

基于粗糙集与 AHP 理论的 IT 项目风险群决策

范敏^{1,2}, 邹平², 朱兴东¹

FAN Min^{1,2}, ZOU Ping², ZHU Xing-dong¹

1.昆明理工大学 理学院 系统理论与应用数学系, 昆明 650093

2.昆明理工大学 管理与经济学院, 昆明 650093

1.Department of System Science and Applied Mathematics, Kunming University of Science and Technology, Kunming 650093, China

2.School of Management & Economics, Kunming University of Science and Technology, Kunming 650093, China

E-mail: fanmin9412@21cn.com

FAN Min, ZOU Ping, ZHU Xing-dong. Method of rough set and AHP theory in IT project risk group decision making. *Computer Engineering and Applications*, 2009, 45(21): 226-228.

Abstract: Aiming at the imprecision and uncertainty of risk factors, this paper analyzes the basic characters of IT project, and points out the problems existing in the group decision making. Meanwhile, the basic steps of RSGDS is put forward. With the rough set theory, the main processes of knowledge reasoning such as attributes reduction and rules mining are discussed, especially introducing the concept of dynamic weight, then the algorithm of rules reduction and mining is optimized and ameliorated.

Key words: rough set; Analytic Hierarchy Process (AHP) theory; IT project risk management; intelligent decision support system; dynamic weight

摘要: 针对 IT 项目的风险管理与群决策过程中存在大量不确定、不完全信息等特征, 在传统决策方法的基础上, 将粗糙集方法与 AHP 理论相结合, 提出了粗集 IT 项目风险群决策的基本步骤, 并借助粗糙集的原理与方法, 通过实例分析探讨了 IT 项目风险群决策中关于知识表达、属性约简、规则挖掘、方案排序等方法, 尤其是引入属性动态权重、专家主观权重、客观权重的概念与计算方法, 进而对整个决策排序算法进行了改进。

关键词: 粗糙集方法; 层次分析法理论; IT 项目风险管理; 智能决策系统; 动态权重

DOI: 10.3778/j.issn.1002-8331.2009.21.065 文章编号: 1002-8331(2009)21-0226-03 文献标识码: A 中图分类号: TP18

1 引言

IT 项目具有高投入、高风险等特点, 如何有效地识别风险、规避风险、及时采取正确决策, 这在 IT 项目实施中具有重要的意义。实际中, 希望通过学习专家经验从而获得项目风险的一般规则, 并在群决策中进行方案的评选。以下三点值得注意: (1) 各专家的评价指标体系是否相同; (2) 各专家对各指标的重要性度量; (3) 各专家的权威性度量。项目风险因素涉及到很多方面, 不同的专家有不同的评价指标体系。而且针对不同项目, 不同的专家对同一指标的重要性认识也并不完全相同。所以, 在群决策中用传统的预先给定评价指标及其权重的方法, 不能真实地反映专家经验。将粗糙集方法与群决策理论相结合, 首先构造一个可增减的指标体系, 接着从各专家评分数据中动态地得到各评价指标的权重, 并进行指标合并, 存入新指标体系; 然后, 根据各评价指标的动态权重定义专家的客观权重, 同时由 AHP 法给出专家的主观权重, 于是得到专家权重。最后, 根据指标动态权重、专家权重及专家评分数据进行方案的排序。给出了规则提取的基本步骤和算法, 建立了有效的

IT 项目风险群决策模型, 为 IT 项目的风险管理提供了一种新的思路与方法。

2 基于粗糙集和 AHP 理论的规则获取原理

2.1 粗糙集的基本方法与原理

粗糙集是由波兰数学家 Pawlak 于 1982 年提出的。它从知识分类的角度出发进行数据推理, 利用已有的信息来判断所有对象在特定条件下的分类及属性重要性, 从而找出隐含知识, 揭示潜在规律。

2.1.1 知识表达系统

设决策系统为 $T=(U, A \cup D, V, f)$, 其中 U 为论域 (非空有限集合), A 为条件属性集, D 为决策属性集, 且 $A \cap D = \emptyset, V =$

$\bigcup_{a \in A} V_a, V_a$ 是属性 a_k 的值域, $f: U \times A \rightarrow V$ 是信息函数, $\forall a \in A, x \in U, f(x, a) \in V_a$ 。

