

校 学 术 经 典 文 库

IT项目风险 决策规则挖掘研究

卢新元 著

A Study on the Risk Decision Rules
Mining of IT Project

湖北长江出版集团
湖北人民出版社

高 校 学 术 经 典 文 库

IT项目风险 决策规则挖掘研究

卢新元 著

A Study on the Risk Decision Rules

Mining of IT Project

湖北长江出版集团
湖北人民出版社

鄂新登字 01 号

图书在版编目(CIP)数据

IT 项目风险决策规则挖掘研究/卢新元著.
武汉:湖北人民出版社,2006.12

ISBN 7-216-05035-5

I. I...

II. 卢...

III. 信息技术—高技术产业—项目管理:风险管理—研究

IV. F49

中国版本图书馆 CIP 数据核字 (2006) 第 014497 号

IT 项目风险决策规则挖掘研究

卢新元 著

出版发行: 湖北长江出版集团
湖北人民出版社

地址:武汉市雄楚大街 268 号
邮编:430070

印刷:武汉中科兴业印务有限公司

经销:湖北省新华书店

开本:850 毫米×1168 毫米 1/32

印张:6.25

字数:147 千字

插页:1

版次:2006 年 12 月第 1 版

印次:2006 年 12 月第 1 次印刷

书号:ISBN 7-216-05035-5

定价:15.00 元

本社网址:<http://www.hbpp.com.cn>

前 言

随着我国“信息化带动工业化”发展战略的不断深入，信息化建设已经成为企业提升核心竞争力的主要途径。而 IT 项目作为信息化的一部分，也因此受到越来越多企业的关注。但是，IT 项目的实施是一项风险事业，其成功率一直都是很低的，项目的成功与否取决于实施过程中对风险进行有效的识别、预测和控制。因此，风险管理问题已成为当前 IT 项目实施中面临的主要问题。

IT 项目的实施过程涉及风险因素较多，具有不确定性和非结构化等特点，难以用单一的模型和算法解决。传统的项目风险分析大多是根据管理者的个人经验来进行的，而且假设条件太多，与实际情况有较大的差别，因此，传统的方法具有一定的主观性和随意性。针对这些问题，本书通过问卷调查，对 IT 项目的风险问题进行了深入系统的研究，提出以粗糙集理论为主、多方法融合的风险决策过程中的规则挖掘方法，其目的是在信息不完整、不精确的前提下，无需传统意义上复杂的建模过程，直接从大量的客观数据中发现隐含知识，实现对 IT 项目的风险等级预测，以及对风险决策规则进行挖掘的目的，从而提高风险决策和判断能力，促进 IT 项目风险管理向着智能化、科学化、定量化的方向发展，并最终提高项目的成功率。

本书主要研究 IT 项目风险分析与决策过程中的规则挖掘问题, 主要内容包括以下几个部分:

本书首先评述了国内外 IT 项目风险管理的现状, 对 IT 项目的范围和定义进行了描述分析, 并通过调查问卷的分析结果, 总结了 IT 项目不成功的主要表现, 对 IT 项目的风险因素进行了归纳与分类。

其二, 利用粗糙集方法对风险因素进行预处理, 主要包括知识分类、分类质量以及属性间的相互依赖性。指出单个属性权重分析所存在的问题, 重点研究了基于粗糙集的组合权重和聚类分析, 通过引入多元属性的组合权重这个概念, 利用欧拉距离公式来构造组合权重的相似矩阵 R_{com} , 分析了属性内部之间的关联, 从而实现对风险因素进行聚类分析。

其三, 研究了基于粗糙集标准分类的风险规则挖掘方法。首先通过实例介绍基于分类一致性的规则挖掘算法 (RICCR), 在此基础上, 利用可识别矩阵 $C_D(i, j)$ 和分辨函数来简化求解的过程, 在规则简化阶段, 利用有效规则权重的方法对规则进行优化与过滤, 实现一致性规则挖掘目的。对于分类不一致的决策系统, 提出了一种改进的规则挖掘方法, 通过上下近似集合分别获取确定性和可能性规则。

其四, 研究了不完全信息下的 IT 项目风险决策规则的挖掘方法。首先构建了基于对象相似的空值估算模型, 并结合实例, 给出了基于粗糙集相似关系的不完全信息规则挖掘方法。然后, 通过约简格与粗糙集相结合的方法, 从样本数据出发, 构建具有不同简化层次的约简格, 对格中的各层节点计算出满足给定可信度阈值 μ 的规则集, 在进行规则推理或决策时, 根据已有的信息在格中逐层进行匹配, 并按照优先级判定算法, 得到最优缺省规则。

