第 24 卷 第 6 期 控 制 与 决 策 2009 年 6 月
Vol. 24 No. 6 Control and Decision Jun. 2009

文章编号: 1001-0920(2009)06-0899-04

优势模糊粗糙模型及其在审计风险评估中的应用

黄 兵[1], 胡作进[1], 周献中[2]
(1. 南京审计学院 信息科学学院, 南京 210029; 2. 南京大学 工程管理学院, 南京 210093)

摘 要: 针对模糊信息系统,建立基于优势关系的粗糙集模型,并给出该模型的多种知识约简定义. 通过构造分辨函数,得到了求解优势模糊粗糙集模型的最大分布约简算法. 最后将该模型应用于审计风险评估,得到了较为合理的评估规则.

关键词: 优势关系; 模糊信息系统; 粗糙集; 约简; 审计风险评估
中图分类号: TP18 **文献标识码**: A

Dominance relation-based fuzzy-rough model and its application to audit risk evaluation

HUANG Bing[1], *HU Zuo-jin*[1], *ZHOU Xian-zhong*[2]
(1. School of Information Science, Nanjing Audit University, Nanjing 210029, China; 2. School of Engineering and Management, Nanjing University, Nanjing 210093, China. Correspondent: HUANG Bing, E-mail: hbhuangbing@126.com)

Abstract: For fuzzy information systems, a dominance relation-based fuzzy-rough model is constructed and several knowledge reduction definitions are proposed. By using discernibility function, a maximum distribution reduction algorithm is devised and applied to audit risk evaluation. Finally, some useful evaluation rules are obtained.

Key words: Dominance relation; Fuzzy information systems; Rough sets; Reduction; Audit risk evaluation

1 引 言

粗糙集理论是一门处理不精确、不确定信息的数学理论. 经过 20 多年的发展,该理论已取得了长足进步,并成功地应用于智能信息处理、模式识别、智能控制和数据挖掘等领域. 粗糙集理论的研究对象是一个二维决策表,称为信息系统. 当信息系统在条件属性和决策属性上的取值均为模糊值时,称为模糊信息系统(FIS).

知识约简是粗糙集理论的精髓. 针对模糊信息系统的知识约简,一般有两种处理方法:其一是文献[1]首次提出的模糊粗糙模型. 利用属性确定的模糊集定义上下近似,通过定义任一模糊集的上下近似和相对正域,借鉴符号信息系统知识约简保依赖度(精度)不变的思想,得到模糊信息系统的知识约简标准,并以精度变化大小为启发信息,给出寻求最小约简的启发式算法[2-4]. 文献[5]深入研究了这种约简算法,发现该算法并不收敛,并提出了新的算法终止条件. 作者在文献[6]中给出了模糊信息系统的多种知识约简语义定义. 这种模糊粗糙集方法的不足之处是没能完全反映粗糙集理论中对象间的不可区分性的重要思想. 另一种方法是利用对象取模糊值的差异定义它们的相似程度,并确定每个对象的邻域. 这种方法的基本思路是去模糊化. 如文献[7,8]通过确定隶属度函数将连续值信息系统转化为模糊信息系统,并根据对象间的相似度给出上下近似和知识约简概念,进而验证属性的必要性并得到知识约简方法. 文献[9]提出了一种新的模糊粗糙集模型,并提出了通过矩阵运算求 FIS 约简的方法.

模糊信息系统中的模糊值在同一属性下的取值大小本身具有一种序的关系. 本文试图结合处理模糊信息系统的两种粗糙集方法,建立基于优势关系的模糊信息系统粗糙集模型,并将该模型应用于审

收稿日期: 2008-06-10; **修回日期**: 2008-10-19.
基金项目: 国家自然科学基金项目(70571032); 教育部人文社科项目(08JC630045); 江苏省高校自然科学基金项目(07KJD120087).
作者简介: 黄兵(1972—),男,四川绵阳人,副教授,博士后,从事粗糙集理论与应用的研究; 周献中(1962—),男,江苏泰兴人,教授,博士生导师,从事智能决策等研究.

