第36卷 第9期
2009年9月
计 算 机 科 学
Computer Science
Vol. 36 No. 9
Sep 2009

基于广义变精度粗糙模糊集模型的知识发现

孙士保[1,2] 吴庆涛[2] 普杰信[2] 秦克云[3]

(北京航空航天大学软件开发环境国家重点实验室 北京 100191)[1]

(河南科技大学电子信息工程学院 洛阳 471003)[2] (西南交通大学智能控制开发中心 成都 610031)[3]

摘 要 介绍了广义Ziarko's变精度粗糙集模型和广义粗糙模糊集模型，找出了它们的不足。基于支集相对错误分类率及误差参数$\beta(0\leqslant\beta<0.5)$，提出了广义变精度粗糙模糊集模型，讨论了模型中$\beta$上、下近似算子的性质；分析了该模型与广义Pawlak's粗糙集模型、广义Ziarko's变精度粗糙集模型和广义粗糙模糊集模型的关系；最后给出了该模型中近似约简的定义和方法，并通过实例分析说明了约简算法的有效性。

关键词 变精度粗糙集，粗糙模糊集，广义变精度粗糙模糊集，近似约简

Knowledge Discovery Approach Based on Generalized Variable Precision Rough Fuzzy Set Model

SUN Shi-bao[1,2] WU Qing-tao[2] PU Jie-xin[2] QIN Ke-yun[3]

(National Lab. of Software Development Environment, Beijing University of Aeronautics & Astronautics, Beijing 100191, China)[1]

(Electronic Information Engineering College, Henan University of Science and Technology, Luoyang 471003, China)[2]

(Intelligent Control Development Center, Southwest Jiaotong University, Chengdu 610031, China)[3]

Abstract Generalized Ziarko's variable precision rough set model and generalized rough fuzzy set model were introduced and the drawbacks of them were found. Based on support relative error ratio and error parameter $\beta(0\leqslant\beta<0.5)$, the generalized variable precision rough fuzzy set model was proposed. The basic properties of approximation operators were investigated. The relation of between this model and generalized Pawlak's rough set model, generalized Ziarko's variable precision rough set model and generalized rough fuzzy set model was analysed in detail. Finally, the definitions and the approaches of approximation reduction were discussed, an example was given to illustrate the validity of the presented algorithms.

Keywords Variable precision rough set, Rough fuzzy set, Generalized variable precision rough fuzzy set, Approximation reduction

1 引言

粗糙集(rough sets)理论是由波兰学者Pawlak提出的一种处理不精确性问题的数学工具[1]。它根据已知数据自身的不可分辨关系，通过一对近似算子，对某一给定概念进行近似表示。粗糙集在信息处理、数据挖掘(DM)和数据库知识发现(KDD)等认知领域都有成功的应用[2,3]。但Pawlak粗糙集模型是有局限性的，一方面它所处理的分类必须是完全正确的，因而它的分类是精确的，亦即只考虑完全"包含"和"不包含"，而没有某种程度上的"包含"和"属于"；另一方面它所处理的对象是已知的，且得到的结论仅适用于这些对象。为了克服这些局限性，Ziarko提出了变精度粗糙集模型[4](VPRS模型，即Variable Precision Rough Set Model)，它的基本思想是在Pawlak's粗糙集模型中引入参数$\beta(0\leqslant\beta<0.5)$，即允许一定程度的错误分类率存在，它可以解决属性间无函数关系的数据分类问题，有利于从看似不相关的数据中发现潜在的相关数据。目前VPRS模型在很多领域得到了广泛的应用[5]。

有关变精度粗糙集模型的研究主要集中在等价关系下的约简方法和模型推广两个方面。等价关系下的约简方法有β约简[6]、β上(下)近似(分布)约简[7]、基于结构的约简[8]等，它们加快了变精度粗糙集理论的应用。但在很多实际问题中，对象之间的等价关系很难构造，或者对象之间本质上就没有等价关系，因此，需要对变精度粗糙集模型进行推广。它的推广是从两个方面来完成的，一方面是把被近似对象从清晰集推广为模糊集[9]，另一方面是把等价关系推广为论域上的模糊关系[9]、一般二元关系[10]或覆盖关系[11]等。文献[9]主要是把变精度思想引入到模糊粗糙集中以探讨变精度模糊粗糙集的理论和应用；文献[10]把变精度粗糙集模型中的等价关系推广为论域U上的一般二元关系R，从而得到广义的变

到稿日期：2008-10-27 返修日期：2009-01-27 本文受国家自然科学基金(60474022)，高等学校博士学科点专项科研基金(20060613007)，河南科技大学人才科学研究基金(09001172)资助。

孙士保(1970—)，男，博士后，副教授，主要研究方向为智能信息处理、机器学习和数据挖掘等，E-mail：sunshibao@126.com；吴庆涛(1975—)，男，博士，副教授，主要研究方向为信息安全和计算机网络等；普杰信(1959—)，男，博士，教授，主要研究方向为智能信息处理和数据挖掘等；秦克云(1962—)，男，博士，教授，博士生导师，主要研究方向为智能信息处理、机器学习和逻辑与不确定性推理等。

