

网络信息计量学的新进展

——评《网络数据分析》

苏 新 宁

(南京大学信息管理学系, 南京, 210093)

武汉大学邱均平教授与黄晓斌、段宇锋、陈敬全等博士撰著的《网络数据分析》一书, 已于 2004 年 6 月由北京大学出版社正式出版。这是我国信息计量学乃至整个信息管理领域的一项重要成果。邱均平教授在《网络数据分析》的前言中指出, “网络数据分析是网络信息计量学的不可缺少的重要组成部分”, 是信息计量在网络环境下和联机分析处理、数据仓库、数据挖掘等新技术条件下的必然延伸。《网络数据分析》则自始至终地体现了“计量分析”这一关键的思想脉络。

《网络数据分析》一书是国家自然科学基金资助项目“网络信息计量学的理论、方法与实证研究”(批准号: 70273032) 及教育部十五规划项目“网络信息计量学研究”的重要研究成果之一。全书共分 9 章, 层次分明、井然有序。第 1 章为网络数据概论; 第 4 章为定性分析; 第 3 章为流量分析; 第 2、5、6、7 各章分别论述数据仓库(DW)、联机分析处理(OLAP)、数据挖掘(DM)等前沿理论与技术工具; 最后的 8、9 两章为数据分析系统的开发及一些典型应用实例。

经认真拜读, 笔者认为《网络数据分析》至少具有以下几个突出的特点。

1 内容新颖

(1) 选题新。网络信息计量学是才诞生不久的新生事物, 而网络数据分析作为网络信息计量学的一个重要组成部分, 更是迫切需要进行研究的热门课题。日益丰富的互联网信息资源与日益求精的用户需求, 向网络数据分析提出了发展的迫切需求。邱均平教授等人敏锐地捕捉到了这个发展契机, 将计量分析的思想引入了对网络信息的考察, 使得研究既具有严谨的科学逻辑, 又紧扣时代脉搏, 符合时代的特色与发展要求。

(2) 观点新。著作中时时可见富有独创性的观点与方法。例如在前言中, 邱均平教授对网络数据

分析与网络信息计量学的关系的论述, 以及对网络信息计量学的概念、对象、内容体系和产生背景等的论述都是首创性的, 并得到了学术界的广泛认同和引用。在第一章“网络数据概论”中, 以一小节专门论述了“网络环境下的数字信息表达”问题, 将其定义为“利用数字化技术将文字、声音、图像等信息转换成以 0、1 表示的计算机可识别的代码, 以实现利用网络对信息进行获取、存储、处理和传递”; 认为“实现原始信息的数字信息表达主要依靠数字化信息生成技术、数字信号处理技术、音/视频处理技术、数字压缩技术、数据库技术、数字存储技术和数字通信技术”。在“网络数据的类型”一节中, 根据实际应用中的不同特点, 将网络数据划分为元数据、结构化与半结构化数据、非结构化数据、流媒体数据; 将元数据与流媒体数据从传统的“三分法”中单列出来。第 4 章在以定量分析为主的总体框架下, 又大胆拓新, 将定性分析引入网络数据分析的内容体系, 详细论述了网络数据的内容分析、抽象分析、归纳分析、相关分析、对比分析等方法。以上这些观点的归纳, 都是极有见地和创新意义的。

(3) 资料新。该书在量化分析思想的基础上, 又大胆将一些最新技术手段如 OLAP、DW、DM 等引入网络数据分析的研究范畴, 进一步拓宽了研究视野。不仅如此, 在对这些部分进行论述时, 都十分注意紧跟理论与技术前沿, 反映和吸收国内外最新研究成果。在第 2 章“基于网络的数据仓库”一节中, 介绍了 MLDB 多层数据库的概念。第 5 章“网络数据的多维分析”介绍了联机分析处理 OLAP 的新发展——联机数据挖掘 OLAM。第 7 章“网络复杂数据的挖掘分析”中涵盖了多媒体数据挖掘、时序数据挖掘、空间数据挖掘等尖端内容。第 8 章“网络数据分析系统的开发”则有一节专门论及“数据挖掘分析语言”, 反映出该领域的关键发展趋势。另外, 该书中还给出了相当多的典型应用实例, 如网络数据分析在网络行为学中的应用, 在股票分析中的应用, 在