2.1.2 可辨识矩阵与属性权重

定义 1 可辨识矩阵

基金项目: 云南省教育厅基金 (No.07Y11385)。

作者简介: 范敏 (1975-), 女, 讲师, 博士生, 主要研究方向为数据挖掘、决策分析; 邹平, 教授, 博士; 朱兴东, 副教授。

收稿日期: 2008-04-23 修回日期: 2008-09-16

$$M(i, j) = \begin{cases} a_k/a_k(x_i) \neq a_k(x_j), D(x_i) \neq D(x_j) \\ \phi, a_k(x_i) = a_k(x_j), D(x_i) \neq D(x_j) \\ 0, D(x_i) = D(x_j) \end{cases}$$

记可辨识矩阵中各元素出现的频率为 $P(a_k), a_k \in B$, 则有

定义 2(属性的重要性) 设 $a_k \in A$, 则属性 a_k 的重要性为 $SIG(a_k) = P(a_k)$ 。

2.1.3 属性约简与规则提取

应用文献[1]中的命题 3: 子集可辨识矩阵 $M^s(i, j)$ 中, 当广义核 C^s 和 $M^s(i, j)$ 中条件属性个数相等时, 称求得决策系统 $T=(U, A \cup D)$ 的一个相对最优约简。

2.2 基于粗糙集与 AHP 理论的群决策规则获取步骤

- (1) 建立相关对象与属性可增减的指标体系(知识表达系统);
- (2) 对各专家知识表达系统进行预处理;
- (3) 对(2)中得到的各知识表达系统, 求动态属性权重(AEDW)并进行属性约简;
- (4) 将各专家的约简进行合并, 形成新规则属性集;
- (5) 求出各专家的主观权重(AHP 法)与客观权重;
- (6) 方案集结与排序。

3 实例分析

为了清晰地说明该文的方法, 将重点分析第 3 至第 6 步骤, 并且首先只考虑四个主要的风险因素(即条件属性), 以及一个决策属性, 且假定这些风险因素都是经过离散化处理的。

如表 1 所示: 属性 a1 表示质量控制, a2 表示需求定位, a3 表示开发工具, a4 表示评估与计划, 决策属性 D 表示风险等级。其中每个属性的取值范围及其意义为: 质量控制(a): 0 表示质量控制能力很强, 1 表示质量控制能力一般, 2 表示质量控制能力很差; 需求定位(b): 0 表示需求定位很明确, 1 表示需求定位一般, 2 表示需求定位不明确; 开发工具(c): 0 表示开发工具很先进, 1 表示开发工具一般, 2 表示开发工具比较落后; 评估与计划(d): 0 表示有详细的评估与计划, 1 表示有初步的评估与计划, 2 表示没有评估与计划; 风险等级(e): 0 表示风险等级低, 1 表示风险等级一般, 2 表示风险等级高。

表 1 专家 1 项目风险知识表达系统

U	a1	a2	a3	A4	D
1	1	1	0	1	0
2	0	1	1	2	0
3	1	0	0	1	1
4	1	0	2	0	1
5	1	1	0	2	2
6	2	2	0	2	2
7	2	0	0	2	2
8	2	2	2	0	2

步骤 1 可辨识矩阵为:

Φ	Φ	a2	a2a3a4	a4	a1a2a4	a1a2a4	a1a2a3a4
Φ	Φ	a1a2a3a4	a2a3a4	a1a3	a1a2a3	a1a2a3	a1a2a3a4
	Φ	Φ	a2a4	a1a2a4	a1a4	a1a4	a1a2a3a4
		Φ	a2a3a4	a1a2a3a4	a1a4	a1a4	a1a2a
			Φ	Φ	Φ	Φ	Φ
				Φ	Φ	Φ	Φ
					Φ	Φ	Φ
						Φ	Φ
							Φ

$$core = \{a_2, a_4\}$$

由于不同专家对同一属性的重要性认识往往不同, 而这种差异就反映在他们对各项目各属性的评分值有所不同, 于是定义:

定义 3(专家 k 动态属性权重(AEDW)) $w_k^D(a_i) = p_k(a_i)$, 其中 $p_k(a_i)$ 表示由专家 k 的信息表计算的可辨识矩阵中属性 a_i 出现的频率。

