

|深|入|浅|出|物|联|网|技|术|丛|书|

曹振丽 著◎



物联网

追溯系统及数据处理



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

深入浅出物联网技术丛书

物联网追溯系统及数据处理

曹振丽 著



电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

数据流处理技术是目前的研究热点,掌握数据流处理技术有助于更好地利用大数据,挖掘出数据背后潜在的价值。

本书围绕物联网追溯系统的研发及数据流处理过程中的聚类、追溯、预测与建模关键技术进行了研究。全书共6章。第1章是绪论,主要介绍物联网的概念、中国农业物联网产业化发展现状,以及国内外大数据的研究现状和热点。第2~5章,主要介绍数据流聚类算法、数据流追溯方法、数据流预测方法、数据流建模方法。第6章,介绍了物联网追溯系统的研发和软硬件环境配置,并对本书的主要内容进行梳理和总结。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

物联网追溯系统及数据处理 / 曹振丽著. —北京: 电子工业出版社, 2019.5
(深入浅出物联网技术丛书)

ISBN 978-7-121-36127-2

I. ①物… II. ①曹… III. ①互联网络—应用—研究②智能技术—应用—研究③数据处理—研究
IV. ①TP393.4②TP18③TP274

中国版本图书馆 CIP 数据核字 (2019) 第 043867 号

策划编辑: 刘志红

责任编辑: 刘志红 特约编辑: 李 姣

印 刷: 三河市鑫金马印装有限公司

装 订: 三河市鑫金马印装有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×980 1/16 印张: 11.75 字数: 149 千字

版 次: 2019 年 5 月第 1 版

印 次: 2019 年 5 月第 1 次印刷

定 价: 69.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: (010) 88254799, lzhmails@phei.com.cn。

作者简介

曹振丽（1979.11-）女，博士，山东烟台人，滨州医学院公共卫生与管理学院教师，长期从事计算机应用技术教学工作，研究方向：物联网、云计算、大数据、人工智能。

本专著为烟台市科技计划“基于物联网与云计算的猪肉安全追溯平台关键技术的研究”（项目编号：2016ZH079）阶段性研究成果之一；滨州医学院博士科研启动基金“物联网环境下的猪肉安全追溯平台关键技术的研究”（项目编号：BY2015KY-QD13）阶段性研究成果之一。



前 言

随着物联网、互联网、云计算等技术的飞速发展,在各个领域出现了大规模的数据增长,信息社会已经进入了大数据时代。大数据主要有数据流和静态数据两种形式,智慧农业中各种传感器产生的数据流是农业大数据的主要来源。农业领域中的数据流来源众多,形式多样,处理复杂,很难有一种计算模式能涵盖不同的计算需求,因此,如何根据数据流的不同数据特征和计算特征,从多样性的计算问题和实际需求中提炼并建立高层抽象模型,是目前数据流研究亟待解决的问题。

本书围绕物联网追溯系统的研发及数据流处理过程中的聚类、追溯、预测与建模关键技术进行了研究,取得了一些成果,主要内容如下。

第 1 章,主要介绍了物联网的概念,以及中国农业物联网产业化发展现状,并阐述了国内外大数据的研究现状和热点。

第 2 章,主要介绍了基于高斯混合模型的数据流聚类算法 Cumicro。使用高斯混合模型作为数据流中不确定数据的基本表示形式,更好地利用存储空间,完成对真实情况的逼近。该算法将时间直接作为数据属性,可直接查询某个时间维度的聚簇。实验证明,该算法在原始数据较密集时,与原有基于离散模型的聚类相比,该算法具有准确度上的优势。

第 3 章,主要介绍了基于不确定数据的数据流追溯方法,将不确定数据引入追溯系统中,搭建追溯模型,解决了数据流背景下无法对可追溯单元一一标识的混合

过程进行表示的问题。利用不确定数据的基本表示和查询方法,解决了多源追溯问题,完成了数据流追溯模型中的一般查询、节点评价和单节点异常推断功能,并给出了多节点异常的求解方法。

第4章,主要介绍了基于时间粒度的自适应调整灰色二阶模型的数据流预测方法,通过实验得出以下结论:随着滑动窗口更新周期的增大,预测的成功率反而下降;随着采样频率的变大,预测成功率降低;随着未来数据窗口宽度的增加,预测的平均相对误差增大。但该方法对近期的数据预测比较准确,满足了系统的需求。

第5章,主要介绍了一个面向养殖环境的猪舍数据流采集与预测为一体的自动化控制系统,提出了在数据流背景下切换到数据视角,以数据为中心来开展业务研究,将时间和空间的事件信息以数据驱动为中心,明确地抽象到编程模型中,进行形式化的描述及一体化建模,打破了传统的建模方法仅局限于时间域内分析的局限性,考虑了计算过程和物理过程通过网络实时交互对系统行为所带来的影响。

