

基于粗糙集理论的隐性知识决策

The tacit knowledge decision making on the basis of rough set theory

伍大清 阳小华 (南华大学 计算机科学与技术系 湖南衡阳 421001)

摘要: 针对目前隐性知识本身的不确切性,本文首先对隐性知识的内涵进行了阐述,进而从如何提高隐性知识管理决策水平的角度,对粗糙集理论应用于隐性知识决策进行了探讨。

关键词: 粗糙集 隐性知识 显性知识 知识管理 决策

在管理领域,知识已经成为企业竞争力的源泉。哈佛大学的学者们认为,当前企业管理已经进入第六个阶段,即全球化和知识化的阶段。知识作为企业竞争的资源和管理的对象,其重要性日益突现。作为越来越依赖积累知识来成长的企业,知识管理特别是隐性知识管理受到越来越多的企业的关注。本文旨在对隐性知识形成、特性分析基础上,从如何提高知识管理决策水平的角度,探讨将粗糙集理论引入到隐性知识管理决策当中,有助于较好地解决隐性知识管理中存在的大量不确定性决策问题^[3]。

1 隐性知识管理的内涵

1.1 隐性知识的形成

隐性知识是指隐含在各知识主体手中和头脑中的,体现为技能、技巧、诀窍、经验、洞察力、心智模式、群体成员的默契等很难用文字、语言来精确表述,大部分不易编码或不能编码的知识。隐性知识的形成过程是一个学习积累与创新的过程,是不同知识主体间相互影响和相互作用的交互过程(Lundvall, 1992; OECD, 1992)。一定的主体从外部各类知识主体或知识库中经过筛选、吸收、积累自身企业所需要的知识(这里指广义的知识,包括科学技术、文化、价值观等),然后在适当的时机激活知识存量、进行提炼、创造出大量存储于主体脑海中的未成文初期隐性知识雏形,在研究、工作中经过反复提炼逐渐形成主体特有的隐性知识体系,这一过程是一个反复试错、论证的思考创造活动,对于外界来讲因信号不明显而具有不可观测性。并且它是企业核心竞争力的基础和源泉^[7]。

1.2 隐性知识的特性

为增强我们对识别、获取机制研究的针对性,有必要结合其产生机理分析一下隐性知识的特性。首先有必要对主体做进一步的解释,它不仅指单独的个人,也指一定的组织,很多情况下隐性知识在本质上更是一种集体共有性的知识(伦德威尔,1995),它与特定的人力资源和社会环境紧密相连。由隐性知识的产生及定义可以发现隐性知识的两大特性: (1) 非格式化。隐性知识以一种非格式化的、未编码的形式存在于人的脑海里或一定的组织形态中,尽管它客观存在,但由于其非结构化和表达、表现形式的模糊性使其难以系统化地为大家共享。它可以是个体对客观世界的心智感悟和判断,也可能是一种特有的文化、价值观,它只能被演示、领悟和练习是获取这类资源的唯一方式。(2) 路径依赖性。隐性知识的产生具有主体属性,这源于它在产生过程中对个体的性格、经历、价值观和组织文化、环境的依赖。隐性知识作为一种积累下来的带经验性质的知识,是个体和组织在工作、学习和科研过程中通过反复试错总结出来的具有规律性的行为倾向或习惯。这就决定了前期开展的学习、工作、研究对隐性知识的形成、内容上的决定作用。而作为隐性知识形成基石的文化、环境本身的营造、构建过程也是一种具有历史积累性质的活动,这就更加强了路径依赖的特征。

2 隐性知识管理决策及其不确定性分析

在知识管理决策过程中,由于隐性知识的产生很大程度上是建立在历史数据、观察、实体体验上的,主

体基于自己喜好、目的、个性会把当时场景中的任何可以得到的能有所帮助的隐性知识元组合起来,内容上缺乏深层次的理性归纳和验证,知识搜索、发展方向上也不一定与目标组织趋同,这些直接涉及到管理者的心理因素、经验和决策水平等,因此可能给决策带来风险。