精度粗糙集模型；在粗糙集模型中，当论域上的等价关系推广为论域上的覆盖时可以得到覆盖粗糙集模型[12]，同样，文献[11]从对象的最小描述方面把论域上的等价关系推广到论域上的覆盖从而得到变精度覆盖粗糙集模型；文献[13]给出了等价关系下变精度粗糙模糊集的定义，由于等价关系很难成立，因此本文在一般关系下讨论了广义变精度粗糙模糊集模型中近似算子的定义、性质以及它与广义 Pawlak's 粗糙集模型、广义 Ziarko's 变精度粗糙集模型和粗糙模糊集模型的关系，最后给出了广义变精度粗糙模糊集模型中近似约简的理论与方法，并通过实例分析说明了广义变精度粗糙模糊集模型的应用。

2 广义 Ziarko's 变精度粗糙集模型

定义 1[10] 设 U 是有限非空的论域，$R\subseteq U\times U$ 是 U 上一个任意二元关系。对于 $\forall x,y\in U$，若 xRy，即 $(x,y)\in R$，则称 y 是 x 的后继，记 $R_S(x)=\{y\in U\mid xRy\}$。称 R 是串行的，若 $\forall x\in U, R_S(x)\neq\emptyset$；称 R 是自反的，若 $\forall x\in U, x\in R_S(x)$；称 R 是对称的，若 $\forall x,y\in U, x\in R_S(y)$ 蕴涵 $y\in R_S(x)$；称 R 是传递的，若 $\forall x,y\in U, y\in R_S(x)$ 蕴涵 $R_S(y)\subseteq R_S(x)$。

定义 2[10] 设 X 和 Y 表示有限论域 U 的非空子集，如果对于每一个 $e\in X$ 都有 $e\in Y$，则称 Y 包含 X，记作 $Y\supseteq X$。令

$$P(X,Y)=\begin{cases}1-\dfrac{|X\cap Y|}{|X|}, & |X|>0\\ 0, & |X|=0\end{cases} \tag{1}$$

其中，$|X|$ 表示集合 X 的基数。称 $P(X,Y)$ 为集合 X 关于集合 Y 的相对错误分类率。

定义 3[10] 设 (U,R) 为广义近似空间，其中 U 是有限非空的论域，$R\subseteq U\times U$ 是 U 上一个任意二元关系，对于任意的 $X\subseteq U$，定义 X 的 β 下近似和 X 的 β 上近似为

$$\begin{cases}\underline{R}_\beta(X)=\{x\in U\mid P(R_S(x),X)\leqslant\beta\}\\ \overline{R}_\beta(X)=\{x\in U\mid P(R_S(x),X)<1-\beta\}\end{cases} \tag{2}$$

当 $\beta=0$ 时，广义 Ziarko's 变精度粗糙集模型就变成广义 Pawlak's 粗糙集模型[14]。

3 广义粗糙模糊集模型

定义 4[7,14] 在 Pawlak's 粗糙集模型中，设 U 是有限非空的论域，用 $F(U)$ 表示 U 的模糊子集全体。$R\subseteq U\times U$ 是 U 上一个任意二元关系，则称 (U,R) 为广义近似空间。记 $R_S(x)=\{y\in U\mid xRy\}$ 为对象 x 的后继。$\forall A\in F(U)$，A 关于近似空间 (U,R) 的上近似 $\overline{R}(A)$ 和下近似 $\underline{R}(A)$ 是 U 上的一对模糊集合，其隶属函数分别定义为

$$\underline{R}(A)(x)=\inf\{\mu_A(y):y\in R_S(x)\}, x\in U \tag{3}$$

$$\overline{R}(A)(x)=\sup\{\mu_A(y):y\in R_S(x)\}, x\in U \tag{4}$$

算子 $\overline{R}$ 和 $\underline{R}$：$F(U)\rightarrow F(U)$ 分别称为粗糙模糊集的下、上近似算子，序对 $(\overline{R}(A),\underline{R}(A))$ 称为粗糙模糊集。若 $\overline{R}(A)=\underline{R}(A)$，则称 A 是可定义的，否则称 A 是不可定义的。容易验证，当 $A\in P(U)$ 时，上述定义退化成广义 Pawlak's 粗糙集。

经过分析可知这种上、下近似算子有以下两个缺点：①上、下近似具有限制的特性，一个相对小的错误分类率可能导致某些不可区分类被排除在下近似之外，同样也可能导致上近似的不必要的增加，这些特性在大型论域(从动态过程中产生)的例子中特别重要；②上、下近似值是由成员函数 $\mu_A(y)$ 决定的，这种近似应该由集合包含的全体值来评价。