计风险评估.

2　基本概念

定义 1　设 FIS $=(U, C, D)$. 其中:U 是论域;C 和 D 分别是条件和决策属性集合,且 $C \cap D=\Phi$; $\forall x \in U, c \in C, d \in D; c(x), d(x) \in[0,1]$.

在 FIS 中,属性集合对论域的划分是通过其中的每个属性确定的模糊划分求交获得的.

定义 2[1]　设由属性子集 $P \subseteq A$ 确定的模糊划分为 $U/P=\{F_1, F_2, \cdots, F_n\}$, $F_i \in U/P$ 相对于论域中任一模糊集 X 的上、下近似 $\mu_{\overline{P}X}(F_i)$ 和 $\mu_{\underline{P}X}(F_i)$ 定义为

$$\mu_{\underline{P}X}(F_i)=\inf_{x \in U} \max\{1-\mu_{F_i}(x), \mu_X(x)\}, \quad 1 \leq i \leq n;$$

$$\mu_{\overline{P}X}(F_i)=\sup_{x \in U} \min\{\mu_{F_i}(x), \mu_X(x)\}, \quad 1 \leq i \leq n.$$

定义 3[1]　论域中对象 x 关于属性集 P 对模糊集 X 的下近似 $\mu_{\underline{P}X}(x)$ 和上近似 $\mu_{\overline{P}X}(x)$ 定义为

$$\mu_{\underline{P}X}(x)=\sup_{F_i \in U/P} \min\{\mu_{F_i(x)}, \mu_{\underline{P}X}(F_i)\},$$

$$\mu_{\overline{P}X}(x)=\sup_{F_i \in U/P} \min\{\mu_{F_i(x)}, \mu_{\overline{P}X}(F_i)\}.$$

定义 4[2-4]　属性集 Q 相对于属性集 P 的相对正域定义为

$$\mu_{\mathrm{POS}_P(Q)}(x)=\sup_{X \in U/Q} \mu_{\underline{P}X}(x).$$

3　优势模糊粗糙模型和知识约简

在 FIS 中,对象在条件属性确定的模糊等价类上取值的大小关系是一种典型的优势关系. 下面给出基于优势关系的模糊粗糙集模型.

定义 5　设 FIS $=(U, C, D)$, $P \subseteq C \cup D$ 确定的模糊划分为 $U/P=\{P_1, P_2, \cdots, P_n\}$. 记 $\mathrm{DR}_P^{P_i}(x)=\{y \in U \mid p_{P_i}(x) \leq p_{P_i}(y)\}$. 其中:$P_i=\bigcap_{p \in P} p_{P_i}$, $p_{P_i}(x)$ 表示对象 x 在模糊等价类 P_i 对应的属性 p 下的取值. 则对象 x 在 $P_i \in U/P$ 下相对于论域中任一模糊集 X 的优势上、下近似 $\mu_{\overline{P_i}X}^{\mathrm{DR}}(x)$ 和 $\mu_{\underline{P_i}X}^{\mathrm{DR}}(x)$ 定义为

$$\mu_{\underline{P_i}X}^{\mathrm{DR}}(x)=\inf_{y \in \mathrm{DR}_P^{P_i}(x)} \max\{1-\mu_{P_i}(y), \mu_X(y)\}, \quad 1 \leq i \leq n;$$

$$\mu_{\overline{P_i}X}^{\mathrm{DR}}(x)=\sup_{y \in \mathrm{DR}_P^{P_i}(x)} \min\{\mu_{P_i}(y), \mu_X(y)\}, \quad 1 \leq i \leq n.$$

与定义 2 不同,定义 5 在计算模糊集相对于模糊等价类的上下近似时,只考虑了优于该对象的对象,而不是所有对象.