第6章,主要介绍了物联网追溯系统的研发和软硬件配置环境,对本书的主要内容进行梳理和总结。

在本书撰写的过程中,中国农业大学孙瑞志教授给予了指导性意见,尹宝全博士、李勖博士、聂娟博士、邓雪峰博士、王雷雨硕士都给予了大量的支持和帮助,提出了很多宝贵意见,在此表示感谢。

由于作者水平有限,不当之处在所难免,恳请广大读者批评指正。

作者

2018年10月

内容简介

数据流处理技术是目前的研究热点，掌握数据流处理技术有助于更好地利用大数据，挖掘出数据背后潜在的价值。

本书围绕物联网追溯系统的研发及数据流处理过程中的聚类、追溯、预测与建模关键技术进行了研究。全书共6章，第1章是绪论，主要介绍物联网的概念、中国农业物联网产业化发展现状，以及国内外大数据的研究现状和热点。第2~5章，主要介绍数据流聚类算法、数据流追溯方法、数据流预测方法、数据流建模方法。第6章，介绍了物联网追溯系统的研发和软硬件环境配置，并对本书的主要内容进行梳理和总结。



目 录

第 1 章 绪论	1
1.1 物联网概述	1
1.2 中国农业物联网产业化发展现状	2
1.3 物联网大数据	5
1.3.1 大数据国内外研究热点	11
1.3.2 国内外相关文献研究	17
1.4 主要研究内容和技术路线	21
1.5 组织结构	24
第 2 章 数据流聚类方法的研究	27
2.1 简介	27
2.2 研究现状及存在的问题	30
2.3 基于高斯混合模型的数据流聚类方法研究	34
2.3.1 数据流聚类算法框架	34
2.3.2 基于高斯混合模型的数据流处理	35
2.3.3 多级队列概要结构	37
2.3.4 时间的属性化处理	38
2.3.5 基于高斯混合模型的数据流聚类	41

2.3.6 微簇合并	42
2.4 实验及分析	44
2.5 相关研究比较	51
2.6 本章小结	52
第3章 数据流追溯方法的研究	53
3.1 简介	53
3.2 研究现状及存在的问题	56
3.3 基于不确定数据的数据流追溯方法的研究	60
3.3.1 不确定数据	60
3.3.2 数据溯源	63
3.3.3 数据流集成处理过程	66
3.3.4 基于不确定数据的数据流追溯查询	70
3.3.5 单节点出错推断	72
3.3.6 多节点出错推断	76
3.4 相关研究比较	77
3.5 本章小结	78
第4章 数据流预测方法的研究	80
4.1 简介	80
4.2 研究现状及存在问题	84
4.3 基于灰色模型的数据流预测方法	87
4.3.1 预测查询处理模型	87

4.3.2	灰色一阶预测模型	92
4.3.3	灰色二阶预测模型	96
4.4	实验及分析	100
4.4.1	实验设置及仪器	101
4.4.2	结果分析	104
4.5	本章小结	108
第 5 章	猪舍环境监控系统的设计	110
5.1	简介	110
5.2	研究现状及存在的问题	113
5.3	猪舍数据流采集监控系统的设计	117
5.3.1	系统组成建模	117
5.3.2	物理世界数据建模	122
5.3.3	传感器数据建模	122
5.3.4	无线网络数据建模	123
5.3.5	计算(控制)单元数据建模	125
5.3.6	执行器数据建模	127
5.4	模型实例	127
5.5	系统的实现与优化	129
5.6	本章小结	135
第 6 章	物联网追溯系统研发	137
6.1	养殖场信息管理系统	137

6.1.1	硬件支持	140
6.1.2	软件支持	141
6.1.3	使用说明	141
6.2	屠宰信息管理系统	152
6.2.1	硬件支持	153
6.2.2	软件支持	153
6.2.3	使用说明	154
6.3	追溯查询信息系统	167
6.4	本章小结	170
6.5	创新点	173
6.6	展望	174
	参考文献	175
	读者调查表	177
	电子工业出版社编著书籍推荐表	178
	反侵权盗版声明	180

第1章 绪论

1.1 物联网概述

物联网 (Internet of Things, IoT) 的概念最早是在 1999 年由麻省理工学院的 ASHTON 教授提出的。2003 年, SUN 公司介绍了物联网的基本工作流程, 并提出解决方案。2008 年 11 月, IBM 提出“智慧地球”的发展战略, 受到美国政府的高度重视, 并得到奥巴马的积极回应。2009 年, 物联网成为继计算机、互联网、移动通信之后新一轮信息产业浪潮的核心领域。物联网是在互联网、移动通信网等通信网络的基础上, 通过智能传感器、射频识别 (RFID)、红外感应器、全球定位系统 (GPS)、激光扫描器、遥感等信息传感设备及系统, 按照约定的协议, 针对不同应用领域的需求, 将所有能独立寻址的物理对象互连起来, 实现全面感知、可靠传输及智能处理, 构建人与物、物与物互连的智能信息服务系统。

