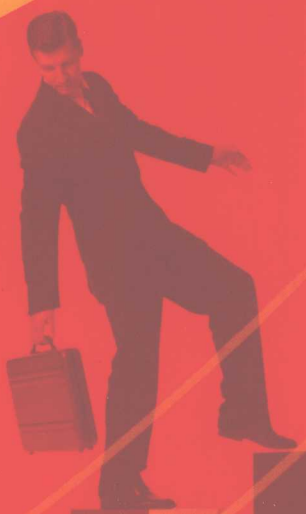
分布式连锁商业数据挖掘模型

FENBUSHI LIANSUO SHANGYE
SHUJU WAJUE MOXING

肖亮 著



航空工业出版社

分布式连锁商业数据挖掘模型

航空工业出版社
北　京

内 容 提 要

在知识成为企业核心竞争力的今天，如何从海量动态商业数据中提炼出有价值的商业知识，指导企业经营管理和科学决策，已经成为连锁商业企业持续健康发展的关键所在。

本书针对连锁商业数据的特性，采用国内外数据挖掘理论的最新热点——分布式商业数据挖掘技术，以连锁商业企业为主要研究和应用对象，全面、系统、深入的探讨了分布式商业数据挖掘理论和技术应用。通过实证研究，验证了这一理论的科学性和实践性。

图书在版编目（CIP）数据

分布式连锁商业数据挖掘模型/肖亮著. —北京：航空工业出版社，2008.12

ISBN 978-7-80243-227-7

Ⅰ.分… Ⅱ.肖… Ⅲ.连锁商店—数据管理 Ⅳ.F717.6

中国版本图书馆 CIP 数据核字（2008）第 185849 号

分布式连锁商业数据挖掘模型

Fenbushi Liansuo Shangye Shuju Wajue Moxing

航空工业出版社出版发行

（北京市安定门外小关东里 14 号 100029）

发行部电话：010-64815615 010-64978486

北京华正印刷有限公司印刷 全国各地新华书店经售

2008 年 12 月第 1 版 2008 年 12 月第 1 次印刷

开本：880×1230 1/32 印张：12.5 字数：320 千字

印数：1—3000 定价：28.00 元

前　言

随着知识经济时代的来临，信息与知识已经成为国家和企业发展的重要战略资源，是提高一个组织乃至一个国家战略竞争力的核心，也是实施科学管理与决策的基础。如何获取信息与发现知识，尤其是如何快速高效地在动态变化和爆炸性增长的分布型海量数据中获取信息和发现知识就成了关键性问题。

作为国民经济和社会发展的重要组成部分，商贸流通业已经在过去的发展中形成了大量的商业数据库，并且其数据流量仍在继续保持增长。尤其是商业领域中大型连锁商业企业（如沃尔玛 Wal－mart、家乐福 Carrefour、麦得龙 Metro、上海百联、解百集团等）通过网络实现上千家连锁分店、配送中心与总店间内部信息的互联，以及与众多供应商、银行等外部信息的交换，形成了分布式的商业数据共享环境。这些分布的数据库以每天几百兆甚至更多的数据记录量的速度增加和流动，形成了动态变化的海量商业数据库。

显然，在知识成为企业核心竞争力的今天，如何从海量动态商业数据中提炼出有价值的商业知识，指导企业经营管理和科学决策，已经成为连锁商业企业持续健康发展的关键所在。因此，针对连锁商业数据的特性，开展网络环境下动态商业数据流的协同知识发现问题的研究，实现复杂环境下的分布式异构动态商业数据的挖掘处理，从而发掘商业企业运行的特征，找出具有共性或规律性的知识，为商业企业提供科学和有效的管理与决策支

持，具有十分重要的理论价值和现实意义。

本书力图追踪国内外数据挖掘理论的最新热点——分布式商业数据挖掘技术，其研究视野居于国内同领域理论研究前沿，研究观点比较全面、系统、深入，具有较强的前瞻性和创新性。全书在深入分析当前连锁商业企业经营特点的基础上，以连锁商业企业为主要研究和应用对象，系统介绍了分布式商业数据挖掘理论和技术应用，并重点研究了以下内容。

