

公共安全

大数据

技术与应用



BIG DATA

齐力 / 主编

从数据源头、技术框架和平台体系的多层角度描绘了公共安全大数据发展的整体框架，合理地制定公共安全大数据的发展规划与建设路线，明确公共安全大数据应用的未来重点，加快促进公共安全大数据在实际的公安行业中的落地应用。



上海科学技术出版社

非外借

公共安全大数据技术与应用

齐 力 主 编



上海科学技术出版社

图书在版编目(CIP)数据

公共安全大数据技术与应用 / 齐力主编. —上海:
上海科学技术出版社, 2017. 12

ISBN 978-7-5478-3748-1

I. ①公… II. ①齐… III. ①公共安全—安全管理—数据处理—研究 IV. ①D035.29-39

中国版本图书馆 CIP 数据核字(2017)第 257850 号

公共安全大数据技术与应用

齐 力 主 编

上海世纪出版(集团)有限公司 出版、发行
上海科学技术出版社

(上海钦州南路 71 号 邮政编码 200235 www.sstp.cn)

苏州望电印刷有限公司印刷

开本 787×1092 1/16 印张 12.5

字数: 360 千字

2017 年 12 月第 1 版 2017 年 12 月第 1 次印刷

ISBN 978-7-5478-3748-1/TP·55

定价: 98.00 元

本书如有缺页、错装或损坏等严重质量问题,
请向工厂联系调换

内容提要



本书的编写目的在于明确公共安全大数据设计的相关技术,明确公共安全大数据相关理念、应用以及发展的方向、路线。从数据源头、技术框架和平台体系的多层角度描绘了公共安全大数据发展的整体框架,合理地制定公共安全大数据的发展规划与建设路线,明确公共安全大数据应用的未来重点,加快促进公共安全大数据在实际的公安行业中的落地应用。最后,本书提出确定了公共安全大数据的架构体系,包括相关公共安全数据的典型运营方式、数据架构体系、平台层次结构等,为进一步研究和推进公共安全大数据确定了指导路线。

本书的主要读者是公共安全领域的技术人员、管理人员,以及高校相关专业的师生。

公 共 安 全 大 数 据 技 术 与 应 用

编委会



主 编

齐 力

编 委

徐 峥 白 翔 顾俊逸 孙傲冰 周亚琴
王潜心 蔡 烜 石宣化 戴 杰 张兆翔

前言



党和国家领导人对大数据,尤其是大数据在公共安全领域的结合应用高度重视。2017年2月17日,中共中央总书记、国家主席、中央军委主席、中央国家安全委员会主席习近平在北京主持召开国家安全工作座谈会并发表重要讲话。习近平强调要筑牢网络安全防线,提高网络安全保障水平,强化关键信息基础设施防护,加大核心技术研发力度和市场化引导,加强网络安全预警监测,确保大数据安全,实现全天候全方位感知和有效防护。

公安部第三研究所建设公安科技领域具有全国领先水平的科研机构为战略目标,以“敢为人先,引领发展,科技强警,励志报国”的核心价值观为引领,按照“科研和实战一体化、科研和产业一体化、科研和检测一体化、科研和教育培训一体化”的指导思想,坚持与时俱进、开拓创新,热情服务,不断破解公安实战中的关键技术难题,为科技强警做出新的贡献。公安部第三研究所主要研究领域包括大数据、信息网络安全、物联网、特种通信、禁毒、反恐防爆、图像处理和传输以及社会公共安全防范技术等。在eID(网络身份认证)、RFID(射频识别)、VSD(视频结构化描述)、PDD(警用数字化单兵)等领域取得显著成绩,多次获得国家级、省部级科学技术奖励。

为了充分发挥公安部第三研究所在公共安全领域国家队的规划、指导、应用实践作用,促进我国公共安全大数据的标准体系建设,推动公共安全大数据在实际的公共安全领域行业中的发展应用,编写组联合了华中科技大学、武汉大学、中国科学院自动化所、北京大数据研究院、中国地质大学(武汉)、中国矿业大学、中国科学院云计算中心、上海大学等多家单位,共同推动本书的编写工作。本书的编写旨在全面、客观、系统地梳理国内公共安全大数据相关的基础研究和实战应用,根据公安行业的最新实践、认识基础、研究进展,总结凝练公共安全大数据的

需求、问题、应用,提出公共安全行业的发展建议。

本书的编写目标在于明确公共安全大数据设计的相关技术,明确公共安全大数据相关理念、应用以及发展的方向、路线。从数据源头、技术框架和平台体系的多层角度描绘了公共安全大数据发展的整体框架,合理制定公共安全大数据的发展规划与建设路线,明确公共安全大数据应用的未来重点,加快促进公共安全大数据在实际的公安行业中的落地应用。