定义 6　设 FIS $=(U, C, D)$, $P \subseteq C \cup D$ 确定的模糊划分为 $U/P=\{P_1, P_2, \cdots, P_n\}$. 论域中对象 x 关于属性集 P 对模糊集 X 的优势上、下近似 $\mu_{\overline{P}X}^{\mathrm{DR}}(x)$ 和 $\mu_{\underline{P}X}^{\mathrm{DR}}(x)$ 定义为

$$\mu_{\underline{P}X}^{\mathrm{DR}}(x)=\sup_{P_i \in U/P} \min\{\mu_{P_i}(x), \mu_{\underline{P_i}X}^{\mathrm{DR}}(x)\},$$

$$\mu_{\overline{P}X}^{\mathrm{DR}}(x)=\sup_{P_i \in U/P} \min\{\mu_{P_i}(x), \mu_{\overline{P_i}X}^{\mathrm{DR}}(x)\}.$$

定义 7　属性集 Q 相对于属性集 P 的相对正域定义为

$$\mu_{\mathrm{POS}_P(Q)}^{\mathrm{DR}}(x)=\sup_{Q_i \in U/Q} \mu_{\underline{P}X}^{\mathrm{DR}}(x).$$

记 $\gamma_{\mathrm{POS}_C(D)}^{\mathrm{DR}}(x)=\{D_i \in U/D \mid \mu_{\mathrm{POS}_C(D)}^{\mathrm{DR}}(x)=\mu_{\underline{C}D_i}^{\mathrm{DR}}(x)\}$. 下面给出 FIS 几种知识约简定义.

定义 8　设 FIS $=(U, C, D)$, $U/D=\{D_1, D_2, \cdots, D_m\}$, $B \subseteq C$.

1) 若 $\forall x \in U$, $\mu_{\underline{C}D_i}^{\mathrm{DR}}(x)=\mu_{\underline{B}D_i}^{\mathrm{DR}}(x)$ $(\mu_{\overline{C}D_i}^{\mathrm{DR}}(x)=\mu_{\overline{B}D_i}^{\mathrm{DR}}(x), \mu_{\underline{C}D_i}^{\mathrm{DR}}(x)>0 \Leftrightarrow \mu_{\underline{B}D_i}^{\mathrm{DR}}(x)>0, \mu_{\overline{C}D_i}^{\mathrm{DR}}(x)>0=\mu_{\overline{B}D_i}^{\mathrm{DR}}(x)>0, 1 \leq i \leq m)$,则称 B 为 FIS 的优势模糊下分布(上分布、下分配、上分配) 协调集;若 B 为 FIS 的优势模糊下分布(上分布、下分配、上分配) 协调集而其任意真子集都不是优势模糊下分布(上分布、下分配、上分配) 协调集,则称 B 为 FIS 的优势模糊下分布(上分布、下分配、上分配) 约简.

2) 若 $\forall x \in U$, $\gamma_{\mathrm{POS}_C(D)}^{\mathrm{DR}}(x)=\gamma_{\mathrm{POS}_B(D)}^{\mathrm{DR}}(x)$,则称 B 为 FIS 的优势模糊最大分布协调集;若 B 为 FIS 的优势模糊最大分布协调集而其任意真子集都不是优势模糊最大分布协调集,则称 B 为 FIS 的优势模糊最大分布约简.

4　优势模糊粗糙模型知识约简方法

下面讨论优势模糊粗糙模型的最大分布约简方法.

定义 9　设 FIS $=(U, C, D)$,条件属性集确定的模糊类 $U/C=\{C_1, C_2, \cdots, C_m\}$,决策属性集确定的模糊类 $U/D=\{D_1, D_2, \cdots, D_s\}$. $\forall x \in U$,设