第一，分布式连锁商业数据源管理研究。在分析连锁商业企业组织网络分布特点和数据管理组织行为的基础上，对分布式连锁商业企业数据源的概念、特点、形式化表达和组织体系进行了系统考察，提出了基于元数据、对象数据、过程数据和知识数据的分布式连锁商业数据管理概念模型，以及分布式商业数据运算关系及其规则表达，并探讨了支持分布式连锁商业数据挖掘的逻辑管理模型、协同管理模型，以及基于 Mobile - Agent 的连锁商业协同挖掘模型实现。

第二，分布式连锁商业数据挖掘系统的体系架构实现。通过对国内外研究成果的总结和研究，面向连锁商业企业的分布式数据挖掘需求可以归纳为支持连锁商业企业全局统筹决策和局部自主决策的两类数据挖掘需求。在此基础上，通过对商业企业经营决策的四维（供应商、财务、消费者、商品）主题的定性和定量分析，提出面向商业企业经营决策的“以客户为中心、商品驱动、供需联动”的商业企业经营决策的系统框架模型——B - DSM，并研究了该模型的框架、工作原理及功能模块的组成、协调与集成等。最后，探讨了基于 SMAS 体系的 B - DSM 系统实现技术、支持 SMAS 系统实现的数据交换关键技术，并以大型零售企业电子商务平台为例进行了实证分析。

第三，分布式连锁商业数据挖掘算法研究。在分析了面向商业经营决策的各种数据挖掘算法的基础上，系统地阐述了分布式

商业数据源分布与管理、分布式商业数据挖掘决策系统架构，以及基于分布式ID3的连锁商业企业挖掘算法（B－DID3）、基于分布式BNS的连锁商业企业挖掘算法、基于分布式关联规则的连锁商业企业数据挖掘算法（B－DAR）和基于C4.5和地域因素的连锁商业企业挖掘算法（B－ZDT）等分布式挖掘算法。

第四，基于大型连锁商业企业商业数据挖掘的实证研究。设计实现了面向杭州某大型连锁商业企业的HZ－DM系统原型和架构以及逻辑层和用户界面层。根据该百货集团客户关系管理领域实际应用需求的分析，将现有比较成熟的统计分析方法应用于该集团的客户数据分析中，并充分考虑了系统对现有成熟分析工具的集成，包括DMiner、Weka等统计分析与数据挖掘工具，在此基础上，将基于C4.5和地域因素的连锁商业企业挖掘算法（B－ZDT）、基于BNs的连锁商业企业挖掘算法（B－BNs）应用到了其客户关系管理系统中，并对应用结果进行了分析。

第五，集成情境的连锁商业数据挖掘系统研究。通过对国内外学者现有研究成果的总结，发现情境对于实现连锁商业数据挖掘算法或模型重用的重要价值，清晰地界定了连锁商业数据挖掘知识情境的内涵和主要内容；在此基础上，进一步提出了集成情境的连锁商业数据挖掘知识产品概念模型，将情境因素以合理的方式嵌入到各类连锁商业数据挖掘知识的概念表述中，并提出了基于情境匹配的连锁商业数据挖掘知识产品的快速发现和重用机制，以及基于Mobile－Agent的分布式连锁商业数据挖掘知识管理系统。最后，以某百货大楼某快速消费品数据为样本，对集成情境的产品生命周期管理模型进行了验证，从而进一步论证了在连锁商业企业各类经营决策中集成情境的必要性。

第六，数据流时代的连锁商业数据挖掘研究展望。随着数据流正在成为连锁商业企业的主流数据形式，连锁商业数据挖掘的重点由传统静态数据逐步转向动态商业数据流。本书在全面回顾