不确定和不可分辨的特性,传统的管理决策方法都是通过采用纯数学上的假设来消除或回避这种不确定性,这使得其在解决具有不可分辨、不确定关系的决策问题时表现出了很大的局限性。但如果正视它,对这些信息、知识进行合适的处理、常常有助于相关实际问题的解决。多年来,研究人员一直在努力寻找科学地处理不完整性和不确定性的有效途径,模糊集和基于概率方法的证据理论是处理不确定信息的两种方法,已应用于一些实际领域,但这些方法有时需要一些数据的附加信息或先验知识,如统计学需要概率分布,模糊集理论需要隶属函数,而粗糙集理论的主要优势就在于它不需要关于数据的任何预备的或额外的信息。波兰学者 Z. Pawlak 提出的粗糙集理论,是一种新的处理含糊性和不确定性问题的数学工具,能有效地分析不精确、不一致、不完整等各种不完备的信息,有效地处理不完整和不确定性的知识管理决策问题。

3 基于粗糙集理论的隐性知识决策

3.1 粗糙集理论的引入

1982 年, Z. Pawlak 发表了经典论文 *Roughest*, 宣告了粗糙集理论 (RS) 的诞生。粗糙集理论的模糊性就是一种基于边界的概念, 即一个模糊的概念具有模糊的不可被明确划分的边界, 在没有掌握所有关于对象域知识的情况下, 我们只能用以逼近来描述对象域上的集合。粗糙集理论已在知识发现、数据挖掘、智能决策、电子控制等许多领域得到了广泛的应用, 下面介绍几个主要概念。

(1) 粗糙集和上下近似集

任一 $X \subseteq U$ 能表达成某些 R 基本范畴的并时, 称 X 为 R 可定义的, 否则, X 为 R 不可定义的, 称作非精确集或粗糙集。实际上, 基本范畴就是 RS 理论用来表达一个特定的概念的基本知识模块; 精确集是可以由已知知识确定地表达的范畴; 粗糙集则只能通过两个精确范畴“上、下近似集”粗略地逼近。称 $POS_R(X) = \underline{R}X$ 为

X 的 R 正域 (下近似); 称 $NEG_R(X) = U - \underline{R}X$ 为 X 的 R 负域 (上近似), 称集合 $BN_R(X) = \underline{R}X - \overline{R}X$ 为 X 的 R 边界域; 当 $BN_R(X) = \Phi$ 时, 即 $\underline{R}X = \overline{R}X$, 称 X 是 R 精确集; 当 $BN_R(X) \neq \Phi$ 时, 即 $\underline{R}X \neq \overline{R}X$, 称 X 是 R 粗糙集^[1]。

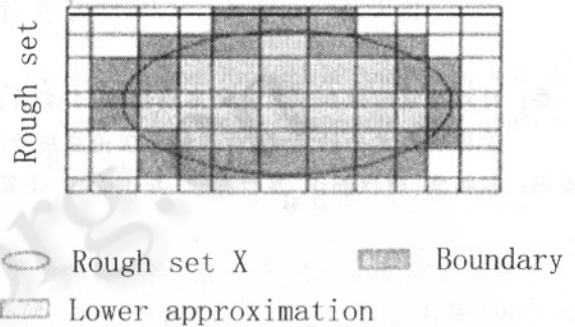


图 1 粗糙集示意图

(2) 近似精度

由等价关系 R 定义的集合 X 的近似精度为:

$$\alpha_R(X) = \frac{|RX|}{|\overline{R}X|}, \text{ 其中 } X \neq \Phi, |X| \text{ 表示集合 } X \text{ 的基数, 下}$$

近似占上近似的比例就是近似精度, 显然有 $0 \leq \alpha_R(X) \leq 1$ 。当 $\alpha_R(X) = 1$ 时, X 是 R 的精确集; 当 $0 \leq \alpha_R(X) < 1$ 时, X 是 R 的粗糙集^[2]。给定两个集合, 我们可以量度它们的距离, 则 $1 - \alpha_R(X)$ 定义为 ROUGH 下和上近似集之间的距离^[9]。

(3) 信息表

信息系统是一种知识表达方式, 知识的表达方式在智能数据处理中有十分重要的地位。信息系统有时也称为知识表示系统。形式上, 四元组 $S = (U, A, V, f)$ 是一个信息系统, 其中: U : 对象的非空有限集合, 即论域; A : 属性的非空有限集合; $V = \bigcup_{a \in A} V_a$, V_a 是属性的值域; $f: U \times A \rightarrow V$ 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 $\forall a \in A, x \in U, f(x, a) \in V_a$ 。

在粗糙集中, 表示集合不精确性的数值是通过现有知识中的两个精确集合定义的。产生不精确性的原因在于我们对论域的现有知识有限, 随着知识粒度的细化, 不精确性会随之降低。

