finally, the article gives out the discussion on the future trend of big data application in public security area.

Keywords

Big Data, Full Text Retrieval, Face Identification, Vehicle Big Data, Mobile Big Data

公共安全领域大数据应用研究

孙家平, 王诗年, 刘 丹

南京电子技术研究所, 江苏 南京
Email: sjpwh@sina.com

收稿日期: 2018年11月21日; 录用日期: 2018年12月6日; 发布日期: 2018年12月13日

摘 要

本文分析了现阶段大数据技术在公共安全领域的需求和应用现状, 介绍了公共安全大数据系统的技术架构, 描述了数据采集、数据存储、系统处理、应用服务的处理流程。后续分析了大数据应用的典型应用, 包含警务全文检索、车辆大数据、人脸大数据等。最后对公共安全领域的大数据发展趋势进行了展望和探讨。

关键词

大数据, 全文检索, 人脸识别大数据, 车辆大数据, 移动大数据

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着信息技术的广泛应用, 大数据已悄然来临。云计算、物联网等科学技术的应用与普及, 各种移动设备、无线传感器等无时无刻不在进行着数据的产生和交叉[1], 极大的催生了数据量的迅猛增加, 数据类型也已经不单纯局限于单一的结构化数据, 现代社会已经步入大数据时代[2]。

公安系统拥有庞大的数据积累和独特的安全需求, 迫切需要利用特别技术进行大数据处理[3]。据不完全统计, 截至 2015 年底, 全国公安机关掌握的数据资源已达数百类、上万亿条、EB 级的大数据规模。同时, 数据产生汇集的速度越来越快, 数据呈阶梯式增长。目前, 公安数据的年增长率超过 50%, 增长速度远超以往任何时期[4]。针对公安日常执法执勤中积累的大量数据, 传统的数据库技术已经无法适应, 相关数据分析和案件侦破的响应速度越来越慢, 建设警务数据分析平台, 深入分析各种信息资源和视频资料, 挖掘信息, 从大型数据库或数据仓库中提取隐含的、有潜在应用价值的信息或数据, 应用到公安领域, 来辅助一线实战。这项工作包括海量数据的应用、挖掘、碰撞比对, 以及各种技术的合成, 通过对数据的挖掘, 提升警务信息化。

随着警务大数据的应用, 为警方解决的问题越来越多。自 2011 年以来, 美国加州桑塔克鲁茨市犯罪情况较为严重。为帮助警方人员采集和分析材料, 提高办案效率, 基于大数据分析工具, 开发了犯罪预测系统, 进行犯罪模式分析。对可能出现犯罪的重点区域、重要时段进行预测, 并安排巡警巡逻。系统购入使用一年后。该市发案率大大减少, 入室行窃减少了 11%, 偷车减少了 8%, 抓捕率大幅上升了 56% [5]。今年来, 我国各地公安机关也在不断探索大数据的实战应用。广东、贵州、江苏、浙江等各地公安机关纷纷利用大数据开展国家安全预测、维稳态势预测、治安形势预测、社会管理预测、民意导向预测和民生服务等方面的应用探索工作[6]。

2. 警务大数据应用需求

随着大数据技术的逐步成熟和警务大数据系统的陆续上线运行, 为公安系统解决了越来越多的业务问题, 主要集中在三个方面: 快速发现目标、预测告警和精确布控。

1) 快速发现目标

通过将人口基本信息登记系统、视频监控系统、高速卡口记录系统、指纹登记系统等公安内部系统以及社会各行业系统中的数据进行整合, 提取有关涉案人员的相关信息, 实现对个人身份信息、视频、指纹甚至 DNA 等数据的融合处理、辨析[7], 实现在海量数据中快速寻找目标线索, 极大提高了相关案件的侦办效率。

2) 预测告警

充分利用大量的历史数据, 建立大数据分析模型, 对海量数据进行汇总分类、深度挖掘、分析研判, 搜索隐藏在海量数据中的高价值信息, 对将可能发生的事件进行预测。同时可以对互联网上的传输的数据进行深度分析, 采集特定对象的信息, 发现危险情报及时告警, 预防恶性事件发生。

3) 精确布控

通过在重要的通道或场所，部署人脸识别或车辆识别系统，对需要进行跟踪或抓捕的对象进行识别监控，一旦发现布控对象，将及时进行提示告警，实现对布控对象的及时处置。

3. 警务大数据架构

警务大数据系统一般按照分层设计的思想进行设计，完成数据的采集、存储、处理、应用过程，架构分为硬件支撑层、数据源层、接入采集层、存储管理层、数据处理层和业务应用层，系统整体框架如图 1 所示：

