

企业信息系统数据复杂性的 度量及应用

RESEARCH ON THE DATA
COMPLEXITY AND APPLICATION OF
ENTERPRISE INFORMATION SYSTEM

尹 隽 著

“数据复杂性”概念及理论模型

数据复杂性度量方法

模型与方法的有效性

数据复杂性对信息系统维护的影响

适应性维护策略



企业信息系统数据复杂性的 度量及应用

尹 隽 著



江苏大学出版社
JIANGSU UNIVERSITY PRESS

镇 江

图书在版编目(CIP)数据

企业信息系统数据复杂性的度量及应用 / 尹隽著

. — 镇江 : 江苏大学出版社, 2018. 12

ISBN 978-7-5684-1042-7

I. ①企… II. ①尹… III. ①企业管理—管理信息系统—研究 IV. ①F270.7

中国版本图书馆 CIP 数据核字(2018)第 291611 号

企业信息系统数据复杂性的度量及应用

Qiye Xinxi Xitong Shuju Fuzaxing De Duliang Ji Yingyong

著 者/尹 隽

责任编辑/仲 蕙

出版发行/江苏大学出版社

地 址/江苏省镇江市梦溪园巷 30 号(邮编:212003)

电 话/0511-84446464(传真)

网 址/http://press.ujs.edu.cn

排 版/镇江文苑制版印刷有限责任公司

印 刷/虎彩印艺股份有限公司

开 本/890 mm×1 240 mm 1/32

印 张/8.125

字 数/219 千字

版 次/2018 年 12 月第 1 版 2018 年 12 月第 1 次印刷

书 号/ISBN 978-7-5684-1042-7

定 价/48.00 元

如有印装质量问题请与本社营销部联系(电话:0511-84440882)

**RESEARCH ON THE DATA
COMPLEXITY AND APPLICATION OF
ENTERPRISE INFORMATION SYSTEM**

本书由以下立项项目资助出版

国家自然科学基金面上项目“基于数据复杂性的大中型制造企业信息系统运维服务定价机制研究”(71271104)

国家自然科学基金重点项目“基于云的管理信息系统再造研究”(71331003)

国家自然科学基金面上项目“多层复杂网络视角下开源软件社区集体智慧涌现机制研究”(71871108)

江苏省普通高校研究生科研计划项目“数据视角的信息系统复杂性及其应用研究”(KYZZ16_0202)

江苏高校品牌专业建设工程一期项目“江苏科技大学信息管理与信息系统专业”(PPZY2015B166)

江苏高校“青蓝工程”优秀教学团队项目“信息管理与信息系统专业教学团队”

江苏省高等教育教改研究课题“面向船舶制造业的信管专业创新性应用人才培养模式探索”(2017JSJG118)

前 言

准确测量信息系统的复杂性是解决系统成本估算、合理报价及维护优化的基础,现有的研究成果或将信息系统视为软件技术系统,从软件的代码、结构或对象等方面探索系统的复杂性;或拓展到信息管理系统视角,关注信息系统设计、开发阶段的复杂程度,但就当前众多研究成果而言,鲜有学者关注信息系统运行阶段的复杂性。为弥补当前的研究缺陷,本书基于数据视角,提出构建能够全面衡量企业信息系统设计、开发和运行阶段复杂性的模型和度量方法,并进行应用探索,以期为信息系统复杂性的理论研究和实践管理提供借鉴和参考。

本书共七章,第1~2章提出研究问题和研究意义,并概述本书的相关理论基础;第3章基于表征理论,从信息系统的内涵出发,提出“数据复杂性”的概念,构建相应的理论模型;第4~5章基于上一章提出的模型开发复杂性度量方法及其对应的度量工具,并将所提出的模型和度量方法与传统的模型和方法进行对比验证,论证模型和方法的有效性;第6~7章是对上述数据复杂性度量的应用,通过对7家制造企业162个信息系统实际数据的实证分析,验证了企业信息系统数据复杂性对维护成本的影响,接着设计了基于动态数据复杂性的适应性维护策略;第8章在总结工作的同时提出了未来研究工作中值得关注的问题。