本书可以帮助公共安全行业的从业人员用来界定公共安全大数据的范畴和特征,明确公共安全大数据的总体目标和相关的研究基础。通过对公共安全大数据的相关支撑技术的综述,分析了国内公共安全大数据的发展现状,总结明确了目前公共安全大数据所存在的问题和难点,为行业下一步的良性发展理清了思路 and 方向。

最后,本书提出确定了公共安全大数据的架构体系,包括相关公共安全数据的典型运营方式、数据架构体系、平台层次结构等,为进一步研究和推进公共安全大数据确定了指导路线。

作者

2017年9月

目录



.....	第 1 章 绪论	1
	1.1 公共安全大数据的定义和特点	2
	1.1.1 公共安全大数据的定义	2
	1.1.2 公共安全大数据的特征	2
	1.2 公共安全大数据的挑战和关键问题	3
	1.2.1 公共安全大数据的挑战	3
	1.2.2 公共安全大数据的关键问题	3
	1.3 公共安全大数据相关政策法规	4
.....	第 2 章 公共安全大数据采集技术	7
	2.1 采集对象和手段	8
	2.1.1 采集对象	8
	2.1.2 采集手段	9
	2.1.3 外部接口	9
	2.1.4 分析手段	11
	2.2 通用技术	12
	2.2.1 ETL	12
	2.2.2 爬虫	12
	2.2.3 遥感	13
	2.2.4 监控	16

17	2.2.5	传输手段
19	2.2.6	海量数据存储技术
20	2.3	业务应用
20	2.3.1	智能采集缓解警力压力
21	2.3.2	遥感“千里眼”助战实战
22		参考文献
25	第3章 公共安全大数据处理技术	
26	3.1	公共安全大数据的数据类型
26	3.1.1	结构化数据
28	3.1.2	非结构化数据
30	3.1.3	半结构化数据
32	3.2	通用大数据处理技术
32	3.2.1	数据降维与压缩
33	3.2.2	数据去噪
34	3.2.3	数据清洗
34	3.2.4	信息提取
40	3.2.5	多源融合
41	3.2.6	数据处理平台——Spark
42	3.3	业务应用与实例
42	3.3.1	视频侦查
44	3.3.2	司法鉴定
48		参考文献
49	第4章 公共安全大数据分析挖掘	
50	4.1	公共安全大数据分析挖掘分类
50	4.1.1	人工分析
50	4.1.2	智能分析
50	4.1.3	辅助分析
50	4.2	公共安全大数据分析挖掘技术
50	4.2.1	时空分析技术



4.2.2	多源数据融合分析技术	52
4.2.3	视觉信息分析技术	54
4.2.4	语音识别	59
4.2.5	文本分析	60
4.2.6	预测模型	60
4.2.7	行人检测	61
4.2.8	车辆检测与识别	62
4.2.9	文字识别	64
4.3	视频数据智能分析挖掘	71
4.3.1	视频数据快速准确处理需求分析	71
4.3.2	当前视频监控系统依然存在问题	71
4.3.3	视频结构化描述技术概述	72
4.3.4	基于视频结构化描述的视频侦查技术	75
4.4	警务数据智能分析挖掘	83
4.4.1	警务数据知识库构建需求分析	83
4.4.2	警务数据知识库构建方法	84
4.4.3	警务数据知识库的关键技术	85
4.5	以人为中心的人证合一数据分析挖掘	87
4.5.1	人证合一分析挖掘需求分析	87
4.5.2	人脸检测技术	88
4.5.3	人证合一关键技术	91
4.5.4	人证合一识别与验证	92
4.5.5	人证合一场景应用	93
	参考文献	96

第5章 公共安全大数据可视化 101

5.1 大数据可视化基础 102

5.1.1 大数据可视化概念 102

5.1.2 数据可视化详解 103

5.1.3 可视化设计原则 105

6.4.2	基于身份和规则的安全策略	148
6.4.1	安全策略实施原则	147
6.4	大数据安全策略	147
6.3.5	访问控制	144
6.3.4	系统身份认证	143
6.3.3	大数据传输安全	142
6.3.2	数据加密技术	141
6.3.1	数据隐藏	140
6.3	大数据安全技术	140
6.2	大数据安全防护技术	139
6.1.5	安全管理制度	139
6.1.4	大数据渗透模型	135
6.1.3	大数据安全威胁因素	135
6.1.2	大数据安全特点	134
6.1.1	大数据安全的定义	134
6.1	大数据安全	134
6	公共安全大数据安全体系	133