物联网被称之为继计算机、互联网之后, 世界信息产业的第三次革命浪潮。所谓物联网, 是一个基于互联网、传统电信网和传感网等信息承载体, 让所有普通物理对象能够通过信息传感设备与互联网连接起来, 进行计算、处理和知识挖

掘,实现智能化识别、控制、管理和决策的智能化网络。物联网技术的快速发展,促进了农业物联网的迅速崛起。农业物联网是物联网技术在农业领域的应用,是通过应用各类传感器设备和感知技术,采集农业生产、农产品流通及农作物本体的相关信息,通过无线传感器网络、移动通信无线网和互联网进行信息传输,将获取的海量农业信息进行数据清洗、加工、融合、处理,最后通过智能化操作终端,实现农业产前、产中、产后的过程监控、科学决策和实时服务。农业物联网在农业生产过程中的各类应用对于推动信息化与农业现代化的融合、推动精细农业应用与实践等具有至关重要的作用。农业物联网技术应用有利于农业生产力的提高和农业生产经营模式的转型升级,是新一代信息技术渗透进入农业领域的必然结果。

1.2 中国农业物联网产业化发展现状

随着农业信息技术的推广,特别是物联网技术的广泛应用,中国作为一个传统农业大国,在物联网技术大潮下,传统的农业生产逐步实现了从“面朝黄土”到“智能化管理”的转变。目前,农业物联网技术应用已覆盖水产养殖、畜禽饲养、设施园艺、大田作物等多个产业。猪舍的温湿度是多少、氨气浓度是否超标……对这些环境信息,在当今时代,人们可以足不出户,通过计算机和手机,就可以对猪舍的信息了解得一清二楚。

在食品安全方面,可以通过销售条形码制,为农产品销售建立可追溯制度,实现从田间到餐桌的产业链精细化管理、全程透明跟踪。消费者不仅可以查询所购买

产品的生产记录、运输记录、销售记录等信息，甚至可以精确到农产品生产的地理位置，真正从源头上保证了农产品质量安全，让人们吃上放心的产品。充分利用物联网、移动互联网及云平台技术，建立数据传输和格式转换方法，实现农业信息的多尺度可靠传输；最后将获取的海量农业信息进行融合、处理，并通过智能化操作终端实现农业的自动化生产、最优化控制、智能化管理、系统化物流、电子化交易，进而实现农业集约、高产、优质、高效、生态和安全的目标。

在畜禽养殖方面，物联网的应用不仅能有效优化养殖管理模式，而且可以提高动物年产奶量和产奶质量，提高整个养殖场的产肉量，精确模拟畜禽的生长环境，为养殖产业的持续发展注入新的活力。传统的粗放式养殖正在被基于物联网的精细化饲养方式逐步取代，该技术主要用于生产环境监测、畜禽的生理监测、精细饲喂、溯源信息采集等方面。养殖人员根据畜禽不同生长阶段的生理特征，采用不同的饲料配方，降低了饲养成本，有益于畜禽的健康成长，提高了养殖场的经济效益。在养殖场的管理方面，现代化技术手段贯穿于畜牧养殖的管理和经营各个过程，信息技术应用于养殖的各个环节，网络使得人们能通过计算机、手机等科技产品，不受时间和地域的限制，随时获取养殖场的相关信息，为养殖场的科学管理提供数据和决策支持，提升了畜牧养殖管理方式。

在设施农业方面，以往只是埋头苦干，没有技术指导的农民，现在转型为有知识、有文化的现代化种植技术人员，借助物联网技术，彻底改变了过去的以人力为中心，依赖机械、看天吃饭的生产模式，转而以软件和信息技术为核心的生产模式。通过在实施对象周边安放多种传感器和监测控制设备，如光照、温湿度、二氧化碳浓度、土壤 pH 值、土壤离子浓度等传感器，来调控作物生长的环境参数，使其达到最佳，从而实现高产和增产。依靠云平台实现自动整理、分析、存储数据，远程进