其五,从动态的角度出发,研究了IT项目的风险决策规则的挖掘。首先指出基于粗糙集的 γ 准则属性约简方法的不足,即缺乏对属性信息动态变化的考虑,因此提出了对 γ 准则进行修正的方法,即信息熵准则,从 $H^{dc}(d/Q)$ 和 $H^{bc}(d/Q)$ 两个方面来分析基于信息熵的动态规则挖掘方法。然后给出了规则变化和近似动态规则的推理过程,该方法弥补了静态规则获取过程中存在的一些缺陷,以及由于噪声污染等原因引起的信息失真和规则变化所带来的不足。

最后,研究了将粗糙集与贝叶斯原理相结合的方法在规则挖掘中的应用。首先指出了单纯贝叶斯理论和粗糙集处理不完全信息所存在的不足,并详细分析了贝叶斯分类器。然后将粗糙集方法和NB贝叶斯分类器相结合,利用粗糙集对数据进行约简,然后利用NB分类器训练约简后的数据。该方法不仅简化了数据的规模,而且具有对不完全数据分类的能力和增量式学习的能力,既避免了贝叶斯过分依赖先验经验的弊端,也克服了单纯依赖粗糙集进行规则推理的缺陷。实践表明,将不同的方法相结合构成的综合规则挖掘系统可以克服各自的局限性,其功能要比单一系统的功能更加强大。

本书是在作者博士论文的基础上完成的,该博士学位论文是在导师张金隆教授的指导下完成的,整个论文和专著都倾注了导师大量的心血和汗水。衷心感谢导师给我的悉心指导和帮助,同时也感谢华中科技大学管理学院管理信息研究所的全体成员,如王林老师、谢刚博士、丛国栋博士、陈涛博士、王斌博士、张东风博士、刘汕博士等等,本文的研究也得到了他们极大的支持与帮助,在此一并表示感谢。同时,本研究得到国家自然科学基金项目“基于粗糙集的投标风险分析与规避决策模型研究(编号:70271031)”的大力资助。

我要感谢我的家人多年来对我的关心，没有他们的支持，我是没法完成我的博士论文的，也就不可能有本书的出版。

在本书的撰写过程中，参考了许多国内外最新的研究成果，作者尽可能详细地在参考文献中列出各位专家的研究工作，在此对他们的贡献表示深深的谢意。本书也可能应用了某些资料而由于作者的疏忽未能指出文献出处，在此表示万分歉意。由于作者水平所限，文中难免存在不妥之处，敬请读者批评指正。

本书的出版得到了湖北人民出版社的大力支持，尤其是张琦编辑在本书的出版过程中提出了很多指导性建议，给予了极大的帮助，在此表示衷心的感谢！

作 者

2006 年 11 月于武汉

目 录

第 1 章 绪 论	(1)
1.1 背景、目的和意义	(1)
1.2 国内外关于 IT 项目风险管理研究进展	(6)
1.3 与本书研究相关的理论和方法综述	(17)
1.4 本书的研究内容与结构	(28)
第 2 章 研究对象及问题界定	(32)
2.1 项目与项目风险管理	(32)
2.2 IT 项目风险管理	(35)
2.3 IT 项目风险因素调查	(42)
2.4 本章小结	(47)
第 3 章 基于经典粗糙集的 IT 项目风险因素数据分析	(48)
3.1 风险信息的知识表达与分类	(48)
3.2 知识依赖与属性重要性分析	(58)
3.3 风险因素权重分析	(61)
3.4 基于组合权重的风险因素聚类分析	(67)
3.5 本章小结	(71)
第 4 章 基于标准分类的风险决策规则挖掘	(72)
4.1 规则获取与分类一致性	(72)
4.2 基于分类一致性的规则挖掘	(75)
4.3 基于可识别矩阵的规则挖掘	(84)
4.4 基于分类不一致的规则挖掘	(94)

4.5	本章小结	(102)
第 5 章	不完全信息下 IT 项目风险规则挖掘	(103)
5.1	不完全信息系统	(103)
5.2	基于相似关系的不完全信息规则挖掘	(105)
5.3	基于约简格的缺省规则挖掘模型	(118)
5.4	本章小结	(129)
第 6 章	基于粗糙集和信息熵的动态规则挖掘	(130)
6.1	引言	(130)
6.2	基于信息熵的属性约简	(131)
6.3	基于信息熵准则的动态属性约简	(135)
6.4	规则变化与近似动态规则挖掘	(140)
6.5	本章小节	(144)
第 7 章	基于粗糙集和贝叶斯理论的风险规则挖掘	(146)
7.1	引言	(146)
7.2	基于粗糙集的贝叶斯分类器	(147)
7.3	基于粗糙集的贝叶斯规则获取	(154)
7.4	本章小结	(165)
第 8 章	全书总结与研究展望	(167)
8.1	全书总结	(167)
8.2	研究展望	(171)
参考文献		(174)