电子商务与网络营销中的应用等,内容充实而活泼。

2 结构合理

整部著作严格遵循着“理论—方法—应用”的逻辑思路来展开论述,脉络异常清晰。

(1)整体结构分明。该书的第1章“网络数据概论”、第2章“网络数据仓库”是作为基本的理论准备。主体部分第3章“网络数据流量分析”、第4章“网络数据的定性分析”、第5章“网络数据的多维分析”、第6章“网络数据的挖掘分析”、第7章“网络复杂数据的挖掘分析”论述了若干具体的数据分析方法与流程,以及一些可使用的工具。最后第8章介绍“网络数据分析系统的开发”,第9章为“网络数据分析的应用与实例”。这样,“理论—方法—应用”的整体结构一目了然。

(2)内部条理清楚。在书中每一章的内部,也大多遵循着同样的思维方式:先阐明概念与原理,再论述方法与流程,最后介绍产品和应用。举例来说,第2章“网络数据仓库”在内容安排上,第1节“数据仓库概述”含数据仓库的定义、特征、类型、发展、结构、意义等内容,是基于“理论”的层面;第2节“基于网络的数据仓库”主要讲述了网络数据仓库的组建、实现技术、体系结构、问题和局限等,是基于“方法”的层面;第3、4节介绍了数据仓库的典型产品,如 Decision Base、DB2 UDB、SQL Server 等,以及如何评价与选择数据仓库系统,很明显是属于“应用”层面的问题。类似的逻辑思路也反映在第3、4、5、6、7、8各章的内容结构中。

可以认为,《网络数据分析》是由“道”而“法”、而“术”、而“器”,溯源析流,纲举目张。通过对内容的精心取舍与合理组织,做到了既概览全局,又突出重点;既全面、系统,又深入、透彻。

3 实用性强

面向应用、便于操作是《网络数据分析》的重要特点。通过方法的讲解、工具的介绍以及案例的分析,大大强化了这部学术专著的实用性。

(1)方法实用。该著作对于各种网络数据分析的基本原理与方法、流程的论述较为透彻,并不乏著者自己的心得体会。以网络数据的定性分析为例,该书对内容分析、抽象分析、归纳分析、相关分析、对比分析等各种定性分析的方法逐一进行了详细的剖

析。对于内容分析法,又从方法的产生、概念、特点、方法、步骤、类型、评价及在网络数据分析中的应用与局限等方面作了全方位的论述。读者完全可以按照这些方法的指引进行操作。

(2)工具实用。如在第2、5两章中,分别介绍了 Oracle、Microsoft SQL Server、IBM DB2、Informix、Decision Base、NCR、SAS、Hyperion Essbase 等多种国内外主流的数据仓库与联机分析处理产品。另外还对如何选择数据仓库产品的评价标准进行了有益的探讨。又如第3章“网络数据的流量分析”中也列举了许多常用的商用网站分析软件及下载网址、提供第三方流量认证的公司及其网址等,这些信息对于读者来说都是十分实用的。

(3)案例实用。该著作的最后一章及各章内部都给出了许多示范性的案例。读者一来可以通过这些案例树立信心和得到启发,二来可立即查找线索和模仿应用,便于实际操作。

4 适应面广

网络数据分析与网络信息计量、信息分析与预测、知识管理、竞争情报、企业经营决策与市场战略等均有着密切的联系,因而具有广泛的应用前景。仅在该著作中就列举了多个方面的应用实例:①在网络信息资源管理中的应用;②在网络行为学研究中的应用;③在线股票分析中的应用;④在企业经营及市场分析与预测中的应用;⑤在电子商务与网络营销中的应用;⑥在网络传播中的应用;⑦在网络信息计量学中的应用;⑧在科研评价中的应用。

总而言之,《网络数据分析》是一部佳作、力作。它从理论、方法、应用多个角度,全方位地考察网络数据分析的问题,注重理论与实践相结合。全书思路清晰,结构合理,内容丰富,观点新颖,资料翔实;既反映和吸收国内外的最新进展,又融入了著者自己的系统研究成果,具有较强的创新性、科学性、系统性和实用性;既适合于高等院校的新闻与传播、信息管理与科学评价、信息计量学与科学计量学、网络设计与管理、管理科学与工程、电子商务、网络营销等专业的师生教学使用,也可供广大信息工作者、网络工作者、科技工作者和有关管理人员学习参考。

(收稿日期:2004-10-20)