3.2 实例分析

传统的知识决策模型的一个局限性是它所处理的类别归属必须是完全肯定的, 只有“包含”或“不包

含", 缺乏一定的柔性; 而没有某种程度上的"包含"或"属于"; 这样, 将对公司发展有较大影响某些指标和有较小影响的指标都同等对待地归入边界域, 就不合理; 另外, 传统模型当出现噪声数据时, 抗干扰能力差, 对新对象的预测能力降低。新的决策模型一定程度上补偿了原始定义的生硬, 柔化了边界, 使决策管理更准确, 下面以某区域科技社会经济协调发展决策系统为例进行说明, 其它决策模型类似。

设 $S = \langle U, E, V, f \rangle$ 是一区域科技、社会经济协调发展决策系统, 其中:

$U = \{u_1, u_2, \dots, u_n\}$ 是发展事业的集合, 如经济发展事业, 科技发展事业, 社会发展事业, ……;

$E = \{e_1, e_2, \dots, e_n\}$ 是事业发展指标实体, 如重工业, 轻工业, ……是经济事业的指标实体; 又如, 计算机技术水平, 医药工程水平, 现代通信技术水平, ……是科技事业的指标实体; 再加社会福利, 人口素质, 社会风尚, 环境质量, ……是社会发展事业的指标实体; 如此等等;

$V = \{ \langle e_1: \text{指标实体 } 1, [(1-d/2)\alpha_1, (1+d/2)\alpha_1] \rangle,$

$\langle e_2: \text{指标实体 } 2, [(1-d/2)\alpha_2, (1+d/2)\alpha_2] \rangle,$

$\langle e_i: \text{指标实体 } i, [(1-d/2)\alpha_i, (1+d/2)\alpha_i] \rangle,$

其中指标实体 e_i 是反映第 i 种发展事业水平的指标之一, 如农业综合技术水平是反映科技事业发展的指标之一。 α_i 是对第 i 种发展事业的发展水平的影响率, 它是一个相对近似数, 即同一项指标实体在不同的单位或部门对同一种发展事业的发展水平的影响率是不同的, 所以专家对影响率 α_i 的评估或经验值, 显然都不尽相同。如此, 用 Rough 上下近似集的距离函数 d 与 α 联合起来做求精运算得到 Rough 数, 即区间值 $[(1-d/2)\alpha, (1+d/2)\alpha]$, 以此区间值表示影响率自然更为科学、合理, 因为它是一个区间, 多值而不局限于单个值。

如果 R 被定义为指标实体集 E 上的等价关系, 也就是说反映同一发展事业的那些指标实体 e_{k_1} 和 e_{k_2} 有 R 关系, 记成 $e_{k_1}(\alpha_{k_1}) R e_{k_2}(\alpha_{k_2})$ 。这样, 在 E 上可按如此定义的等价关系 R 划分成若干个互不相交的等价类 $E_k = \{e_{k_1}(\alpha_{k_1}), e_{k_2}(\alpha_{k_2}), \dots, e_{k_l}(\alpha_{k_l})\}$ 。

f 是赋值函数, 即对发展事业影响其发展水平的指标实体赋予近似 Rough 数, 一个区间值: $[(1-d/2)\alpha, (1+d/2)\alpha]$, 它是 $U \times E \rightarrow V$ 的一个映射, 即 $f: U \times E \rightarrow V$ 。

如果 $u_i \in U$ 是第 i 种发展事业, $e_{k_l}(\alpha_{k_l}) \in E_k = \{e_{k_1}(\alpha_{k_1}), e_{k_2}(\alpha_{k_2}), \dots, e_{k_l}(\alpha_{k_l})\}$, 则 $f(u_i, e_{k_l}(\alpha_{k_l})) = \{\alpha_{j_{k_l}}\} \subseteq [(1-d/2)\alpha_{j_{k_l}}, (1+d/2)\alpha_{j_{k_l}}]$ 。

