

基于粗糙集的软件项目风险管理研究

丁剑洁^{1,2} 郝克刚¹ 侯红¹ 郭小群¹

(西北大学信息科学与技术学院 西安 710127)¹ (陕西教育学院数理工程系 西安 710061)²

摘要 软件项目的复杂性导致软件风险涉及的数据具有很大的不确定性和模糊性。由此基于粗糙集理论,分析了软件项目风险管理的历史数据;利用等价类和分辨矩阵等概念提出软件项目风险分析系统,在此基础上给出了两个风险因子的相关性系数以及不可分辨矩阵的计算方法。此项工作能为风险控制提供策略,有助于发现风险管理活动中的薄弱环节,达到改进过程的目的。

关键词 软件风险,风险管理,粗糙集,等价类,分辨矩阵

中图分类号 TP311 **文献标识码** A

Research of Software Project Risk Management Based on Rough Set Theory

DING Jian-jie^{1,2} HAO Ke-gang¹ HOU Hong¹ GUO Xiao-qun¹

(Dept. Information Science and Technology, Northwest University, Xi'an 710027, China)¹

(Dept. of Math and Physics, Shaanxi Education Institute, Xi'an 710061, China)²

Abstract The complexity of software process leads to that risk data is uncertain and rough. Based on the rough set theory, this paper analyzed history data of software risk management. Using concept of equivalence class and discernible matrix, put forward risk factors of correlation analysis system, and defined calculate formula of dependence coefficient and risk factor discernible matrix. The work provides policy to risk control and help organization to find their shortcomings in risk management and improve their process.

Keywords Software risk, Risk management, Rough set, Equivalence class, Discernible matrix

快速开发满足用户需求的软件新产品,是每个软件组织追求的目标。但是由于软件项目具有很大的风险性,达到这个目标成为困扰很多软件组织的难题^[1]。因此,软件风险管理一直吸引了很多学者的研究。

从文献看^[2-11],软件风险研究多是从风险管理的角度出发,主要涉及风险管理的内容、过程以及活动等。另有国内外学者运用数学工具研究软件项目风险管理中的问题^[12-15]。本文侧重于后者。Sun-Jen Huang 利用簇分析法分析了软件项目持续时间与风险暴露之间的关系^[12];Neumannt 提出一种把模式识别、多变量统计以及神经网络相结合对风险因子进行分类的技术^[13];Xu 则提出了一种机遇模糊专家系统的风险识别和分析系统^[14]等,但是尚没有风险因子相互关系方面的研究。本文提出了一种探索风险因子之间关系的方法,以提高风险管理效率,协助寻找过程管理的薄弱环节,从而达到节约资源和实现过程改进的目的。

CMMI 的推广在很多软件组织中留下了大量的历史数据,例如软件项目风险数据库。但由于风险的不确定性,导致很多关于风险的数据都是模糊和不确定的,这个特性决定了运用粗糙集理论来度量软件风险数据是适合的。本文首次应用粗糙集理论表示一组软件项目的风险数据,寻找风险因子之间的相互关系,并通过实例予以说明。

1 软件风险管理

软件风险管理旨在制订一系列可行的原则和实践,规范化地控制影响项目成功的风险,目的是识别、分析和消除风险因素,以免它们威胁软件的成功运作^[4]。在此基础上,软件风险管理致力于两个活动:其一是识别风险因子并分析其重要性,通过评估风险发生的概率和它对项目的影响来确定;其二是制定策略去管理、控制已确定的风险因子。

风险因子是软件项目管理中的核心概念。关于软件项目的风险列表在文献和实践中很常见,它为组织识别风险项目提供了很好的指导,例如 Boehn 提出的“top ten”风险列表^[4,6]。Addison, Vallabh, Barki 和 Schmidt 等人也对风险列表进行了深入描述^[2,7,8]。同时,也有调查表明^[8-10],因为文化、管理模式造成的环境不同,导致这些风险列表并不适用于所有的软件组织。所以,探究不同的风险因子之间的相互关系是很重要的。例如,如果两个风险因子总是同时出现,并且发生的概率都很高,就应该引起注意。这些潜在的信息能帮助风险分类,应采取不同的策略去控制它们。当然,这些工作需要保留了大量和全面的历史数据的基础上开展。