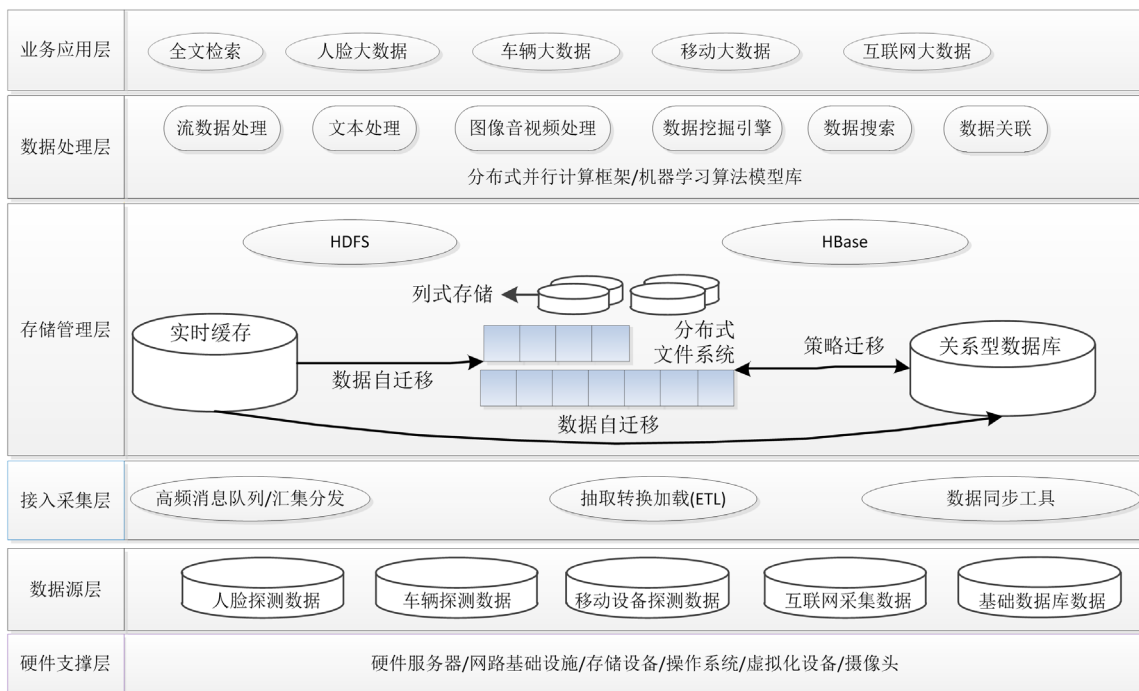


Figure 1. Architecture of police big data system
图 1. 警务大数据系统架构

3.1. 硬件支撑层

大数据基础数据平台由硬件服务器、网络基础设施、存储设备、时统设备、虚拟化设备等组成。通过核心交换机将多台硬件服务器、存储设备、时统设备等各类设备进行互联，拓扑成内部网络。由时统系统为各类设备提供高精度的授时服务。根据数据平台的需要，在设备上构建虚拟机，部署对应的操作系统，形成内部集群。为上层结构提供基础的支撑环境。

3.2. 数据源层

数据源层描述了各类数据源，包括人脸探测数据、车辆探测数据、移动设备探测数据、互联网采集数据以及基础数据库数据。基于公安相关业务平台，收集包含公安网、图像专网、业务专网、互联网平台上的各类数据。

3.3. 接入采集层

接入采集层实现数据从数据源到管理存储层的抽取、转换、加载和预处理。接入采集的方式分为消

息队列/汇集分发、ETL、数据库同步三种方式。在数据接入采集过程中支持自定义的数据订阅发布、自定义策略的数据同步、自定义预处理规则的 ETL 加载。

3.4. 存储管理层

数据的存储形式主要有两种：基于 HBase 的非关系型数据和基于 HDFS 非分布式文件系统。HBase 主要用来存储对随机、实时访问(读/写)性能要求较高的大数据，HDFS 主要用于对视音频、图像、大文档等大数据量文件的存储。HBase 和 HDFS 都具有数据备份功能以及良好的可扩展性。利用管理服务，设置数据迁移策略，支持存储结构之间的数据按策略迁移。