本书的主要内容系国家自然科学基金项目“基于数据复杂性的大中型制造企业信息系统运维服务定价机制研究”(项目编号:71271104)、“基于云的管理信息系统再造研究”(项目编号:71331003)、“多层复杂网络视角下开源软件社区集体智慧涌现机

制研究”(项目编号:71871108),以及江苏省普通高校研究生科研计划项目“数据视角的信息系统复杂性及其应用研究”(项目编号:KYZZ16_0202)、江苏高校品牌专业建设工程一期项目“江苏科技大学信息管理与信息系统专业”(项目编号:PPZY2015B166)、江苏高校“青蓝工程”优秀教学团队项目“信息管理与信息系统专业教学团队”、江苏省高等教育教改研究课题“面向船舶制造业的信息专业创新性应用人才培养模式探索”(项目编号:2017JSJG118)联合资助的研究成果,上述一些研究成果已发表在《管理工程学报》《系统工程理论与实践》《管理评论》《管理学报》等国内一流学术期刊上。

尽管笔者非常期望奉献给读者一部满意的书稿,但鉴于水平有限,有些观点、理论及方法尚不成熟,书中难免存在疏漏或不足之处,恳请广大学者、同仁多提宝贵意见,以便今后进一步深入研究。在此,对书中引用的研究成果和文献作者表示真诚的感谢和崇高的敬意,对支持本书出版的同事、朋友及出版社的编辑们表示诚挚的谢意。

目 录

第1章 绪 论 001

- 1.1 研究背景及研究意义 001
 - 1.1.1 研究背景 001
 - 1.1.2 研究问题 010
 - 1.1.3 研究意义 012
- 1.2 文献综述 012
 - 1.2.1 文献检索和分析 012
 - 1.2.2 信息系统的复杂性研究框架 016
 - 1.2.3 软件技术视角研究 019
 - 1.2.4 信息系统管理视角研究 032
 - 1.2.5 现有研究评述 040
- 1.3 本书结构 042
 - 1.3.1 研究内容 042
 - 1.3.2 技术路线和研究方法 044
- 1.4 研究创新点 046

第2章 理论基础 048

- 2.1 信息系统及信息系统数据 048
 - 2.1.1 信息系统 048
 - 2.1.2 信息系统数据 052
- 2.2 复杂性 054
 - 2.2.1 复杂性的定义和内涵 054
 - 2.2.2 复杂性理论、复杂系统与复杂性 056

- 2.2.3 复杂性的度量 058
- 2.3 信息系统的复杂性 063
- 2.4 信息系统复杂性的度量研究 066
 - 2.4.1 软件设计阶段的软件复杂性度量方法 066
 - 2.4.2 软件开发阶段的软件复杂性度量方法 073
 - 2.4.3 信息系统开发阶段的信息系统复杂性度量方法 077
- 2.5 信息系统复杂性的实证研究 079
 - 2.5.1 Banker 等的复杂性研究模型 079
 - 2.5.2 Bowen 等的复杂性研究模型 080
- 2.6 本章小结 081

第3章 数据复杂性的理论模型构建 083

- 3.1 信息系统的表征理论 083
- 3.2 数据复杂性 086
 - 3.2.1 信息系统的内涵表征——数据视图 086
 - 3.2.2 数据复杂性的提出 089
 - 3.2.3 数据复杂性的定义 092
- 3.3 数据复杂性的理论模型构建 093
 - 3.3.1 框架 093
 - 3.3.2 框架应用 094
 - 3.3.3 模型构建 095
 - 3.3.4 模型概述及边界条件 105
- 3.4 数据复杂性理论模型的理论贡献 107
 - 3.4.1 基于表征理论和数据的新视角 107
 - 3.4.2 理论的拓展 109
- 3.5 本章小结 110