第 1 章 绪 论

1.1 背景、目的和意义

1.1.1 背景分析

20 世纪 80 年代以来,随着信息技术的发展,IT 产业作为一个新兴的知识型产业,已经成为世界各国的主要经济增长点,而信息技术和 IT 产业的发达程度则体现了一个国家的综合国力,并在一定程度上决定着一个国家未来的国际竞争力。在这样一个大的环境和发展趋势下,信息技术成为继工业革命后企业提升自身竞争力的又一个有力途径,这必将导致 IT 项目走向一个前所未有的发展高潮。

人们对 IT 项目的了解,更多的是从 Brooks 分别于 1975 和 1987 年出版了两本里程碑式的软件工程知识著作《人月神话》和《没有银弹》开始的^[1,2],让人们从思想深处认识到了软件工程和 IT 项目的本质和内涵,尤其是《人月神话》发表后的 30 多年里,人们在 IT 项目管理领域取得了很大的进步和提高。但是,由于 IT 项目具有规模大、周期长、投入资金大等特点,导致大量不完全信息的产生,其实施过程是一项复杂的系统工程,充满了巨大的风险。而影响 IT 项目的风险因素众多,这些风险因素涉及到很多方面,同时,各风险因素之间的关系错综复杂,所引起后果的严重程度迥异。因此,IT 项目

能否取得预期效果在很大程度上取决于对项目实施过程中所涉及风险因素进行有效的识别、分析、判断,并给出风险规避与规避方案。

据斯坦迪什咨询集团公司(Standish Group)2000 的调查发现:国外大型的公司在一个信息技术开发项目上平均投入230 万美元;中型公司要花费130 万美元以上,小型公司也要花费43.4 万美元以上。调查同时发现,在已经完成的项目中,有50%超过预算的60%~190%,而只有25%的项目在预算和进度之内提供给客户满意的产品。在中国IT 项目中,有30%~45%在完成之前就已经失败了,这种失败主要表现在:约有70%的IT 项目超出预定的开发周期,20%~50%的大型项目超出计划交付时间,90%以上的软件项目开发费用超出预算。众多项目的失败,使许多企业对IT 企业缺乏信任,在是否进行信息化的决策时犹豫不决。这种局面若不改变,将会使中国企业的信息化进程受阻,企业丧失信息化带动工业化的良机,企业的商机也将越来越少。由此可见,风险管理与控制是IT 项目管理中一个不可缺少的环节,有效的IT 项目风险管理对许多企业来说是一项挑战性的工作^[5,6]。

Keil. M 与Wallace. L(2000)认为^[7],风险管理的投资回报率一般在700%~2000%之间,软件系统中90%的风险可以检测到,50%~70%的风险可以避免。而一项KPMG 的研究发现,55%的项目失控(Runaway Projects)根本没有进行风险管理,38%的失控项目只做了一些风险管理工作,但是其中有一半在项目进行过程中没有处理已经发现的风险,而另外7%的失控项目并不知道是否进行了风险管理工作。这项研究表明,实施风险管理对提高项目成功率、防止项目失控是非常重要的。

目前,中国IT业正处在高速发展、并急需规范管理的环境中。一方面,企业要不断面对各种各样IT项目的实施和管理等活动;另一方面,客户需求的不断提高导致产品生命周期缩短,项目开发的难度不断增加,新技术的更新换代也导致了开发项目需求的增加等。因此,在技术含量高、变化速度快、资源有限的环境中,如何对企业、项目、资源实施科学的管理,加强团队能力,使软件项目的生产规范化、规模化、国际化,是当前我国IT企业面临的最大挑战。

遗憾的是,我国现有的IT项目风险管理体系还不是很完善,风险管理实践还仅仅处在起步和摸索阶段。不可否认,经过多年的摸索、总结和努力后,企业对于IT项目风险管理的意识得到了极大的提高,很多企业已经明确将风险管理作为IT项目实施和管理的一部分,并且制定了合理的计划与安排。然而,在现阶段的IT项目实施过程中,风险分析与管理大多数还停留在仅凭经验和直觉的阶段,尤其是对于大量不完全信息的IT项目风险决策,往往都带有很大的主观性,缺乏一定的理论依据。而传统的概率统计、模糊数学等方法存在不同程度上的局限性,尤其是现代IT项目实施过程中产生的大量不确定、不完全信息,更是给风险决策带来了相当的难度,因此,针对IT项目所面临的这类非结构化和半结构化的不完全信息风险决策问题,必须将定性与定量方法相结合,才能制定正确的风险规避与决策方案。