3.5. 数据处理层

针对关系复杂、资源消耗大的数据资源处理，设计分布式并行计算框架，充分利用服务器的计算资源。以分布式并行计算框架做支撑为复杂应用提供机器学习算法模型库以及各类处理引擎，包括流数据处理引擎、文本处理引擎、图像音视频处理引擎、数据挖掘引擎、数据搜索引擎和有用的数据关联处理。

3.6. 业务应用层

业务应用层分为数据访问服务和数据应用服务：数据访问服务在数据统一存储和管理的基础上，以中间件的形式提供规范的数据访问接口；数据应用服务则提供常用的应用操作接口，如检索、统计、比对，以及各种分析操作。应用领域包含全文检索、车辆大数据、人脸大数据、移动大数据和互联网大数据应用。

4. 大数据处理设计

大数据处理包含数据采集、数据存储和数据处理。数据采集完成数据的收集、分类和存储；数据存储按照应用业务整合业务数据并进行汇总存储；数据处理完成数据的分析计算。图 2 是典型的大数据处理流程：

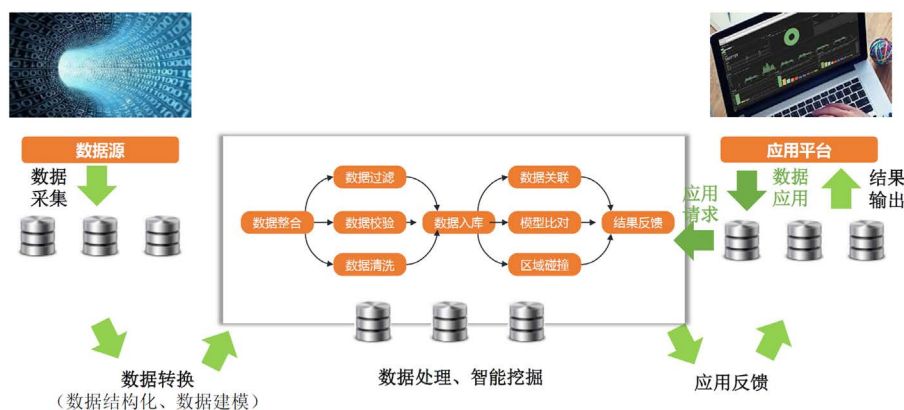


Figure 2. Typical large data processing flow
图 2. 典型大数据处理流程

4.1. 数据采集

数据采集提供对来源于公安系统的平台业务库、公安网、业务专网、图像网的内部数据源数据以及来源于第三方提供的数据和互联网数据的外部数据源数据进行采集、结构化提取以及入库功能。数据采集支持对视频、音频、文档、图片等数据的实时采集，同时支持设定定时任务进行数据定时采集。数据采集流程如图 3 所示：

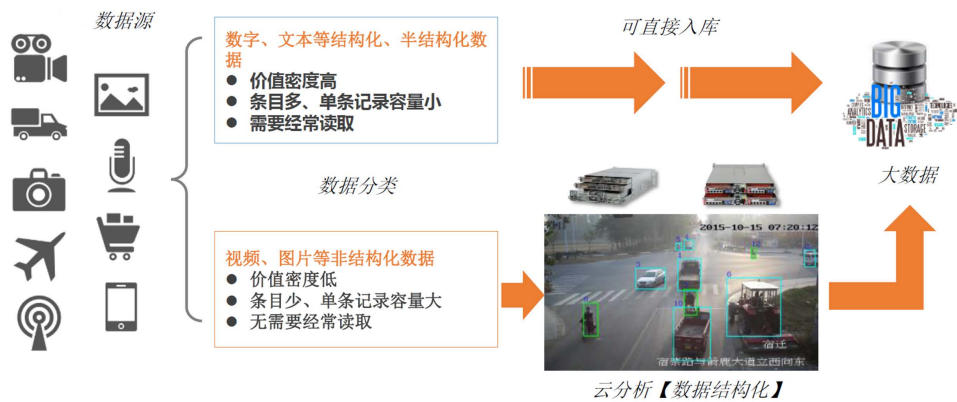


Figure 3. Data acquisition process
图 3. 数据采集流程

数据采集由数据分类、数据结构化提取以及数据入库几部分组成：

1) 数据分类

将数据分类为结构化数据、半结构化数据及非结构化数据。结构化数据、半结构化数据价值密度高、条目多、单条记录容量小、读取频率高，非结构化数据价值密度低、条目少、单条记录容量大、读取频率低。