第 4 章 企业信息系统数据复杂性的度量 111

- 4.1 数据复杂性的度量模型 112
 - 4.1.1 数据结构复杂性的度量模型 113
 - 4.1.2 数据操作范围复杂性的度量模型 117
 - 4.1.3 数据操作强度复杂性的度量模型 118
 - 4.1.4 数据量复杂性的度量模型 120
 - 4.1.5 数据复杂性的综合度量模型 121
- 4.2 数据复杂性的度量方法 124
 - 4.2.1 数据复杂性的信息熵计算方法 124
 - 4.2.2 数据复杂性的格空间计算方法 129
- 4.3 数据复杂性度量工具的设计与开发 132
 - 4.3.1 现有信息系统复杂性度量工具 132
 - 4.3.2 数据复杂性度量工具的设计 134
 - 4.3.3 数据复杂性度量工具的开发 141
- 4.4 数据复杂性度量的算例分析 143
 - 4.4.1 X 公司物资库存管理和成本管理系统数据复杂性度量 143
 - 4.4.2 A 市中小企业公共服务平台系统数据复杂性度量 150
- 4.5 本章小结 154

第 5 章 企业信息系统数据复杂性度量方法的有效性验证 156

- 5.1 信息系统的体系结构复杂性与数据复杂性比较的可行性分析 157
- 5.2 案例研究 158
 - 5.2.1 X,Y,Z 公司物资管理信息系统对比分析 158
 - 5.2.2 X 公司物资和成本管理信息系统对比分析 163
- 5.3 讨 论 168
- 5.4 本章小结 169

第6章 企业信息系统数据复杂性对维护成本的影响 170

- 6.1 问题提出 171
- 6.2 理论基础和研究假设 172
 - 6.2.1 理论基础 172
 - 6.2.2 研究假设 174
- 6.3 实证研究设计 177
 - 6.3.1 数据来源与样本描述 177
 - 6.3.2 变量定义和测量 178
 - 6.3.3 模型构建 180
- 6.4 实证研究与分析 181
 - 6.4.1 样本特征及描述性统计 181
 - 6.4.2 假设检验与结果分析 182
- 6.5 结论与启示 185
- 6.6 本章小结 188

第7章 基于动态数据复杂性的企业信息系统维护策略设计 189

- 7.1 问题提出 190
- 7.2 基于企业信息系统动态数据复杂性的维护策略设计 191
- 7.3 案例分析 193
 - 7.3.1 样本企业选择 193
 - 7.3.2 样本数据获取 194
 - 7.3.3 样本企业信息系统动态数据复杂性应用特征及规律分析 195
 - 7.3.4 样本企业信息系统维护策略设计 200
- 7.4 本章小结 201

第 8 章 结论与展望 203

8.1 研究结论 203

8.2 理论意义与实践价值 206

8.3 研究局限 208

8.4 进一步的研究方向 209

参考文献 211

第 1 章 绪 论

本章首先介绍本书的研究背景,结合信息系统复杂性的研究脉络,从而引申出研究问题和研究意义,然后进行系统的研究综述,分析已有研究中已经解决和尚未解决的问题,梳理本书的研究与已有研究之间继承和发展的关系,在此基础上,明确本书的研究内容、技术路线和研究方法,并指出可能的创新点。

1.1 研究背景及研究意义

1.1.1 研究背景

(1) 现实背景

在复杂多变的市场环境下,越来越多的企业依赖信息系统满足其提高运营效率、创造竞争优势、创新企业的需求^[1]。Gartner 报告指出^[2],2016 年全球的企业信息系统总投资高达 3 330 亿美元,较 2015 年增长 5.9%;2017 年预计全球企业信息系统的总投资将达到 3 550 亿美元,较 2016 年增长 6.8%。从上述数据可以看出,目前企业信息系统的投资越来越大,信息系统为众多企业创造了前所未有的机遇和价值^[3, 4],如沃尔玛、亚马逊、谷歌、汉莎航空^[5]、星展银行^[6]、阿迪达斯^[7]、阿里巴巴等企业都是应用信息系统获得竞争优势的典型案例。但与此同时,随着信息技术的发展,社会需求日趋多样化,信息系统的复杂性也越来越高^[8-10]。日益复杂的信息系统给企业带来了诸多困扰,如系统资源投入不断上升、部分系统应用效果低下甚至失效、系统维护价格昂贵却

难以满足维护要求^[11-13]等。因此,理解和把握信息系统的复杂性,并以此为依据提出优化信息系统管理的机制和方法,包括制定合理的系统维护价格、优化系统管理过程等,具有重要的理论意义和实践价值。