专家 1 动态属性权重(AEDW1)为:

$$w_1^D(a_1) = \frac{14}{56} = 0.25, w_1^D(a_2) = \frac{16}{56} = 0.27$$

$$w_1^D(a_3) = \frac{11}{56} = 0.2, w_1^D(a_4) = \frac{15}{56} = 0.28$$

约简为: $\{a_1, a_2, a_4\}$ 。

步骤 2 其他专家的属性约简集与专家动态属性权重同理可得, 将所有的约简集合并, 得规则属性集 $\{a_1, a_2, a_4\}$ 。

解释为: 由规则属性集知“需求定位”、“评估与计划”、“质量控制能力”完全可以确定 IT 项目的风险等级。

步骤 3 专家权重的确定

对各个专家的权采取主观权重和客观权重相结合的方法。假设有 n 个专家, r 个对象, m 个属性。

(1) 客观权重: 设 $S_k(x_i)$ 是第 $k(k=1, \dots, n)$ 位专家对第 $i(i=1, \dots, r)$ 个对象的评分:

$$S_k(X_i) = \sum_{l=1}^m w_k^D V(a_l) \tag{1}$$

则第 i 个对象平均评分为:

$$\bar{S}(x_i) = \frac{1}{n} \sum_{k=1}^n S_k(x_i) \tag{2}$$

设 D_k 表示第 k 位专家的偏差, 且

$$D_k = \sum_{i=1}^r [S_k(x_i) - \bar{S}(x_i)]^2 \tag{3}$$

由此, 定义客观权重为:

$$\omega_k^1 = 1 - \frac{D_k}{\sum_{k=1}^n D_k} \tag{4}$$

(2) 主观权重: 首先构造专家互评矩阵, 假设共有 n 个专家, E_{ij} 表示第 i 位专家对第 j 专家的分值, $E_{ij}=0$, 当 $i=j$ 时; 利用 Saaty (1980) 根据一般人的认知习惯和判断能力给出了属性间相对重要性等级表, 则可构造出专家的互评矩阵, 最终即可得到专家的主观权重:

$$\omega_k^2 = (\omega_1^2, \omega_2^2, \dots, \omega_n^2) \tag{5}$$

(3) 专家权重: 根据决策群体的偏好程度, 可以得到:

$$\omega_k^* = \alpha \cdot \omega_k^1 + (1-\alpha) \cdot \omega_k^2 \tag{6}$$

其中, $0 \leq \alpha \leq 1$, 表示决策人对客观权重的偏好程度。 $1-\alpha$ 为决策人对主观权重的偏好程度; 在不同的系统中, 人们对主客观权重的偏好程度各不相同。一般, 当评价系统是一个新系统时, 人们比较看重数据, 故此时可取客观权重 $\alpha \geq 0.5$ 。

步骤 4 综合评价模型的建立

综合评价模型可采用多种集结算子, 集结算子适用的条件各不相同, 而且算子的选择会影响到排序结果, 这部分内容将在以后的文章中进行讨论。

表3 9个项目得分及其平均值

学校专家	1	2	3	4	5	6	7	8	9
1	0.724 8	0.752 7	0.754 6	0.799 3	0.775 9	0.739 6	0.728 0	0.733 1	0.812 7
2	0.744 6	0.753 8	0.753 9	0.800 0	0.763 5	0.745 0	0.737 2	0.731 0	0.811 4
3	0.728 6	0.753 8	0.753 6	0.801 7	0.776 9	0.744 0	0.738 4	0.732 3	0.810 3
4	0.738 3	0.757 7	0.756 7	0.807 3	0.785 5	0.736 6	0.731 6	0.732 4	0.824 3
5	0.675 9	0.723 8	0.711 5	0.800 0	0.745 6	0.712 0	0.737 4	0.736 6	0.836 9
平均值	0.714 267	0.745 1	0.740 6	0.803 0	0.769 333 333	0.730 867	0.735 8	0.733 767	0.823 833

当评价系统的指标之间是可以补偿的,那么选择

$$C_i = \sum_{k=1}^n \omega_k^* \cdot S_k(x_i) \quad (7)$$

最后,可以依据 C_i 的大小进行排序。

采用了某 IT 公司 5 位专家对 9 个 IT 项目的风险信息数据(5 个指标)如表 2、表 3。

表2 专家风险信息数据表

项目	风险指标				
	1	2	3	4	5
1(4211)	7.11	3.94	4.6	0.20	0.50
2(4115)	8.01	3.93	4.4	0.210	0.52
3(5521)	7.20	3.56	4.7	0.180	0.60
4(7289)	7.50	3.42	5.5	0.190	0.54
5(2156)	5.60	2.64	8.9	0.220	0.51
6(1907)	7.36	3.10	4.8	0.176	0.52
7(1697)	6.48	2.75	5.5	0.240	0.59
8(1830)	4.77	2.60	6.6	0.210	0.53
9(1737)	8.44	3.30	5.9	0.20	0.45