数据流现有研究成果的基础上，对连锁商业数据流挖掘的研究现状进行了总结，并对未来连锁商业数据流挖掘研究工作进行了展望。

本书是浙江工商大学管理科学工程研究所多年研究成果积累的综合反映，受国家自然科学基金（70671094）、国家社科基金（05BTJ019）21世纪人才资助计划（2005），高等学校博士学科点专项科研基金（教技发中心函2005－216）、浙江省哲学社会科学基金（07CGGL015YBQ）、浙江省科技计划（2008C23002）等项目的资助；也是以琚春华教授为带头人包括刘东升、王冰、王蓓、周怡、张捷等人在内的团队多年研究成果，在此表示衷心感谢。同时，感谢浙江工商大学商贸研究中心对本书出版的大力支持。

本书写作过程中，参考了不少资料，作者已尽可能详细地在参考文献中列出，在此对这些专家学者们表示深深的谢意。同时对于本书虽然引用但是由于疏忽而没有指出资料出处的情况，表示诚挚的歉意。

由于作者水平有限，对分布式连锁商业数据挖掘系统的认识和研究不够深入，本书论述难免出现谬误。在此，作者真心希望同行、读者提出意见。

肖　亮

2008年9月于浙江工商大学

目　录

第1章 绪 论

1.1 背景意义

1.1.1 连锁商业企业发展现状与趋势

连锁经营是当前商业企业发展的一大趋势，已经成为商业企业的主流发展模式。据调查，迄今我国共有各类连锁商业企业2300多家，连锁业态超过40个，连锁店铺38000多间，全年销售额过3000亿元，占全国社会消费品零售总额的8%左右。随着我国零售业的全面开放，国际零售业巨头以不同的业态形式展开了对中国市场的全面争夺。为应对国际零售业巨头的竞争压力，近年来国内一些知名的大型零售企业已经通过并购、重组迅速实现了连锁化，其门点数量和经营规模持续扩大，见表1-1。典型的如国内最大的连锁商业企业联华超市，截至2006年底，总门店数目已经达到3716家，遍布全国20个省（区）及直辖市，建有2个常温配送中心和1个生鲜配送中心。北京最大的连锁企业物美在近三年，先后兼并了超市发和美廉美，截至2007年6月30日，物美共有大型超市86家，便利超市418家。可以预见，在未来相当长一段时间内，国内外连锁商业企业将继续保持强劲的增长势头，主导商业领域发展的格局将进一步深化。

表 1－1　国内部分连锁企业门店数量扩张情况　　家

连锁企业名称＼年份	2002	2003	2004	2005
百联集团有限公司	3175	4390	5493	6345
百胜（中国）餐饮投资	902	1100	1400	1757
农工商超市有限公司	702	1207	1232	1572
上海可的便利店有限公司	706	946	1079	1182
深圳海王星辰医药	320	460	668	1115
苏果超市有限公司	940	1159	1345	1503
华润万家有限公司	397	467	476	610
青岛维客集团公司		712	715	726
内蒙古小肥羊餐饮连锁		657	703	716
物美控股集团有限公司	355	518	608	656
北京国美电器有限公司	64	139	227	426
苏宁电器集团有限公司	134	148	193	363

1.1.2　连锁商业企业信息化现状与趋势

连锁商业企业的快速发展产生了对广泛覆盖、高效传输的信息系统的巨大潜在需求。随着企业分店铺越来越多，覆盖范围越来越大，必然出现管理链条的拉长和管理规模的扩大，对全地区乃至全国范围内各分店之间信息的高效传递、分析、处理能力提出了很高的要求。显然，对一个多门店、大规模、跨地区的连锁企业来说，缺乏一套有效的管理系统就不仅意味着无法实现具有统一规范的、大量数据信息的查询决策分析，也就是在经营管理的决策过程中只能盲目判断，而且也无法实现规范化的管理和有效的控制。很多连锁商业企业的经验表明，要实现连锁化经营与一体化管理的统一，依托的就是信息技术，在很多大型连锁商业集团，已广泛采用了商品条码技术、电子数据交换（EDI）技