(2) 理论背景

按照 Simon 对系统复杂性的定义,当系统“由多个部分组成,且通过非简单的方式交互,则认为这个系统是复杂的”^[14]。而一个信息系统通常由多个模块组成,需支持和协调不同类型的业务功能^[3],并由用户通过与系统功能的交互来完成相应的业务^[15],这种交互过程还需要适应快速变化的内外环境^[9],因此,信息系统被认为是典型的复杂系统^[16-18]。

多项研究表明,复杂性已成为信息系统最突出的特点^[19],并成为贯穿整个信息系统生命周期的重要因素。对信息系统复杂性的理解是当前实践和理论界的共同话题,就实践而言,组织或个人需要面对和处理信息系统的复杂性^[20];就理论而言,当前的研究重点主要关注了两个方面:“如何运用复杂性科学来解释信息系统(Using Complexity Science in Information System)”和“如何刻画分析信息系统的复杂性特征(Analysis Complexity in Information System)”。第一个方面的研究又被称为信息系统自身的复杂性研究,旨在探索如何采用复杂性科学的理论和分析方法来解释信息系统这一复杂系统的相关问题,如 Benbya 等^[21]研究如何通过复杂性理论来改善信息系统与业务的匹配,Shojaati 等^[22]尝试通过复杂性理论分析信息系统软件开发过程,结合复杂适应系统理论分析集体层面的信息系统使用,Chae^[23]将 IT 使能的服务视为复杂适应系统,通过构建理论框架分析 IT 使能的服务创新过程等。第二个方面的研究又被称为信息系统的复杂性研究,主要来源于实践的驱动,旨在刻画和分析信息系统这个复杂系统的复杂性特征,探索信息系统的复杂性程度。本书的研究聚焦于第二个方面,即如何认识信息系统的复杂性特征,以及

如何应对这些复杂性。

“刻画分析信息系统的复杂性特征(Analysis Complexity in Information System)”的相关研究在过去的40余年获得了深入而有意义的进展,那“什么是信息系统复杂性?能不能刻画这些复杂性?如何应对和管理其复杂性?”若要回答这些问题,就需要从“信息系统复杂性的本质是什么?信息系统复杂性包含哪些?信息系统复杂性能否定量测度?信息系统的复杂性会造成怎样的影响?如何把握和驾驭这些影响?”等问题出发。学者们从“技术”层面探索了如何刻画信息系统的复杂性,特别是从软件复杂性度量方面探索。据学者 Zuse^[24]统计,从1976年到1990年的15年时间,学者们共提出了90多种不同的复杂性度量方法,并基于前期丰富的复杂性度量研究,开展了一系列有意义的实证探索;进一步地,相关研究从软件技术层面拓展到了考虑信息系统组织因素的管理层面,尤其是 Xia 和 Lee^[25]提出的信息系统开发项目(Information System Develop Project, ISDP)复杂性研究框架,为基于项目的信息系统复杂性研究奠定了理论基础。

具体的研究伴随着信息技术和管理需求的演进,可分为5个阶段,即面向软件代码的复杂性度量研究阶段、面向软件结构的复杂性度量研究阶段、多维复杂性探索及面向功能体系的复杂性分析研究阶段、面向软件对象复杂性的分析及实证研究阶段和信息系统管理视角的复杂性探索研究阶段。

第一个阶段:面向软件代码的复杂性度量研究阶段(20世纪60年代至20世纪70年代初期)

20世纪60年代,学术界开始关注软件系统的复杂性,Rubey等^[26]在1968年发表的论文可能是最早的关于软件复杂性的论文。由于该阶段软件技术的核心是程序代码,因此,学者提出可以用代码行(Lines of Code, LOC)作为软件复杂性的衡量指标^[27],即行数越多,系统越复杂,认为代码长度能预测程序的特性,如可靠性、易维护性等。这种方法是最早也是最简单直观的