2) 数据结构化提取

对非结构化数据进行特征提取，采用全量抽取、增量抽取等方式进行数据抽取，将非结构化数据转换为能够处理的结构化数据。

3) 数据入库

将结构化、半结构结构化数据以及非结构化数据进行入库或结构化提取入库。

4.2. 数据存储

数据存储提供按照应用业务整合明细业务数据，对采集的内、外部各资源数据和业务数据进行汇总存储的功能，同时具备数据备份功能。支持对音频、图像、大文档等大数据量文件的存储。

数据存储由关系型数据库存储、NoSQL 数据库存储、内存数据库存储以及分布式数据库存储组成。

1) 关系型数据库存储

对需要频繁进行事务处理的业务系统或已经按照传统技术架构建设完毕的业务系统产生的数据以及统一建设的应用系统产生的数据存储到关系型数据库。

2) NoSQL 数据库存储

对需要海量存储和高并发访问的数据存储到高扩展性、高可用的 NoSQL 数据库。

3) 内存数据库存储

对布控、可疑信息、热点信息等需要快速响应、实时性要求高的业务逻辑或数据存储到内存数据库。

4) 分布式数据库存储

对警务公安卡口、人口、视频信息等需要海量存储、高速读写、高效管理的各类数据存储到分布式数据库，存储在多个节点上，通过任务调度模块，将一个大的任务分解到多个节点上执行，提升系统的计算和传输性能。

4.3. 数据处理

数据处理提供对海量数据进行数据整合、数据计算处理分析的功能。数据处理由数据预处理、数据计算和数据分析和应用分析组成。数据处理流程如图 4 所示：

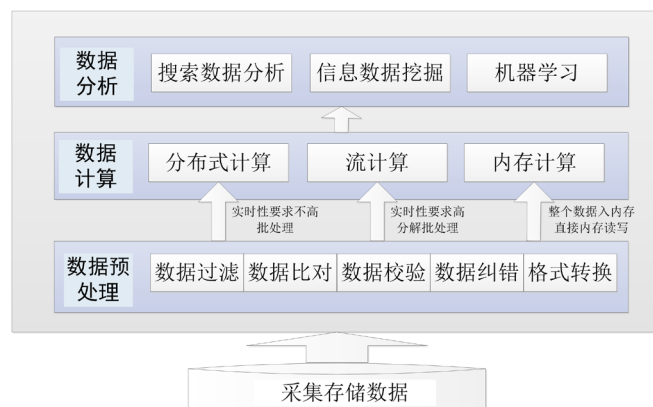


Figure 4. Data processing flow

图 4. 数据处理流程

1) 数据预处理

对采集接入的数据进行数据过滤、数据比对、数据校验、数据纠错等预处理操作等，提高数据的质量和关联性，为统一的存储和集中的管理提供基础，同时具备对需要报送的数据进行相应的格式转换的功能，转换为报送要求的格式。

数据过滤：根据用户定义的规则对数据进行过滤，过滤后的数据要求必须是合法、正确和可用的。

数据比对：根据定制的比对规则对数据进行比对，发现不同部门之间的信息存在不一致性时，将这些信息分别记录下来进行比对，生成数据比对报告。

数据纠错：根据定制的纠错规则对数据进行自动纠错。纠错过程中，对出错的数据进行记录进行纠错，生成数据纠错报告。

数据校验：根据定制的校验规则(如数据格式、长度等)对数据进行自动校验。校验包括两个方面：数据本身的合法性和数据的逻辑错误。

格式转换：根据系统需要报送的数据格式要求完成信息之间格式的转换。

2) 数据计算

在海量数据情况下，采用计算引擎来应对线下及线上的计算需求，采用计算框架来为数据构建模型，并对数据进行处理，提高计算能力。

分布式计算：通过调度批量任务操作静态数据，进行大规模数据集的并行运算，采用分布式计算架构进行日志分析统计、大规模索引、海量数据排序、词频统计和历史数据挖掘分析等实时性要求不高的业务操作，用于批处理。

流式计算：通过软件算法在传输数据时对其进行分析，将具有高优先度的数据传输给针对解决复杂、快速变化的问题优化的应用，将计算分解成一系列短小的批处理作业，中间结果保存在内存中，再对中间的结果进行叠加，用于实时处理。