2 基于粗糙集的软件项目风险度量

2.1 粗糙集的基本概念

到稿日期:2009-05-12 返修日期:2009-07-23 本文受 863 国家重点基金项目(2007AA010305)资助。

丁剑洁(1979-),女,博士生,主要研究方向为软件工程,E-mail:dingjianjie@yeah.net;郝克刚(1936-),男,博士生导师,主要研究方向为软件体系结构等;侯红(1966-),女,副教授,主要研究方向为软件过程改进等;郭小群(1972-),女,讲师,主要研究方向为形式化方法。

粗糙集理论是 Pawlak 提出的一种强大的数学分析工具,用来处理不精确知识,已经在人工智能、数据挖掘和模式识别等领域得到了广泛应用^[17-19]。此处先介绍几个基本定义。

定义 1 设 $S=\langle U,A,V,F\rangle$ 是一个信息系统,其中 $U=\{x_1,x_2,\dots,x_n\}$ 是论域; $A=C\cup D,C\cap D=\emptyset$, C 称为条件属性集, D 称为决策属性集; $V=\cup V_a,a\in A,V_a$ 是属性 a 的值域; f 表示 $U\times A\rightarrow V$,是一个信息函数,它为每个对象的每个属性赋予一个信息值,即 $\forall a\in A,x\in U,f(x,a)\in V_a$,这样具有条件属性和决策属性的信息系统称为决策表。

在定义 1 的基础上,每个属性子集 $E\subseteq A$,决定了一个二元不可区分关系(等价关系) $IND(E)$ (简记为 E),即 $IND(E)=\{(x,y)\in U\times U|\forall a\in E,f(x,a)=f(y,a)\}$,其中 $U|IND(E)$ (简记为 $U|E$)表示 U 中的一个划分, $U|E$ 中的任何元素 $[x]_E$ 称为等价类。

定义 2 设 $S=\langle U,A\rangle$ 为一信息表, $U=\{x_1,x_2,\dots,x_n\}$, $B\subseteq A$,定义 S 中 B 的分辨矩阵为

$$M_B(S)=(\delta_B(x_i,x_j))_{n\times n}$$

其中, $\delta_B(x_i,x_j)=\{a\in B;a(x_i)\neq a(x_j)\}$ 。如果 B 已知, $\delta_B(x_i,x_j)$ 可简记为 $\delta(x_i,x_j)$ 。

分辨矩阵的元素 $\delta_B(x_i,x_j)$ 是用于区分对象 x_i 和 x_j 的属性集合。对于任意 x_i 和 x_j , $\delta_B(x_i,x_j)$ 满足下列性质:

- (1) $\delta_B(x_i,x_i)=\emptyset$;
- (2) $\delta_B(x_i,x_j)=\delta_B(x_j,x_i)$,即 $M_B(S)$ 是对称阵;
- (3) $\delta_B(x_i,x_j)\subseteq\delta_B(x_i,x_k)\cup\delta_B(x_k,x_j)$ 。

2.2 软件风险因子分析模型

基于粗糙集理论中等价类和分辨矩阵的定义,给出以下定义。

定义 3 $RS=\langle U,A,V,F\rangle$ 是软件项目风险分析信息系统,其中 $U=\{x_1,x_2,\dots,x_n\}$ 是一组软件项目; $A=R\cup D,R\cap D=\emptyset$, $R=\{r_1,r_2,\dots,r_n\}$ 是一组风险因子, $D=\{d_1,d_2,\dots,d_n\}$,其中 d_i 是软件项目 x_i 采用的风险管理策略, V 是 R 的值域, $f:U\times R\rightarrow V$ 是信息函数。特别地, $V=(u_{ij}),1\leq i\leq n,1\leq j\leq n$, u_{ij} 表示在项目 x_i 中风险因子 r_j 发生的可能性, $u_{ij}\in\{1,2,3\}$,见表 1 和表 2。

表 1 软件项目风险分析信息系统

	r_1	r_2	...	r_j	...	r_m	d
x_1	u_{11}	u_{12}	...	u_{1j}	..	u_{1m}	d_1
x_2	u_{21}	u_{22}	...	u_{2j}	...	u_{2m}	d_2
...
x_i	u_{i1}	u_{i2}	...	u_{ij}	...	u_{im}	d_i
...
x_n	u_{n1}	u_{n2}	...	u_{nj}	...	u_{nm}	d_n