复杂性度量方法,但正因为简单,反而受到了学术界的质疑,学者 Basili 等^[28]建议只将 LOC 作为与其他度量相比较的评估指标。

这个阶段的学者将信息系统视为软件技术系统,明确了可以通过度量软件中的“元素”或“属性”来衡量软件的复杂性^[29],即“元素”指标的上升意味着系统整体复杂性的提高,但是对系统复杂性的具体研究还处于起步阶段,甚至还没有给出信息系统复杂性的定义。此后随着软件开发方法的不断演进,复杂性的研究特别是度量研究开始成为学者研究的热点。

第二个阶段:面向软件结构的复杂性度量研究阶段(20 世纪 70 年代中后期至 20 世纪 80 年代)

1965 年 Dijkstra 首次提出了结构化程序设计的概念,由此进入了软件技术的程序系统阶段。在该阶段,由于高级语言的发展,软件作为一种产品开始被广泛使用,业界对软件的认识也从“技巧”和“个体行为”向“工程”和“群体协同工作”转化,因而该阶段逐步形成了广泛应用的结构化程序设计方法。相应地,20 世纪 70 年代中后期至 20 世纪 80 年代,学者尝试给出信息系统复杂性的定义,如 Curtis 等^[30]提出“系统复杂性是指软件系统中的过程特点,这些特点使得系统难以理解和改变”,进一步地,学者们探索了基于软件结构的多种复杂性度量方法,主要是通过软件中的模块^[31, 32]、控制流^[33-35]、数据流^[36]、程序组块^[37]及程序数据结构^[38]等要素来测算系统的整体复杂性。在这一阶段中,被引用最多的是 McCabe 的环复杂性(Cyclomatic Complexity)^[34]和 Halstead 的软件科学(Software Science)^[36]复杂性方法,此外还有 Woodward 等^[33]提出的基于控制流的节点(Knot)复杂性度量方法,Davis^[37]提出的基于数据流的组块(Chunk)复杂性度量方法,这些经典度量方法的提出,也为后续学者设计新的度量方法提供了可靠的比较依据。

这个阶段的学者开始试图给出软件复杂性的相关定义,并基于软件的代码结构对复杂性进行度量,主要思路是通过软件结构

中的控制流、数据流等元素来分析系统的整体复杂性。该阶段提出的诸多方法为此后的研究奠定了基础,但是上述成果还只是基于软件系统中的单一元素进行度量的结果,缺乏对复杂性的多维刻画和分析。

第三个阶段:多维复杂性探索及面向功能体系的复杂性分析研究阶段(20世纪80年代至20世纪90年代)

如前所述,在20世纪80年代之前,大多数学者是基于软件系统中的单一元素,从不同的角度提出了多种复杂性的度量方法,但也有学者认为,从单一维度度量,难免会遗漏信息,更好的度量应包含更多的要素,以更好地揭示系统的复杂性^[39],而且复杂性本质上是希望降低问题的维度^[40],因此应先从多个维度构建复杂性模型,再通过相应算法进行分析和度量^[32, 40-42]。如Konsynski等^[43]提出可依据软件的3个性质——体积、分布和位置,将复杂性分为软件要素复杂性、结构复杂性及接口复杂性3个方面;Basili和Hutchens^[32]则从软件结构角度,提出应从软件的结构规模、控制组织及数据组织3个维度来度量软件复杂性;Adamov和Richter^[42]也从软件结构层面研究,认为软件是由各要素相互连接形成的结构,其复杂性会受内部复杂性、位置复杂性、接口复杂性3个方面影响。

与此同时,20世纪80年代,软件技术进入软件工程阶段。在该阶段,研究者认为单一的程序系统很难保证达到系统的预期目标,需要重视开发之前的需求分析和设计,该阶段技术的方法论开始向需求分析前移,而自从1987年Zachman^[44]提出软件体系结构的概念,软件的体系结构开始受到重视,相应的复杂性度量研究也转为关注软件系统分析和设计阶段。此后,陆续有学者提出以软件的功能、体系结构等为依据来度量软件的复杂性,以解决软件系统的设计、优化和重构等方面的问题。软件功能的复杂性度量起始于Albrecht^[45]提出的功能点(Function Point, FP)方法,功能点来源于需求说明书,这种方法可以用于软件生命周期