粗糙集理论(Rough Set)作为一种处理不确定、不完备信息的新的数学工具,已经被广泛地应用到数据挖掘、知识发现、决策支持、专家系统、规则推理和模式识别等方面^[8,9,10,11,12]。粗糙集理论的出发点是根据问题的描述对研究论域进行划分,通过引入不可分辨关系定义了上、下近似等概

念,因而成为知识发现、知识获取、机器学习等领域的一种重要研究方法。粗糙集理论具有一些独特的观点,如知识的粒度性,粗糙集理论认为知识的粒度性是造成使用已有的知识不能精确地表示某些概念的原因。

粗糙集的主要特点是:不需要任何预备或额外的有关数据信息,如统计学中的概率分布、模糊集理论中的隶属度或隶属函数等,而是直接从给定的信息出发,通过不可分辨关系确定问题的近似域,找到隐含在数据中的内在规律,从而避免了主观因素的影响。

而 IT 项目的实施过程中产生大量的不完全、不确定的信息,利用传统的方法处理这类非结构化的决策问题效果一直都不是很理想,因此,对于 IT 项目的风险决策问题,目前大多还停留在定性分析的基础之上,或者仅仅只是依靠个人经验来做出决策。因此,将粗糙集引入 IT 项目风险分析是必然的考虑,也是一项颇具开拓性的基础应用研究。本论文正是在这种背景下,把 IT 项目作为研究对象,着眼于 IT 项目风险分析与粗糙集方法的结合,并综合运用数据挖掘、知识发现、人工智能等多种方法,对 IT 项目风险进行全面系统的分析,重点研究了风险评估与决策过程中的规则挖掘与推理问题。

1.1.2 本书研究的目的和意义

IT 项目是一个风险与收益并存的过程,但是据相关调查显示,当前只有不到 1/3 的 IT 项目是成功的,大多数都是以失败而告终。由此可见,在当前的实际情况下,IT 项目的风险是远远大于其收益的,因此,如何降低并规避风险、提高 IT 项目的成功率,是当前 IT 项目实施过程中所面临的主要问题。因此,风险管理自然就成为 IT 项目实施中的一个重要组

成部分,而风险决策尤其是不完全信息下的 IT 项目风险决策规则的挖掘,则是 IT 项目风险管理中一个及其重要的课题。

本书所研究的主要目的是解决风险决策中的规则挖掘问题,具体如下:

1) 针对目前 IT 项目风险决策过程中存在的问题及困难,本书以企业信息化项目的实施效果为实际背景,对涉及到 IT 项目实施的风险因素进行充分调研,探讨基于粗糙集的风险知识表达和知识获取,利用粗糙集与其他方法相结合的模型对风险决策中的不确定数据进行有效的分类、约简与推理,从原始数据集中发现风险因素之间的关系,获取有用的风险决策规则。

2) 在风险决策与规则优化过程中,本书充分考虑将粗糙集与其他方法相结合,例如信息熵、约简格、贝叶斯理论等,有效互补,综合各种方法的优势,扩大粗糙集在风险分析与规则挖掘领域的应用范围。

3) IT 项目风险管理研究和实践具有重大的学术和经济价值,但是健康有效地开展这项活动必须有正确的体系指导,本书的研究目的就是为广大 IT 项目管理人员提供正确的风险管理指导体系。

本书的研究属于基础性应用研究,具有十分重要的理论和现实意义,主要表现在以下方面:

1) 其理论意义在于:有力地推动了粗糙集在不确定决策系统中的应用,尤其是将粗糙集理论成功地应用到 IT 项目风险管理中去,并融合了其他方法和知识,通过对不确定性的原始数据进行有效约简、分析、归纳和推理,从中发现 IT 项目风险因素相互之间的关系,并获取有用的风险决策规则,从而丰富了 IT 项目风险管理理论体系,为 IT 项目的风险分析与

管理提供了新的思路与方法。

2) 其现实意义在于:通过对项目实施过程中表现出来的各种风险因素进行分析,科学合理地对决策规则进行挖掘,为 IT 行业和其他企业在 IT 项目实施过程中,如何有效地控制与规避风险,提供了可操作性的建议和方法;同时也为项目管理人员的风险决策提供直接依据,保证项目实施的顺利进行,从而提高 IT 项目的成功率,促进企业信息化的快速发展。