大数据的采集是指利用多个数据库来接收来自客户端（Web、App）或者传感器

行作物生长环境信息的智能优化、综合调控,关键数据在监控端实时可视化显示,最终实现精确感知、精准操作、精细管理,提高生产效率。

在水产养殖业方面,通过物联网技术实现对水产品环境监测、水产品精细投喂智能决策、自动化投饲、水产品疾病诊治等全过程监控和智能化管理。通过部署溶解氧、pH、水位、水温、光照传感器和智能球等传感器设备,将实时采集的环境变量值通过网络进行传输,实现水产养殖的全过程信息化管理。用户可以登录计算机或手机,查看养殖池塘的信息,自行设置环境阈值,可实现达到或超过阈值时自动向手机发送预警短信提醒的生产报警功能;当溶解氧达到阈值下限时,可实现自动打开增氧机生产设备的智能调控功能,人工智能终端开关增氧机、投料机生产设备的远程控制功能;远程调整无线智能球的角度,通过视频图像,实时查看养殖池塘、生产设备等状况和周边环境情况,使水产品标准化的养殖环境下生长,最终达到提高水产品质量、降低养殖风险的目的。

在农业电子商务方面,物联网技术的普及推广大大推动了农业电子商务的迅猛发展。集网上商城、销售系统、仓库管理系统、运输管理系统于一体的电子商务平台可将传统的农业贸易模式转变为现代信息流模式,使农产品流转环节清晰透明,降低和节约物流成本和农产品交易成本,拓宽农产品销售渠道,增加农户的收入,促进农村经济发展。基于物联网的农业电子商务,实现信息数字化、生产自动化、管理智能化、高效、高产、绿色生态的现代农业商品体系,实现了产、学、研的高效结合。

在食品安全方面,可以通过销售系统,为农产品销售建立可追溯制度,实现从田间到餐桌的产业链精细化管理、全程透明跟踪。消费者不仅可以查询所购买

1.3 物联网大数据

随着物联网、云计算、互联网、社交媒体等技术的飞速发展，在智慧农业、精准农业、医疗、传感器、用户数据、互联网和金融公司交易等遍布全球的各个领域都出现了大规模的数据增长，我们每天都处在数据的环境中，大量数据实时地影响我们的工作、生活乃至社会发展。大数据是继高性能计算机、互联网、网格计算、云计算之后的又一被大众所关注的技术术语。全球知名咨询公司麦肯锡提出大数据时代已经到来，称：“数据已经无孔不入地渗透到各个行业和领域，成为重要的生产因素。对于海量数据的挖掘与运用，预示着新的生产率增长和消费者盈余浪潮的到来。”大数据的涌现，不仅影响人们的生活方式、工作方式、企业的运作模式，甚至还引起科学研究模式的根本性改变。据国际数据资讯（IDC）公司监测，全球数据量大约每两年翻一番，预计到2020年，全球将拥有35ZB的数据量。

大数据，是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。大数据技术，就是从各种类型的数据中快速获得有价值信息的技术。大数据领域已经涌现出了大量新技术，它们成为大数据采集、存储、处理和呈现的有力武器。大数据处理关键技术一般包括：大数据采集、大数据预处理、大数据存储及管理、大数据分析及挖掘、大数据展现和应用等。

大数据的采集是指利用多个数据库来接收发自客户端（Web、App 或者传感

器形式等)的数据,通过这些数据进行简单的查询和处理工作。在大数据的采集过程中,其主要特点和挑战是并发数高,因为同时有可能会有成千上万的用户来访问和操作,比如火车票售票网站和淘宝网,它们并发的访问量在峰值时达到上百万,所以需要在采集端部署大量负载均衡的数据库来支撑。

虽然采集端本身会有很多数据库,但是如果要对这些海量数据进行有效的分析,还是应该将这些来自前端的数据导入到一个集中的大型分布式数据库,或者分布式存储集群中,并且可以在导入的基础上做一些简单的清洗和预处理工作。也有一些用户会在导入时使用来自 Twitter 的 Storm 对数据进行流式计算,来满足部分业务的实时计算需求。导入与预处理过程的特点和挑战主要是导入的数据量大,每秒钟的导入量经常会达到百兆,甚至千兆级别。

统计与分析主要利用分布式数据库,或者分布式计算集群来对存储于其内的海量数据进行分析 and 分类汇总等,以满足大多数常见的分析和需求。在这方面,一些实时性需求会用到 EMC 的 GreenPlum、Oracle 的 Exadata,以及基于 MySQL 的列式存储 Infobright 等,而一些批处理,或者基于半结构化数据的需求可以使用 Hadoop。该操作涉及的数据量大,其对系统资源,特别是 I/O 有极大的占用。

与前面统计和分析过程不同的是,数据挖掘一般没有什么预先设定好的主题,主要是在现有数据基础上进行基于各种算法的计算,从而起到预测的效果,实现一些高级别数据分析的需求。比较典型的算法有用于聚类的 k-means、用于统计学的支持向量机和用于分类的朴素贝叶斯分类器等,主要使用的工具有 Hadoop 的 Mahout 等。

大数据需要特殊的技术,以有效地处理数据。适用于大数据的技术,包括大规模并行处理(MPP)数据库、数据挖掘、分布式文件系统、分布式数据库、云