表 2 风险因子发生概率级别含义表

等级	可能性
1	不可能
2	一般
3	很可能

根据等价关系的定义,可以得到集合 U/r_j ,其中包含的每个元素对于 x_i 都是不可分辨的。不难看出 $U/r_j=\{A_{1j},A_{2j},A_{3j}\},j=1,2,\dots,m,A_{1j}=\{x_i|u_{ij}=1,j=1,2,\dots,m\}$,实质上 A_{1j} 是风险因子 r_j 发生的概率为 1 级(即不可能)的所有软件项目集合。同理, $A_{2j}=\{x_i|u_{ij}=2,j=1,2,\dots,m\}$,表示风险因子 r_j 发生的概率为 2 级(即一般)的所有软件项目集

合。 $A_{3j}=\{x_i|u_{ij}=3,j=1,2,\dots,m\}$ 是风险因子 r_j 发生的概率为 3 级(即很可能)的所有软件项目集合。

通过对各个软件项目进行等价类划分,可以反映出组织在风险管理过程中的潜在规律。例如在数据分析的过程中,如果出现 $A_{1j}=U,A_{2j}=\emptyset,A_{3j}=\emptyset$,即对于风险因子 r_j ,所有的软件项目都记录此风险发生的概率为 1 级(不可能),那么在一定程度上表明该组织在风险因子 r_j 的控制活动是比较完善的;同理,如果出现 $A_{3j}=U,A_{1j}=\emptyset,A_{2j}=\emptyset$,即对于风险因子 r_j ,所有的软件项目都记录此风险发生的概率为 3 级(很可能),反映了组织在与风险因子 r_j 相关的过程中存在问题,应该以此作为切入,探寻原因,实施过程改进。

定义 4 设 $RS=\langle U,A,V,F\rangle$ 是一个给定的软件项目风险分析信息系统, $A=R\cup D,r_i\in R,r_j\in R,r_i\neq r_j,U/r_i$ 和 U/r_j 分别是关于风险因子 r_i 和 r_j 的等价关系集合,如果 $U/r_i=U/r_j$ 成立,则称 r_i 和 r_j 是完全相关的。特别地,如果 $A_{1i}=A_{1j},A_{2i}=A_{2j},A_{3i}=A_{3j}$,那么 r_i 和 r_j 是完全正相关的;如果 $A_{1i}=A_{3j},A_{2i}=A_{2j},A_{3i}=A_{1j}$,那么 r_i 和 r_j 是完全负相关的。

简单地讲,如果两个风险因子完全正相关,表明它们在各项目中总是被评估为相同的发生概率,在风险预防与跟踪中可以合并两个风险因子进行管理,以节约资源,提高效率;相反,如果两个风险因子完全负相关,即一个风险因子以较低的概率发生时,总伴随另一风险因子发生的概率很高,就需要组织采取折中的策略,避免集中预防一个风险因子而忽略另一个的发生。

在软件风险管理中,完全相关的两个风险因子是很少见的,需要对两个风险项之间的相关性做定量分析。

定义 5 设 $RS=\langle U,R,V,F\rangle$ 是软件风险分析信息系统, $r_i\in R,r_j\in R,r_i\neq r_j,U/\{r_i,r_j\}$ 是关于 $\{r_i,r_j\}$ 的等价类。 $D(r_i,r_j)$ 是风险因子 r_i 和 r_j 的相关系数,计算方法如下:

$$D(r_i,r_j)=\begin{cases} 1, & r_i \text{ 和 } r_j \text{ 完全相关时} \\ 1-\frac{|U/\{r_i,r_j\}|}{n}, & \text{其它} \end{cases}$$

$|U/\{r_i,r_j\}|$ 表示集合 $U/\{r_i,r_j\}$ 中的元素个数。 $D(r_i,r_j)\in[0,1]$, $D(r_i,r_j)$ 值越接近 1,表明 r_i 和 r_j 的相关性越高。只有当两者的相关性较高时,才可以考虑采用合并或折中的管理策略。另外,组织的管理模式可能会导致某些风险因子存在较高的相关性,需要进一步探寻其中的原因进行修正,达到过程改进的目的。