内存计算：对于信息比对、专题分析、数据挖掘等功能，将整个数据放入内存，使每个事务在执行过程中没有 I/O，直接进行内存数据读写，完成大批量的运算。

3) 数据分析

搜索数据分析：按照人、车、组织单位、地址、案事件等信息及相关信息进行全文搜索，快速从系统内部查找并列有效信息。

信息数据挖掘：从数据中自动地抽取模式、关联、变化、异常和有意义的结构，进行关联分析、分类分析、聚类分析、序列分析等，建立挖掘模型，利用已有的海量数据，分析案事件关联、串并案，对高危人群及地区犯罪进行预警。

机器学习：警用相关的机器学习主要包含行为分析和智能推荐。机器学习行为分析是通过警察或公众行为日志，分析用户的平台内访问路径、访问深度及页面停留时间，细分用户群体。机器学习智能推荐主要是基于警察行为分析结果，按照相同角色、相同地区、相同职务、相同业务构建推荐模型，给警察推荐可能感兴趣的功能模块。

4) 应用分析

应用分析提供在数据统一存储、管理、访问的基础上，对警务不同领域大数据应用的定制分析，对长期积累的数据和实时获取的海量数据进行深度挖掘，获取高价值信息。应用领域主要包含警务综合检索、车辆分析、人脸识别和互联网应用等。

5. 典型警务大数据应用

典型的警务大数据应用领域包含全文检索应用、车辆应用、人脸应用、互联网应用等。

1) 全文检索应用

提供大数据服务，以警务档案系统为基础，通过信息的有效串联，提供各层次的通用搜索服务。

建立档案：建立人、案、物档案信息，建立与主体相关联的各类关联信息。

通用搜索：以人、案/事件、物品、网页、地图、文库等要素进行通用搜索。

轨迹查询：以地点、时间为线索检索轨迹存在交集的人员，实现人员轨迹的横向关联。

由浪潮公司开发的警务千度系统，提供了典型的全文检索应用，民警在搜索框中输入任意搜索词，搜索后，系统会根据这些搜索词在全文检索的分布式索引库中进行检索，得到的搜索结果会按照搜索词与结果的匹配程度从高到低进行排序。民警对搜索结果还可进一步分类筛选。



Figure 5. Full text retrieval example
图 5. 全文检索示例

图 5 中，使用者可输入限定的条件(如年龄、籍贯、民族等)以缩小搜索范围。系统可以提供模糊检索等功能。获取人员信息后，可以进一步获得该人员的基本信息、人际关系、轨迹信息等。人员轨迹信息示例如图 6：

非 轨迹信息			
时间轴			
火车搭乘 (3)			
		不限	至
		不限	查询
日期	事项	类别	功能
2013-12-10	乘坐K1362次, 从青岛前往济南。	火车	专题应用
2013-12-08	乘坐1566次, 从济南前往青岛。	火车	专题应用
2013-02-08	乘坐5011次, 从济南前往东营。	火车	专题应用
飞机搭乘 (2)			
		不限	至
		不限	查询
日期	事项	类别	功能
2012-11-23	乘坐SC4762次, 从济南遥墙前往重庆江北。	飞机	专题应用
2012-08-27	乘坐SC4943次, 从济南遥墙前往重庆江北。	飞机	专题应用

Figure 6. An example of personnel trajectory information
图 6. 人员轨迹信息示例

2) 车辆应用

通过交通采集设备收集的信息，以车辆为主体，对车辆信息进行查询分析，确定车辆轨迹或涉案信息。

数据查询：对海量过车数据进行快速检索，如车辆查询、普通过车查询、布控报警查询、异常牌照查询、行车轨迹查询等。

以图搜图：基于图片的搜索模式，通过图片建模后特征向量的比对分析，同时支持被搜索图片中选择特征区域进行二次比对，从图片库中检索出符合条件的图片，并根据相似度返回比对结果。

数据研判：基于不同的研判算法对车辆数据进行综合研判，如初次入城研判、频繁出入车辆研判、行车轨迹研判、违法多发地点研判等。

联防布控：根据车牌(黑名单、红名单)或车辆特征，对车辆进行联防布控，可设定布控时间段。

以海康车辆大数据产品为例，首先通过摄像头对车辆进行前端抓拍，快速提取车辆特征，包含车牌、车型、品牌、车身颜色等，对车辆进行建模。可实现对车辆的统计分析、违法识别以及智能布控等功能。下图是特征搜车示例，用于对车牌污损或人为拆除牌照的车辆以图搜图：