现假设在上述系统中, 所讨论的个体域是 $E = \{e_1(\alpha_1) \text{——重工业发展水平}; e_2(\alpha_2) \text{——轻工业发展水平}; e_3(\alpha_3) \text{——农业发展水平}; e_4(\alpha_4) \text{——商业发展水平}; e_5(\alpha_5) \text{——建筑业发展水平}; e_6(\alpha_6) \text{——邮电电业发展水平}; e_7(\alpha_7) \text{——社会结构水平}; e_8(\alpha_8) \text{——人口素质水平}; e_9(\alpha_9) \text{——生活质量水平}; e_{10}(\alpha_{10}) \text{——社会风尚水平}; e_{11}(\alpha_{11}) \text{——环境质量水平}; e_{12}(\alpha_{12}) \text{——计算机技术水平}; e_{13}(\alpha_{13}) \text{——食品工业水平}; e_{14}(\alpha_{14}) \text{——医药工业水平}; e_{15}(\alpha_{15}) \text{——现代通信技术水平}; e_{16}(\alpha_{16}) \text{——环境工程水平}; e_{17}(\alpha_{17}) \text{——生活工程水平}; e_{18}(\alpha_{18}) \text{——农业综合技术水平}\}$ 上被定义的二元关系 R , 意味着反映同一发展事业发展水平的两个指标实体随同它们的影响率都相同, 则它们有关系。显然如此定义的 R 是一个等价关系, 于是 E 按如此的 R 可划分成等价类:

$E_1 = \{e_1(\alpha_1), e_2(\alpha_2), e_3(\alpha_3), e_4(\alpha_4), e_5(\alpha_5), e_6(\alpha_6)\}$, 影响经济发展事业发展水平的指标实体集;

$E_2 = \{e_7(\alpha_7), e_8(\alpha_8), e_9(\alpha_9), e_{10}(\alpha_{10}), e_{11}(\alpha_{11})\}$, 影响社会发展事业发展水平的指标实体集;

$E_3 = \{e_{12}(\alpha_{12}), e_{13}(\alpha_{13}), e_{14}(\alpha_{14}), e_{15}(\alpha_{15}), e_{16}(\alpha_{16}), e_{17}(\alpha_{17}), e_{18}(\alpha_{18})\}$, 影响科技发展事业发展水平的指标实体集。

任取一子集 $X = \{e_1(\alpha_1), e_7(\alpha_7), e_{12}(\alpha_{12}), e_{13}(\alpha_{13}), e_{14}(\alpha_{14}), e_{15}(\alpha_{15}), e_{16}(\alpha_{16}), e_{17}(\alpha_{17}), e_{18}(\alpha_{18})\} \subseteq E$, 则

$R^*(X) = E_3 = \{e_{12}(\alpha_{12}), e_{13}(\alpha_{13}), e_{14}(\alpha_{14}), e_{15}(\alpha_{15}), e_{16}(\alpha_{16}), e_{17}(\alpha_{17}), e_{18}(\alpha_{18})\}$

$R^*(X) = E_1 \cup E_2 \cup E_3 = E$

所以 X 在近似空间 $A = (E, R)$ 上的近似精度为

$\mu_A(X) = K(R^*(X)) / K(R^*(X)) = 7/18$

而 Rough 下和上近似集 $R_*(X)$ 和 $R^*(X)$ 的距离函数为

$d(R_*(X), R^*(X)) = 1 - \mu_A(X) = 1 - 7/18 = 11/18$

设 u_3 是社会发展事业, $e_{13}(\alpha_{13})$ 是食品工业水平指标实体, 它对 u_3 发展事业的影响率可通过函数 f 求

得: $f = (u_2, e_{13}(\alpha_{13})) = [(1 - 11/18/2) \alpha_{13}, (1 + 11/18/2) \alpha_{13}]$

如果, $e_{13}(\alpha'_{13})$ 出现在影响科技发展事业 u_2 发展水平的指标实体 E_2 中, 则它对 u_2 发展水平的影响率同样可以通过函数 f 求得:

$$f = (u_2, e_{13}(\alpha'_{13})) = [(1 - 11/18/2) \alpha'_{13}, (1 + 11/18/2) \alpha'_{13}]$$

此处 $\alpha'_{13} \neq \alpha_{13}$

实践证明, 用这种 Rough 集方法求得的区间值作为影响率比通过专家打分作为单值影响率要好得多, 因为它更切合实际且富于理论。表 1 是各项指标实体对科技发展影响乘数 (或称影响率或称判断系数) 矩阵。其中 SO 表示社会发展, “:” 表示“比”, 其左边数字表示左竖列上指标实体对 SO 的影响率, 右边数字是顶横行上指标实体对 SO 的影响率。