5.4	业务应用——可视化情报分析	131
5.3.4	VR/AR技术	129
5.3.3	交互技术	128
5.3.2	数据分析	128
5.3.1	数据管理	127
5.3	可视分析系统框架设计	126
5.2.5	可视化展示手段	124
5.2.4	跨媒体数据可视化	118
5.2.3	时空数据可视化	118
5.2.2	层次和网络数据可视化	112
5.2.1	文本可视化	109
5.2	可视化分类与方法	109
	参考文献	132



6.4.3 综合访问控制策略 148

6.5 公安业务安全 149

6.5.1 网络安全防范 150

6.5.2 业务安全 151

参考文献 151

第7章 公共安全大数据案例 153

7.1 视频结构化描述实战平台 154

7.1.1 视频数据实战管理需求分析 154

7.1.2 视频结构化描述技术平台架构 154

7.1.3 视频结构化描述云平台 157

7.1.4 视频结构化描述技术应用实例 158

7.2 基于云计算的大数据实战 160

7.2.1 智慧城市大数据开放共享平台 160

7.2.2 生态安全大数据创新应用平台 164

7.2.3 公共服务类大数据创新应用 165

参考文献 166

第8章 公共安全大数据标准 167

8.1 大数据标准体系 168

8.1.1 需求分析 168

8.1.2 大数据标准孵化过程 168

8.1.3 标准体系结构 172

8.2 公安元数据标准 173

8.2.1 需求分析 173

8.2.2 标准体系结构 174

8.2.3 标准明细介绍 174

8.3 公安视频图像描述标准 175

8.3.1 需求分析 175

8.3.2 标准体系结构 176

8.3.3 标准明细介绍 176

8.4 公安视频智能分析技术的标准数据集及评测标准	177
8.4.1 需求分析	177
8.4.2 标准体系结构	178
8.4.3 标准明细介绍	179
参考文献	182

.....	缩略语表	183
-------	-------------	-----

.....	图表索引	185
-------	-------------	-----



绪 论

1.1 公共安全大数据的定义和特点

1.1.1 公共安全大数据的定义

公共安全大数据是指围绕社会公共安全需求,国家政策法规允许的,用于支持公共安全保卫的所有数据。

按照数据采集方式来区分,公共安全大数据的主要数据来源有以下三类:

第一类是对象被动产生的数据。这类数据主要是通过强制的法规或者各种手段,公共安全案事件涉及对象产生的数据。例如宾馆住宿时需要登记身份证信息,乘坐飞机高铁需要进行安检等。

第二类是对象主动产生的数据。这类数据主要是公共安全案事件涉及对象在案事件过程中,为了达到犯案目的,在犯案过程中所主动产生的数据。例如同伙之间的通联数据,案发现场留下的生物特征信息等。

第三类是对象自动产生的数据。这类数据主要是从对象身上自动获取的数据。例如人的定位信息,车辆的定位信息等。

公共安全大数据涉及的技术是指针对公共安全大数据,采用挖掘、分析、提炼等手段获取其相应的价值,并且进行有效的展示与研判的一系列技术与方法,包括数据采集、预处理、存储、分析挖掘、可视化、数据安全等过程。公共安全大数据的应用,是针对特定的公共安全大数据集,采用特定的技术方法,获取特定相关应用的有效数据价值的过程。

1.1.2 公共安全大数据的特征

公共安全大数据具有一般大数据的特征,包含以下四个方面:

(1) 数据量巨大(volume): 公共安全大数据的数据量规模巨大,单以视频监控举例,视频数据有着非常巨大的容积,以一个城市为例,安装了多台摄像头,每台摄像头每天收集超过 GB 数据量级的高清视频数据。

(2) 多样性复杂(variety): 公共安全大数据的数据类型多样,数据来源众多,数据模态多类。

(3) 数据产生速度快(velocity): 公共安全大数据产生的大多是实时性数据,需要极快



的处理速度,同时由于案件的快速分析需求,需要对数据的分析也需要极快的速度。例如视频数据,需要及时的处理与分析。

(4) 数据价值密度低(value): 公共安全大数据产生的大量数据是无价值的,有价值的数据往往需要及时的处理与分析。

公共安全大数据除了具有上述一般大数据的4V特征之外,还包含以下四个方面(简称公共安全大数据的4P特征):

(1) 强政策性(policy): 公共安全大数据的采集、处理、分析等过程,高度依赖于国家相应的法规政策。在法规政策范围允许内的数据,才可以被采集。

(2) 强私密性(privacy): 区别于一般数据,公共安全数据很大一部分是与对象相关的隐私数据,例如地理位置信息、通联记录等。因此,公共安全大数据具有隐私性,通过统计方法或其他数据挖掘技术来提取隐藏的信息和相关性。而提取出的价值与相关性要平衡于与公共利益、群体利益无关而且个人或团体不愿意被外界所知的信息。