术、EOS 技术、电子资金转账系统（EFT）技术等信息技术。信息技术的日益广泛应用使单店的成功商业模式得以大量复制，并为商业企业的连锁化、规模化经营奠定了强大的基础。

国外发达国家连锁经营的发展实践也表明，实现连锁经营持续健康发展的关键之一在于条形码技术、销售时点系统、电子转账作业系统、电子订货系统等先进计算机与网络技术的发展。与国内连锁商业企业总投资占零售总额的比例还不到 0.2% 相比，国际零售巨头这一比例一般要到 2% 以上[2]。国际零售巨头的 IT 投资一方面用于门店信息系统的升级换代，另一方面是与国际知名的数据仓库、商业信息系统供应商（惠普、思科等）合作，引入新的智能技术、数据挖掘技术、决策方法来提高信息系统的智能决策支持能力和效率。如美国西尔斯·罗巴克百货公司投入巨资建立起有数百台小型计算机和 5 万多台销售时点系统全日制工作的计算机控制系统，并引进最先进的多媒体技术、卫星通信网络，充分利用电脑使总部与各地的连锁分店以及供应商传递各种信息，做到其所属的连锁分店都处于实时控制管理之下。这一趋势在著名连锁企业沃尔玛的成长过程中表现得更为突出。沃尔玛不仅可以利用其庞大数据仓库中的销售数据、库存数据、配送数据、退货数据等来分析控制各门店的 CPFR 联合计划预测补货系统，也可以将 POS 系统数据按不同地域、不同市场、不同季节进行分析得出不同类型商品的销售趋势等。同时，通过网络信息共享系统——零售链（Retail Link），沃尔玛还能够实现与供应商之间互利的信息共享，供应商可以使用用户名和密码通过 Internet 连接登录零售链，对自己商品的销售情况进行追踪、分析。

目前，我国已经有 70% 以上的连锁企业建立了系统开发的前台 POS 销售系统和后台 MIS/ERP 管理系统，30% 左右的企业率先进入了商业自动化技术、现代通信技术和网络信息化技术相

结合的数字化管理系统集成的阶段。信息技术已经成为国内大型连锁企业实现连锁效应和效益的重要途径，这主要反映在以下几点。一是增加商品销售规模每年达20%以上，也就是600亿元以上；二是减少采购、配送、通信、理货的人工直接费用达40%；三是提高管理绩效、减少库存积压、提高商品资金周转率节约的间接费用达50%。按这个发展水平计算，信息化对我国连锁商业企业的直接收益贡献率达到40%以上，企业因采用信息技术而节约成本、增加销售而产生的直接利润就是每年30亿元以上。典型的如联华超市股份有限公司通过与IBM公司合作，建立了包括电子订单处理、网上对账及结算、数据分析决策等在内的先进供应链管理系统，为联华超市进一步向超级市场和便利店的业态延伸提供强大的支持。同时，联华超市还建成了国内首家大型智能化配送中心，先进的计算机信息系统已经覆盖了联华所有的门店、配送中心和数千家供应商，实现了商业管理的信息化、自动化和现代化。物美集团的WINBOX@SAP项目进入系统实现阶段，完成了对采购、物流、门店、财务等业务部门的SAP系统配置，与专业零售POS软件商以色列Retalix公司合作，在门店实施新的POS软件，以简化店铺营运，增加更有力的促销方式，加快统一会员管理。

1.1.3 连锁商业企业经营决策需求分析

随着连锁企业逐步摆脱价格战的困扰，未来连锁企业间的竞争将表现为：集约式的价值竞争取代粗放式的价格竞争，流通领域企业间单独依靠价格、拼数量或始终如一地以打江山时的创业品牌打“持久战”、“吃老本”、“拼优惠条件”等的竞争将被视市场需求变化、不断展开营销创新、品牌创新以及开发不同获利定位的高价值、附加价值的商品或优良投资环境等集约式竞争所取代；开放式竞争将进一步取代封闭式竞争，流通领域提供商品