$$\begin{aligned}
&\mu_{\mathrm{POS}_C(D)}^{\mathrm{DR}}(x)=\sup_{D_j \in U/D} \mu_{\underline{C}D}^{\mathrm{DR}}(x)= \\
&\mu_{\underline{C_{i_1 j_1}}D_{j_1}}^{\mathrm{DR}}(x)=\mu_{\underline{C_{i_2 j_1}}D_{j_1}}^{\mathrm{DR}}(x)=\cdots=\mu_{\underline{C_{i_{j1} j_1}}D_{j_1}}^{\mathrm{DR}}(x)= \\
&\mu_{\underline{C_{i_1 j_2}}D_{j_2}}^{\mathrm{DR}}(x)=\mu_{\underline{C_{i_2 j_2}}D_{j_2}}^{\mathrm{DR}}(x)=\cdots=\mu_{\underline{C_{i_{j2} j_2}}D_{j_2}}^{\mathrm{DR}}(x)= \\
&\cdots= \\
&\mu_{\underline{C_{i_1 j_k}}D_{j_k}}^{\mathrm{DR}}(x)=\mu_{\underline{C_{i_2 j_k}}D_{j_k}}^{\mathrm{DR}}(x)=\cdots=\mu_{\underline{C_{i_{jk} j_k}}D_{j_k}}^{\mathrm{DR}}(x),
\end{aligned}$$

其中 $\mu_{\underline{C_{i_w j_v}}D_{j_v}}^{\mathrm{DR}}(x)$ 表示对象 x 在条件属性确定的模糊等价类 C_{i_w} 和决策属性的模糊类 D_{j_v} 上取得下近似 $\mu_{\mathrm{POS}_C(D)}^{\mathrm{DR}}(x)$. 则 x 的优势模糊最大分布分辨函数构造如下:

$$f(x)=\bigwedge_{1 \leq v \leq k}\left(C_{i_1 j_v} \bigvee C_{i_2 j_v} \bigvee \cdots \bigvee C_{i_{j_v} j_v}\right).$$

整个 FIS 的优势模糊最大分布分辨函数为

$$f(U) = \bigwedge_{x \in U} f(x).$$

将 $f(U)$ 中的合取范式转化为析取范式,每个析取范式中的水平便构成一个约简结果.

因为对象可能在多个决策属性确定的模糊类上取得下近似,而每个这样的模糊类又可能对应多个条件属性确定的模糊类,根据优势模糊最大分布约简定义,这些条件属性确定的所有模糊类只需保留其中之一,就能保证该对象的最大决策之一,故在 $f(x)$ 表达式中每一项应采用"析取";同时要保证每个对象的所有最大决策,$f(x)$ 中项与项之间应采用"合取".对于后者,若是求下近似约简,则只需保留一个这样的条件属性确定的模糊类即可,即应采用"析取".故有结论:FIS 的优势模糊最大分布约简分辨函数应保证所有对象的优势模糊最大分布不变.

5　审计风险评估算例

风险导向审计模式下,会计师事务所或注册会计师是否承接业务,承接后采取何种审计策略,实施何种审计程序,何时实施以及在多大范围内实施,完全取决于对审计风险的评估.因此,风险导向审计的核心是评估审计风险.目前,审计风险评估方法主要有:风险因素分析法、分析性审核法、定性风险评价法、模糊综合评价法、层次分析法和风险率评价法等[10].

2004 年 12 月开始实施的新的国际审计风险准则中,将审计风险模型修改为

审计风险 = 重大错报风险 ×检查风险.

注册会计师在承接审计业务时,会事先进行调查并凭借经验确定重大错报和检查风险率(往往得到的是一个模糊值),进而计算出审计风险值.由于审计前所获取的重大错报和检查风险率数值具有很大的主观性,通过审计风险模型计算出的实际审计风险值就不太可靠.即事先估计的重大错报风险率和检查风险率与实际审计风险率之间是一种不确定的关系.

假设有 9 个已完成的审计案例.用 a 表示重大错报风险,有 3 个水平:高(A_1),中(A_2)和低(A_3);b 表示检查风险,有两个水平:高(B_1)和低(B_2);d 表示审计风险,有3个水平:高(X),中(Y)和低(Z).其中:a 和 b 的取值是事先估计值,d 的取值是审计结束后最终确定的审计风险值(见表 1).