(3) 高精准性(precision): 公共安全大数据的挖掘分析结果需要极高的精准性,公共安全事关到人民群众的最高利益,因此必须做到最精准的处理。

(4) 高时效性(promptness): 公共安全的趋势主要为事中快速响应,事后准确溯源,事前精准预防预警,因此公共安全大数据的分析、挖掘要求极高的时效性。

1.2 公共安全大数据的挑战和关键问题

1.2.1 公共安全大数据的挑战

大数据本身是一把“双刃剑”,对于公共安全行业来说,既带来了前所未有的机遇,也相伴而生了许多挑战。

(1) 大数据带来了公共安全领域数据处理成本与收益之间的矛盾。大数据的一个重要特性是海量性,而数据规模越大,必然导致存储成本的上升。由于大数据强调在全量数据中进行挖掘分析而非传统的抽样调查,因此增加了处理成本。如何快速地过滤无价值的数,对于公共安全数据进行准确的处理是一个重要的挑战。

(2) 大数据带来了公共安全数据互联互通需求与管理体制之间的矛盾。大数据的重要特性是建立数据之间的关联,通过关联挖掘提取数据的价值。但是当前的管理体制是由于各类安全数据之间缺乏统一的标准,现有组织、部门、制度间的分割以及信息管理理念的滞后,往往导致“数据孤岛”现象的出现。

(3) 大数据带来了公共安全数据处理技术的快速发展与缺乏相应的标准之间的矛盾。大数据技术发展迅速,但是仍然缺乏相应的标准对其进行规范。

1.2.2 公共安全大数据的关键问题

为了应对公共安全大数据的几个挑战,本书提出了公共安全大数据所涉及的几个关键

问题。

(1) 如何将数据由存不起转变为存得了。大数据的重要观点是对全量数据进行分析,在公共安全领域迫切的是需要解决数据存储安全与空间成本的问题。数据存储多久,如何存储,采用分布式还是集中式,都是亟需解决的问题。

(2) 如何将数据由联不通转变为互联互通。大数据的重要观点是对数据进行关联分析,然后从中获取数据的价值。由于体制、数据类型、数据模态等多种问题,公共安全相关的数据依然无法做到有效的互联互通。如何对建立数据之间的联通机制、如何对数据进行有效的关联融合,也是亟需考虑的问题。

(3) 如何将数据由找不准转变为找得到,看得准,挖得深。目前对于公共安全相关的数据处理仍然缺乏非常有效的手段。例如视频,依然无法做到非常精准的对象识别,因此仍需要采用有效的数据分析手段,把原始的非结构化的数据转变为结构化的可理解、可分析的数据。

1.3 公共安全大数据相关政策法规

进入 21 世纪,尤其是党的十八大以来,以习近平同志为核心的党中央站在时代科技发展的前沿,为大数据指明发展方向,扫清各种障碍。2016 年 10 月 9 日,习近平总书记在中央政治局会议上提出,要以数据集中和共享为途径,建设全国一体化的国家大数据中心,推进技术融合、业务融合、数据融合,实现跨层级、跨地域、跨系统、跨部门、跨业务的协同管理和服务。在政策、法律等制度层面,党中央和国务院为国家大数据战略的实施作出了一系列重要的规划和部署。2015 年 3 月,十二届人大三次会议提出实施“互联网+”战略,加快制定有关行动计划;2015 年 11 月,党的十八届五中全会提出《促进大数据发展行动纲要》,将大数据直接上升为国家战略。2015 年 12 月,十二届人大十八次会议通过《反恐怖主义法》和《网络安全法》,明确规定电信业务经营者和互联网服务供应商应当依法为公安机关、国家安全机关防范、调查恐怖活动提供技术接口和解密等技术服务,为反恐领域的协作提供了法律依据。

有关大数据发展的法律法规包括:

- (1) 《2006—2020 年国家信息化发展纲要》(2006 年 5 月);
- (2) 《“十三五”国家信息化规划》(国发〔2016〕73 号);
- (3) 《中华人民共和国国民经济和社会发展第十三个五年规划纲要》(2015 年 11 月);
- (4) 《促进大数据发展行动纲要》(国发〔2015〕50 号,2015 年 8 月);
- (5) 《国家信息化发展战略纲要》(2016 年 7 月);
- (6) 《中华人民共和国反恐怖主义法》(2016 年 1 月);
- (7) 《中华人民共和国网络安全法》(2017 年 6 月)。

另外,国务院发布的《促进大数据发展行动纲要》提出,要加快政府数据开放共享,推动