1.1.3 本书研究的课题来源

本书的选题和出版得到了国际自然科学基金项目“基于粗糙集的投标风险分析与规避决策模型研究(项目编号:70271031)”的大力资助,而本书的研究内容主要为资助项目中关于“IT 项目风险评估与风险决策过程中的规则挖掘”部分的内容。

1.2 国内外关于 IT 项目风险管理研究进展

1.2.1 国外 IT 项目风险管理研究综述

风险管理是人们对潜在的、意外的损失进行计划、识别、分析、应对、跟踪和控制的过程,风险管理是项目管理的重要组成部分,它贯穿于项目生命周期的始终。早在 20 世纪初,德国人在第一次世界大战后就提出了一套风险管理的方案,包括风险的控制、风险分散、风险补偿、风险预防、风险回避等思路,最后逐步传输到欧洲其他国家,并形成了一门综合性学科。而将风险管理应用到 IT 项目中则是 20 世纪 70 年代后,随着信息技术的不断发展,IT 项目风险分析逐渐成为 IT 行业

的研究热点,并不断发展壮大,成为了当前各国学者的重点研究对象。以美国为例,从1984年至2000年,与IT项目风险管理研究相关的论文就不下200篇,并不断总结出IT项目风险管理的一些基本方法,如Boehm^[13,14](1989、1991)、Humphrey(1989)、Raftery^[22](2003)、Akintoye和Macleod(1997)归纳了IT项目风险分析的主要方法,如专家打分法、概率分析、决策技术、敏感性分析、随机控制、蒙特卡罗模拟(Monte Carlo)、CIM模型、影响图方法、灰色系统理论(Grey system theory)、层次分析法(AHP)、期望值法(EMV)、头脑风暴法(Brainstorming)等^[23,24,25,26]。这些方法,可以分为:定性分析方法,如专家判断、风险事项跟踪等方法;定量分析方法,如特征值方法、影响矩阵、决策树(Decision tree)方法。

Boehm^[6](1989)认为依照风险源清单(risk checklist),可以发现项目中大多数关键的风险因素,该思想的核心是利用专家经验和统计数据,列出现阶段最可能出现的前若干项重要风险因素,据此在一定程度上免除实际工作者的风险辨识工作。该思想的优点是简单、方便且成本低廉,缺点是风险清单需要随时间、应用领域而及时调整。由于IT应用范围日益扩大,致使风险清单规模不断膨胀。

Marvin^[20](1993)提出了“基于分类的风险辨识”的思想,他认为项目风险是在不断改变的,而风险清单是对风险结构的静态描述,因此,传统的方法不足以真实地反映风险的状况。他按风险来源构造了一棵分类树,共分3层,顶层分为“产品工程”、“开发环境”和“方案约束”3个结点,并辅以配套的详细问卷调查表。IT项目的风险辨识过程,就是修剪和量化这棵分类树的过程。该思想提高了辨识的灵活性,弱化了辨识的

结构性,然而同时对辨识成本和配套手段的要求也有一定程度的提高。

Haimes^[21] (1995) 等认为属于复杂系统的 IT 项目风险结构不能以单一视角、用单一模型来描述,但若分别用多个独立的模型描述则不利于模型间的通讯,需要采用“层次全息模型(HHM)”将复杂系统以互补、协作的方式分解为部件、子系统等层次,比如从时间维、系统维、工具维、地域维、方法维和人力资源维来综合分析一个 IT 项目。但是,本来深刻地说明了风险结构的这一思想,却因为思想提供结构构建步骤极为灵活,对实施者的知识和经验要求过高,以及辨识的可重复性和有效性对实施者的依赖程度过大,所以应用范围不广。

Cooper 和 Chapman^[22] (1987) 按着风险的特性将风险分为技术风险与非技术风险。Perry 与 Hayes 于 1985 年, Mustafa 与 Albahar 于 1991 年分别分析了 IT 项目的核心风险,并对传统的风险因素进行了适当的补充。1996 年 Wirbaetal 将 Tahetal、Cooper 和 Chapman 的研究成果进行了综合,按照 HRBS (Hierarchical Risk Breakdown Structure) 方法对风险进行分类。

Chin—Feng Fan^[27] (2004) 将 BBN (Bayesian belief networks) 理论引入到 IT 项目的风险管理中,并通过 BBN 建立起有效的反馈系统,从而预测潜在的风险,并进行动态调整,该方法提高了风险决策的可靠性,而且在实际操作中比较实用。

T. Raz 和 E. Michael^[28,29] (2001) 通过调查问卷的形式,总结了当前 IT 项目风险管理中所用到的主要方法,并对这些方法进行了详细的分类,尤其是在风险管理的几个关键阶段,通过数据来重点阐述了如何合理地利用现有的方法。