表 1 中,$U = \{x_0, x_1, x_2, \cdots, x_8\}$,$C = \{a, b\}$,$D = \{d\}$,$U/C = \{A_i \cap B_j \mid i \le 3, j \le 2\}$,$U/D = \{X, Y, Z\}$.为简便起见,记 $A_i \cap B_j = A_iB_j = P_{ij}$(见表 2).

Step1:计算每个对象的优势模糊下近似

表 1　审计风险评估表

对象	a			b		d		
	A_1	A_2	A_3	B_1	B_2	X	Y	Z
x_0	0.3	0.7	0.0	0.7	0.3	0.9	0.1	0.0
x_1	0.0	1.0	0.0	0.7	0.3	0.8	0.2	0.0
x_2	0.0	0.7	0.3	0.6	0.4	0.7	0.3	0.0
x_3	0.8	0.2	0.0	0.2	0.8	0.6	0.3	0.1
x_4	0.5	0.5	0.0	0.0	1.0	0.6	0.8	0.0
x_5	0.0	0.2	0.8	0.0	1.0	0.0	0.3	0.7
x_6	1.0	0.0	0.0	0.2	0.8	0.7	0.4	0.0
x_7	0.1	0.8	0.1	0.7	0.3	0.6	0.4	0.0
x_8	0.3	0.7	0.0	1.0	0.0	0.8	0.3	0.0

表 2　表 1 条件属性求交后的结果

对象	$a \otimes b$					
	$A_1 \cap B_1$	$A_1 \cap B_2$	$A_2 \cap B_1$	$A_2 \cap B_2$	$A_3 \cap B_1$	$A_3 \cap B_2$
x_0	0.3	0.3	0.7	0.3	0.0	0.0
x_1	0.0	0.0	0.7	0.3	0.0	0.0
x_2	0.0	0.0	0.6	0.4	0.3	0.3
x_3	0.2	0.8	0.2	0.2	0.0	0.0
x_4	0.0	0.5	0.0	0.5	0.0	0.0
x_5	0.0	0.0	0.0	0.2	0.0	0.8
x_6	0.2	0.8	0.0	0.0	0.0	0.0
x_7	0.1	0.1	0.7	0.3	0.1	0.1
x_8	0.3	0.0	0.7	0.0	0.0	0.0

$$\mu^{DR}_{POS_C(D)}(x_1) = \mu^{DR}_{POS_C(D)}(x_5) = \mu^{DR}_{POS_C(D)}(x_6) = 0.7,$$
$$\mu^{DR}_{POS_C(D)}(x_0) = \mu^{DR}_{POS_C(D)}(x_2) = \mu^{DR}_{POS_C(D)}(x_3) = 0.6,$$
$$\mu^{DR}_{POS_C(D)}(x_4) = 0.5,$$
$$\mu^{DR}_{POS_C(D)}(x_7) = \mu^{DR}_{POS_C(D)}(x_8) = 0.6.$$

Step2:构造对象的分辨函数

$$f(x_0) = f(x_1) = f(x_2) = f(x_7) = f(x_8) = P_{21},$$
$$f(x_3) = f(x_4) = f(x_6) = P_{12},$$
$$f(x_5) = P_{32}.$$

Step3:计算 FIS 的分辨函数

$$f(U) = P_{21} \wedge P_{12} \wedge P_{32}.$$

故有唯一优势模糊最大分布约简$\{P_{12}, P_{21}, P_{32}\}$.

因为条件属性构成 6 个水平,决策属性构成 3 个水平,计算的花费主要在确定每个对象的下近似在哪些条件属性水平上取得,故算法时间复杂度为 $18n^2$(n 为对象个数).

进一步可得如下规则:

If $x \in P_{12}$ 的程度不低于 0.5 或 $x \in P_{21}$ 的程度

不低于 0.6,

Then $x \in X$ 的程度不低于 0.6.

即当重大错报风险和检查风险都是“高”的程度至少为0.6 时,审计风险“高”的程度至少是 0.6.