Figure 7. Characteristic search example
图 7. 特征搜车示例

图 7 中，车辆未挂牌照，系统对车辆抓拍图片的框选区域(区域内有年检标签和饮料瓶)进行建模，对带牌照的车辆进行比对，找到有牌照的卡口抓拍照片，通过车牌进一步追查该车辆的登记信息。

3) 人脸识别应用

在人脸目标库建立基础上，支持进行人脸图像检索、人脸图像验证，快速、高效的辨别犯罪嫌疑人，实现人脸布控。

人脸图像检索：1:N 的检索，直接确认人员身份；

人脸图像验证：1:1 的检索，将不同案例视频中的人脸进行截取、比对、判断不同视频中的人是否为同一人。

人脸布控：在交通要道、室外场所、娱乐场所、住宿场所、商业场所进行布控，对人脸进行识别，对布控名单人员进行告警。

与车辆大数据产品类似，首先通过摄像头对人脸进行前端抓拍，快速提取人脸特征，对人脸进行建模，形成脸谱库。将抓拍的人脸照片放在静态库中进行比对，可获取人员的身份信息；同样可以脸搜脸，分析目标人物的轨迹和落脚点等。

图 8 为人脸区域碰撞分析。

人脸区域碰撞分析可用于公安串并案分析，按相似度依次罗列两个案发地点出现的相同人员，辅以人工筛选，可发现在案发时间都出现在案发现场附件的嫌疑人。



Figure 8. An example of face region collision analysis
图 8. 人脸区域碰撞分析示例

4) 互联网应用

互联网应用主要通过通信拦截方式或者公开网络信息收集方式获取数据，对数据进行分析和解码，提取出有价值的情报。可分为网络监听和舆情分析：

网络监听：通过技术手段记录和统计网民的上网行为，对网络进行全面监控，对互联网上传输的信息进行分析，包括采集特定对象的信息，发现危险人员情报，预防恶劣事件的发生。

舆情分析：通过互联网海量数据采集技术、中文语义分析技术和数据挖掘技术，全时全天候对境内外公开互联网信息进行监测，具备舆情发现、分析、预警和报告等功能。

监听一般通过代理程序对目标的 HTTPS/HTTP 协议数据、邮件或社交聊天数据进行拦截，收集后发送至后台记录，然后对数据进行综合分析处理后，输出数据。对于有价值数据线索，采用短信或邮件方式进行告警。

舆情分析主要使用网络爬虫工具，对公开的新闻、论坛信息进行检测，及时发现舆论动向和敏感话题，对未来形势发展进行早期预测。

6. 结束语

大数据的核心就是预测，这点与公安工作异曲同工。将大数据技术与预测警情相结合，必定会对公安预防工作的方式和内容产生极大影响。对公安工作而言，随着警务人员面临的问题日趋复杂化、信息规模化，能否充分运用大数据正日益成为制约我国公安工作顺利发展的关键因素[8]。

本文研究了大数据技术在警务系统中的思路，包含技术体系架构和关键技术解决思路和方法。同时对警务大数据的典型应用现状进行描述。随着警务大数据系统的广泛深入应用，将会为公安系统提供越来越重要的辅助手段，并将在公安实际工作中起到更大的作用。

参考文献

- [1] 姜浩端. 大数据的本质及其可能的影响[J]. 中国经济报告, 2013(6): 16-22.
- [2] 张文元, 马丽雅. 警务大数据在公安实践中的应用[J]. 中国人民公安大学学报(自然科学版), 2016(2): 51-54.
- [3] 吕益民, 宫夏屹, 聂春华. 公安大数据处理平台的探索[J]. 广东公安科技, 2014, 22(3): 13-18.

-
- [4] 王楠韩, 鹏霄. 公安大数据应用研究[J]. 警察技术, 2016(5): 3-7.
- [5] 安晖. 美国大数据维稳镜鉴[J]. 人民论坛, 2014(12): 61-63.
- [6] 刘向荣, 农忠海, 陈雅. 公安大数据应用研究的几点思考[J]. 产业观察, 2016(11): 38-41.
- [7] 冯冠筹. 大数据时代实施预测警务探究[J]. 广东公安科技, 2014, 22(1): 23-27.
- [8] 张鹤飞. 大数据技术与公安工作: 应用与挑战[J]. 现代警务, 2015(2): 77-81.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2330-4677, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jsst@hanspub.org