定义 6 设 $RS=\langle U,A,V,F\rangle$ 是软件项目风险分析信息系统,各符号含义同定义 3,则 RS 中的风险因子分辨矩阵为 $M_R(S)=(\delta_R(x_i,x_j))_{n\times n}$,其中 $\delta_R(x_i,x_j)=\{a\in R;a(x_i)\neq a(x_j)\}$ 。

风险因子分辨矩阵体现了一组风险因子在不同的软件项目中发生概率的区别,为制定风险管理计划等提供参考。例如,若 $\delta_R(x_i,x_j)=\emptyset$,表明风险因子集合 R 在软件项目 x_i 和 x_j 上发生的概率分布相同,那么风险策略 d_i 和 d_j 必有重复之处,可参考执行以节约资源。

3 实例分析

某公司在进行 CMMI 的三级评估时,建立了风险管理库,主要内容有 30 余项风险因子、发生概率、重要程度以及应对措施等。各个项目小组根据上述内容识别和管理各自的风

险因子。表3和表4是其中一个小组在2008年提供的风险管理数据。

首先,根据等价类的概念,得到以下划分:

$$\begin{aligned}
 U/r_1 &= \{\{x_1, x_2\}, \{x_4\}, \{x_3, x_5, x_6\}\}; \\
 U/r_2 &= \{\{x_5, x_6\}, \{x_1, x_4\}, \{x_2, x_3\}\}; \\
 U/r_3 &= \{\{x_1, x_2\}, \{x_4\}, \{x_3, x_5, x_6\}\}; \\
 U/r_4 &= \{\{x_1, x_2, x_3, x_4, x_5, x_6\}, \emptyset, \emptyset\}.
 \end{aligned}$$

表3 风险因子含义说明表

序号	风险因子
r_1	客户要求多变性
r_2	项目计划的低质量
r_3	系统需求不清晰
r_4	团队人员缺乏有效沟通

表4 实例数据表

	r_1	r_2	r_3	r_4
x_1	1	2	1	1
x_2	1	3	1	1
x_3	3	3	3	1
x_4	2	2	2	1
x_5	3	1	3	1
x_6	3	1	3	1

从划分结果中很容易看到,所有软件项目中风险因子 r_4 的发生概率都为1级(不可能),在一定程度上反映该项目中的人员能保持良好的沟通机制。进一步得到下列划分:

$$\begin{aligned}
 U/\{r_1, r_2\} &= \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6\}\}; \\
 U/\{r_1, r_3\} &= \{\{x_1, x_2\}, \{x_3, x_5, x_6\}, \{x_4\}\}; \\
 U/\{r_1, r_4\} &= \{\{x_1, x_2\}, \{x_3, x_5, x_6\}, \{x_4\}\}; \\
 U/\{r_2, r_3\} &= \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5, x_6\}\}; \\
 U/\{r_2, r_4\} &= \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_5, x_6\}\}; \\
 U/\{r_3, r_4\} &= \{\{x_1, x_2\}, \{x_3, x_5, x_6\}, \{x_4\}\}.
 \end{aligned}$$

根据定义5,计算两两风险因子之间的相关系数,得到 $D(r_1, r_2)=0.167, D(r_1, r_3)=1, D(r_1, r_4)=0.5, D(r_2, r_3)=0.167, D(r_2, r_4)=0.5, D(r_3, r_4)=0.5$ 。由定义4可知,风险因子 r_1 和 r_3 是完全正相关的,即客户需求的多变性和系统需求不清晰之间存在直接联系。

最后由定义6给出了风险因子的分辨矩阵,结合其性质,将它表示为如下三角阵:

$$M = \begin{pmatrix} r_2, r_1, r_2, r_3 & r_1, r_3 & r_1, r_2, r_3 & r_1, r_2, r_3 \\ & r_1, r_3 & r_1, r_2, r_3 & r_1, r_2, r_3 \\ & & r_1, r_2, r_3 & r_2 & r_2 \\ & & & r_1, r_2, r_3 & r_1, r_2, r_3 \\ & & & & \emptyset \end{pmatrix}$$