If $x \in P_{32}$ 的程度不低于 0.8,

Then $x \in Z$ 的程度不低于 0.7.

即当大错报风险和检查风险都是“低”的程度至少为 0.8 时,审计风险“低”的程度至少是 0.7. 等等.

6 结 论

针对模糊信息系统,本文建立了基于优势关系的粗糙集模型并提出了知识约简的多种定义;将分辨函数引入模糊信息系统优势模糊最大分布约简,定义了对象及系统的分辨函数. 这种方法可以类推到本文定义的其他约简. 最后将优势模糊粗糙集模型应用于审计风险评估,得到了较为可信的评估规则.

参考文献(References)

[1] Dubois D, Prade H. Putting rough sets and fuzzy sets together[C]. Intelligent Decision Support. Dordrecht: Kluwer Academic Publishers, 1992: 203-232.

[2] Richard Jensen, Qiang Shen. Fuzzy-rough attribute reduction with application to web categorization [J]. Fuzzy Sets and Systems, 2004, 141(3): 469-485.

[3] Qiang Shen, Richard Jensen. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring[J]. Pattern Recognition, 2004, 37(7): 1351-1363.

[4] Richard Jensen, Qiang Shen. Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches [J]. Knowledge and Data Engineering, 2004, 16(12): 1457-1471.

[5] Rajen B, Bhatt M Gopal. On fuzzy-rough sets approach to feature selection[J]. Pattern Recognition Letters, 2005, 26(7): 965-975.

[6] Huang B, Zhou X Z, Jiang X Y. Knowledge reductions in fuzzy information systems[C]. Proc of 2006 Int Conf on Machine Learning and Cybernetics. Dalian, 2006, 5: 4169-4172.

[7] Hong T P, Wang T T, Wang S L. Mining fuzzy β certain and β possible rules from quantitative data based on the variable precision rough-set model[J]. Expert Systems with Applications, 2007, 32(1): 223-232.

[8] Hu Q, Xie Z, Yu D. Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation[J]. Pattern Recognition, 2007, 40(12): 3509-3521.

[9] Wang X, Tsang E C C, Zhao S, et al. Learning fuzzy rules from fuzzy samples based on rough set technique [J]. Information Sciences, 2007, 177(20): 4493-4514.

[10] 陈力生,朱亚兵,高前善. 审计风险管理研究[M]. 上海: 立信会计出版社, 2005.
(Chen L S, Zhu Y B, Gao Q S. Research on audit risk management[M]. Shanghai: Lixin Accounting Press, 2005.)

(上接第 898 页)

[8] Kim Y, Kim J H, Han K H. Quantum-inspired multiobjective evolutionary algorithm for multiobjective 0/1 knapsack problems [C]. Proc of 2006 IEEE Congress on Evolutionary Computation. Vancouver, 2006: 2601-2606.

[9] 李阳阳, 焦李成. 求解 SAT 问题的量子免疫克隆算法[J]. 计算机学报, 2007, 30(2): 176-183.
(Li Y Y, Jiao L C. Quantum-inspired immune clone algorithm for SAT problem [C]. Chinese J of Computers, 2007, 30(2): 176-183.)

[10] Nebro A J, Durillo J J, Luna F, et al. A cellular genetic algorithm for multiobjective optimization[C]. Proc of NICSO 2006. Granada, 2006: 25-36.

[11] Van Veldhuizen D A, Lamont G B. Multiobjective evolutionary algorithm research: A history and analysis [R]. Ohio: Air Force Institute Technology, 1998.

[12] Zitzler E, Thiele L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach [C]. IEEE Trans on Evolutionary Computation, 1999, 3(4): 257-271.

[13] 游晓明, 帅典勋, 刘升. 基于免疫原理的量子进化算法及收敛性研究[J]. 控制与决策, 2007, 22(7): 749-754.
(You X M, Shuai D X, Liu S. Research on quantum evolutionary algorithm based on immune theory and its convergence[J]. Control and Decision, 2007, 22(7): 749-754.)