由公式(1)可以计算出各个专家给 9 个项目的打分情况,

由 $\bar{S}(x_i) = \frac{1}{n} \sum_{k=1}^n S_k(x_i)$ 计算 9 个项目得分的平均值,如表 3 所示。

根据表 3 由公式 $D_k = \sum_{i=1}^f [S_k(x_i) - \bar{S}(x_i)]^2$ 可以计算出 5 位专家的偏差:

$$D_1=0.000 683, D_2=0.001 58, D_3=0.000 874$$

$$D_4=0.001 328, D_5=0.003 882$$

$$\text{则有 } \sum_{k=1}^5 D_k = 0.008 346.$$

由公式 $\omega_k^1 = 1 - D_k / \sum_{k=1}^5 D_k$ 可以计算出各个专家的客观权重

分别为: $\omega_1^1 = 0.918 2, \omega_2^1 = 0.810 7, \omega_3^1 = 0.895 3, \omega_4^1 = 0.840 9, \omega_5^1 = 0.534 9$ 。归一化后权重为: $\omega_1^1 = 0.229 6, \omega_2^1 = 0.202 7, \omega_3^1 = 0.223 8, \omega_4^1 = 0.210 2, \omega_5^1 = 0.133 7$ 。

(1) 计算各个专家的主观权重,首先要建立 5 个专家的互评矩阵,如表 4 所示(各个专家按表 1 Saaty(1980)根据一般人的认知习惯和判断能力给出了属性间相对重要性等级表,根据专家的认知情况,构造专家的互评矩阵,即其他专家在自己心目中的重要性),专家设定的矩阵如表 4 所示。

根据 AHP(层次分析法)可以计算出相应的权向量和 λ_{\max} 值, A 中每行元素连乘并开 n 次方:

$$\omega_i^* = \sqrt[n]{\prod_{j=1}^n a_{ij}} \quad i=1, 2, \dots, n \quad (8)$$

表4 5个专家的互评矩阵

	1	2	3	4	5
1	1	1	1	4	1
2	1	1	2	4	1
3	1	1/2	1	5	3
4	1/4	1/4	1/5	1	1/3
5	1	1	1/3	3	1

求权重:

$$\omega_i = \frac{\omega_i^*}{\sum_{i=1}^n \omega_i^*} \quad i=1, 2, \dots, n \quad (9)$$

A 中每列元素求和:

$$S_j = \sum_{i=1}^n a_{ij}, \quad j=1, 2, \dots, n \quad (10)$$

计算 λ_{\max} 的值:

$$\lambda_{\max} = \sum_{i=1}^n \omega_i S_i \quad (11)$$

此时 $n=5$, 用上述方法计算出来的 λ_{\max} 的值为 5.31, 表 1 中 5 阶矩阵的临界值为 $\lambda'_{\max} = 5.54$ 。且 $\lambda_{\max} < \lambda'_{\max}$, 通过一致性检验。此时对应的主观权重为 0.232 9, 0.267 5, 0.264 1, 0.059, 0.176 5。

(2) 综合主客观权重,可以得到最终的权重,运用式子 $\omega_k^* = \alpha \cdot \omega_k^1 + (1-\alpha) \cdot \omega_k^2$, 取 $\alpha=1/2$; 则有: $\omega_1^* = 0.231 2, \omega_2^* = 0.235 1, \omega_3^* = 0.243 9, \omega_4^* = 0.134 6, \omega_5^* = 0.155 1$ 。

(3) 根据假设属性之间的完全可补偿性,可以应用加权和法排序,即再根据公式 $C_i = \sum_{k=1}^n \omega_k^* \cdot S_k(x_i)$ 进行排序,由此可计算出:

$$C_1=0.724 6, C_2=0.749 4, C_3=0.747 8, C_4=0.801 2, C_5=0.769 8$$

$$C_6=0.737 3, C_7=0.734 6, C_8=0.732 9, C_9=0.817 1$$

根据 C_i 值的大小, 可以给定 IT 项目风险的优劣次序为: $C_9 > C_4 > C_5 > C_2 > C_3 > C_6 > C_7 > C_8 > C_1$ 。

所以综合(1)、(2)、(3)步骤即可获得 IT 项目风险排序方案,具体为项目 9 的风险最小而项目 1 的风险最大。

4 结束语

探讨了粗糙集方法在 IT 项目风险群决策分析中的应用,并重点讨论了粗糙集方法在属性约简、规则推理、方案排序等过程的应用,有效地解决了智能决策支持系统中关于粗糙集应用的知识推理的基本问题。作为粗糙集理论在人工智能和专家群决策系统这一新的应用领域,还有很多值得进一步探讨和改进的地方,比如如何提高系统的决策效率、增强系统对相关案