表 1 各指标实体对社会发展影响率矩阵

SO	计算机 技术	食品 工程	医药 工程	通信 工程	环境 工程	生物 工程	农业 工程
计算机技术	1:1	1:3	1:5	1:1	1:3	5:1	7:1
食品工程	3:1	1:1	1:3	3:1	3:1	1:1	3:1
医药工程	5:1	3:1	1:1	3:1	3:1	3:1	3:1
通信工程	1:1	1:3	1:3	1:1	5:1	3:1	3:1
环境工程	3:1	1:3	1:3	1:5	1:1	1:1	1:1
生物工程	1:5	1:1	1:3	1:3	1:1	1:1	1:3
农业工程	1:7	1:3	1:3	1:3	1:1	1:3	1:1

他们都是由专家打分给出的, 显然是凭经验, 无理论依据, 为此通过 Rough 上下近似集, 并以此求得其距离函数, 将它与打分得到的影响率作进一步的求精运算, 产生影响率的区间值。左边数字产生的区间值为上行, 右边数字产生的区间值为下行, 于是得到表 2。对表 1 中的影响率作求精运算, 产生区间值, 参与运算的参数 $d = 11/18$ 。

4 结束语

由于隐性知识本身具有很大的不确定和不可分辨的特性, 这就导致人们进行知识管理决策的难度加大, 将粗糙集理论引入到知识管理决策中, 将使管理决策理论进一步得到丰富和完善, 并有助于较好地解决隐

表 2 影响率为区间值的判断系数矩阵

SO	计算机 技术	食品 工程	医药 工程	通信 工程	环境 工程	生物 工程	农业 工程
计算机 技术	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(125/36, 235/36)	(175/36, 329/36)
食品 工程	(25/36, 47/36)	(25/12, 47/12)	(125/36, 235/36)	(25/36, 47/36)	(25/12, 47/12)	(25/36, 47/36)	(25/36, 47/36)
医药 工程	(25/12, 47/12)	(25/36, 47/36)	(25/36, 47/36)	(25/12, 47/12)	(25/12, 47/12)	(25/36, 47/36)	(25/12, 47/12)
通信 工程	(125/36, 235/36)	(25/12, 47/12)	(25/36, 47/36)	(25/12, 47/12)	(25/12, 47/12)	(25/12, 47/12)	(25/12, 47/12)
环境 工程	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)
生物 工程	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)	(25/36, 47/36)
农业 工程	(175/36, 329/36)	(25/12, 47/12)	(25/12, 47/12)	(25/12, 47/12)	(25/36, 47/36)	(25/12, 47/12)	(25/36, 47/36)

性知识管理中存在的大量的不确定性决策问题。本文通过粗糙集基本理论的介绍, 并将其应用于某区域科技社会经济协调发展决策系统之中。通过在不损失原有信息分类的情况下, 从科技、经济协调发展决策系统中求得的不精确值能落入相应区间, 则认为某发展事业水平所达到的相应指标实体是切合实际、符合科学发展规律, 在信息系统规模不大的情况下, 采用本文所提出的决策方法可以获得满意的结果。

参考文献

- 1 Pawlak, Z. Busse, J. G. Roman Slowinski, et al. Rough Sets [J]. Communication of the ACM, 1995, 38(11): 89 - 95.
- 2 张文修等, 粗糙集理论与方法 [M], 科学出版社, 2001.
- 3 Johannessen J. A. Aspects of Innovation Theory Based on Knowledge - Management [J]. International Journal of Information Management, 1999, 19(2): 121 - 139.

(下转第 100 页)

(上接第 80 页)

- 4 王众托, 知识系统工程[M], 北京: 科学出版社, 2004. 9.
- 5 郁义鸿, 知识管理与组织创新[M], 上海: 复旦大学出版社, 2001. 39.
- 6 Rubinstein A. A Theorists View of Experiments [J], European Economic Review, 2001, (4): 615 - 628.
- 7 巢乃鹏, 知识管理——概念、特性的分析[J], 学术界, 2000, (5): 14.
- 8 彼得·F·德鲁克等著, 知识管理[M], 赵云飞译, 北京: 中国人民大学出版社, 哈佛商学院出版社, 1999. 18; 25; 32.
- 9 刘清著, ROUGH 集及 ROUGH 推理[M]. 北京: 科学出版社, 2001, 8(1), 11 - 15.
- 10 胡可云、陆玉昌、石纯一, 粗糙集理论及其应用进展[J], 清华大学学报(自然科学版), 2001, 41(1): 64 - 68.