在矩阵 M 中,若 M_{ij} 所包含的风险因子越多,即越接近风险因子集合 R ,表明风险因子集合在软件项目 x_i 和 x_j 上具有差异很大的发生概率分布;相反, M_{ij} 所包含的风险因子越少,表明风险因子集合在软件项目 x_i 和 x_j 上具有相似的发生概率分布,后面的软件项目可以参考前者的风险管理活动。例如, $M_{56} = \emptyset$,说明软件项目 x_6 可参考 x_5 的风险管理计划。另外 $M_{35} = M_{36} = r_2$,软件项目 x_6, x_5 与 x_2 在风险因子 r_2 的跟踪与管理应有所不同。

结束语 探究软件风险因子之间的关系,是软件风险管理活动中重要的但又被忽略的一个环节。本文基于粗糙集理论主要研究风险因子之间的相关性以及不同项目之间的风险因子分布的相似性。一方面为选择管理策略提供辅助信息,

例如对两个拥有很大相关系数的风险因子采取合并跟踪的策略,两个风险分布相似的软件项目应该实施类似的管理活动;另一方面,组织的文化或管理模式可能导致某些风险因子含有一定的相关性,借此也可以寻找过程管理的薄弱环节,为过程改进提供切入点。文中根据粗糙集理论中的等价类、分辨矩阵等概念,给出了计算两个风险因子之间相关系数以及风险因子的分辨矩阵的计算方法。但是还需要做很多进一步的工作:风险因子之间的关系除了相关性之外,还应深入分析是否含有诱导关系;相关系数的概念应该从两个风险因子扩充到风险因子的集合;寻找更客观的计算风险因子相关系数以及分辨矩阵的方法等等。

参考文献

- [1] Charette R N. Why software fails? [J]. IEEE Spectrum, 2005, 42(9): 42-49
- [2] Addison T E. Commerce project development risks: Evidence from a Delphi survey [J]. International Journal of Information Management, 2003, 23 (1): 25-40
- [3] Aloini D, et al. Risk management in ERP project introduction: review of the literature [J]. Information & Management, 2007, 44(6): 547-567
- [4] Boehm B W. Software risk management: principles and practices [J]. IEEE Software, 1991, 8(1): 32-41
- [5] CMMI Product Team. CMMI for Systems Engineering/Software Engineering Version 1.02 [R]. CMU /SEI-2000-TR-019
- [6] Johnson J, Boucher K D, Conners Y, et al. Project management: the criteria for success [J]. Software Magazine, 2001, 21(1): S3-S11
- [7] Barki H, Rivard S, Talbot J. Toward an assessment of software development risk [J]. Journal of Management Information Systems, 1993, 10(2): 203-225
- [8] Schmit R, Lyytinent K, Keil M, et al. Identifying software project risks: an international Delphi study [J]. Journal of Management Information Systems, 2001, 17(4): 5-36
- [9] Paul L, Bannerma N. Risk and risk management in software projects: A reassessment [J]. The Journal of Systems and Software, 2008, 81: 2118-2133
- [10] Wallace L, Keil M, Rai A. How software project risk affects project performance: an investigation of the dimensions of risk and an exploratory model [J]. Decision Sciences, 2004, 35 (2): 289-321
- [11] 潘春光, 陈英武, 汪浩. 软件项目风险管理理论方法研究综述 [J]. 控制与决策, 2007, 22(5)
- [12] Huang S J, Han W M. Exploring the relationship between software project duration and risk exposure: A cluster analysis [J]. Information & Management, 2008, 45: 175-182
- [13] Neumann D E. An enhanced neural network technique for software risk analysis [J]. Software Engineering, IEEE Transactions on, 2002, 28(9): 904-912
- [14] Xu Z W, Khoshgoftaar E. Operational risk assessment of software using fuzzy expert systems [C] // T M, Allen E B, ed. World Automation Congress, 2002 Proceedings of the 5th Biannual. 2002, 13: 435-442
- [15] 蒋国萍, 陈英武. 基于面向对象贝叶斯网络的软件项目风险评估 [J]. 系统工程与电子技术, 2005, 27(2)
- [16] 冯楠, 李敏强, 寇纪淞, 等. 基于贝叶斯网络的软件项目风险分析过程 [J]. 计算机工程与应用, 2006, 18
- [17] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356
- [18] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data [M]. Dordrecht: Kluwer Academic Publishers, 1991
- [19] 安利平. 基于粗糙集理论的多属性决策分析 [M]. 北京: 科学出版社, 2008: 14-24