4 广义变精度粗糙模糊集模型

为了克服式(3)、式(4)中的缺点①，把广义变精度粗糙集模型的思想用于广义粗糙模糊集模型，从而得到广义变精度粗糙模糊模型。

定义 5[13] 设 A 和 B 表示有限论域 U 上的模糊子集，令

$$P_S(A,B)=\begin{cases}1-\dfrac{|\mathrm{supp}(A\cap Y)|}{|\mathrm{supp}(X)|}, & |\mathrm{supp}(X)|>0\\ 0, & |\mathrm{supp}(X)|=0\end{cases} \tag{5}$$

其中，$|X|$ 表示集合 X 的基数。称 $P_S(A,B)$ 为集合 A 关于集合 B 的支集相对错误分类率。

定理 1 设 U 是有限非空的论域，对于任意的非空清晰集 $A,B\subseteq U$，则 A 关于集合 B 的支集相对错误分类率 $P_S(A,B)$ 等价于相对错误率 $P(A,B)$。

证明：由 $A,B\subseteq U$ 都是清晰集得知：$\forall x\in A\cap B$，有 $\mu_{A\cap B}(x)=1$；$\forall x\notin A\cap B$，有 $\mu_{A\cap B}(x)=0$。因此，$A\cap B$ 与 $\mathrm{supp}(A\cap B)$ 是等价的。结果成立。

定义 6 设 U 是有限非空的论域，$R\subseteq U\times U$ 是 U 上一个任意二元关系，则称 (U,R) 为广义近似空间。记 $R_S(x)$ 为对象 x 的后继。$\forall A\in F(U)$，A 关于近似空间 (U,R) 的 β 上近似 $\overline{R}_\beta(X)$ 和 β 下近似 $\underline{R}_\beta(X)$ 是 U 上的一对模糊集合，其隶属函数分别定义为

$$\underline{R}_\beta(A)(x)=\begin{cases}\inf\{\mu_A(y):y\in M\},\\ \quad \text{if } P_S(R_S(x),A)\leqslant\beta, x\in U\\ 0, \quad \text{otherwise}\end{cases} \tag{6}$$

$$\overline{R}_\beta(A)(x)=\begin{cases}\sup\{\mu_A(y):y\in M\},\\ \quad \text{if } P_S(R_S(x),A)<1-\beta, x\in U\\ 0, \quad \text{otherwise}\end{cases} \tag{7}$$

其中，$M=\mathrm{supp}(R_S(x)\cap A)$ 是 $R_S(x)$ 和 A 交集的支集。

算子 $\underline{R}_\beta$ 和 $\overline{R}_\beta$：$F(U)\rightarrow F(U)$ 分别称为变精度粗糙模糊集的 β 下、上近似算子，序对 $(\underline{R}_\beta(A),\overline{R}_\beta(A))$ 称为变精度粗糙模糊集。若 $\underline{R}_\beta(A)=\overline{R}_\beta(A)$，则称 A 是可定义的，否则称 A 是不可定义的。

定理 2 设 (U,R) 是广义近似空间，$\forall X,Y\in F(U)$，$0\leqslant\alpha,\beta<0.5$ 则广义变精度粗糙模糊集模型中 β 下近似算子 $\underline{R}_\beta(X)$ 和 β 上近似算子 $\overline{R}_\beta(X)$ 满足如下性质：

(1)若 R 是串行的，$\underline{R}_\beta(U)=\overline{R}_\beta(U)=U$，$\underline{R}_\beta(\emptyset)=\overline{R}_\beta(\emptyset)=\emptyset$；

(2)$X\subseteq Y\Rightarrow\underline{R}_\beta(X)\subseteq\underline{R}_\beta(Y)$，$\overline{R}_\beta(X)\subseteq\overline{R}_\beta(Y)$；

(3)$\underline{R}_\beta(X\cap Y)\subseteq\underline{R}_\beta(X)\cap\underline{R}_\beta(Y)$，$\overline{R}_\beta(X\cup Y)\supseteq\overline{R}_\beta(X)\cup\overline{R}_\beta(Y)$，$\underline{R}_\beta(X\cup Y)\supseteq\underline{R}_\beta(X)\cup\underline{R}_\beta(Y)$，$\overline{R}_\beta(X\cap Y)\subseteq\overline{R}_\beta(X)\cap\overline{R}_\beta(Y)$；

(4)$\underline{R}_\beta(X)\subseteq\overline{R}_\beta(X)$；

(5)$\alpha\geqslant\beta\Rightarrow\underline{R}_\alpha(X)\supseteq\underline{R}_\beta(X)$，$\overline{R}_\alpha(X)\subseteq\overline{R}_\beta(X)$；

(6)$\underline{R}_\beta(X)=\sim\overline{R}_\beta(\sim X)$，$\overline{R}_\beta(X)=\sim\underline{R}_\beta(\sim X)$。

证明：(1)由定义可知，对于 $\forall x\in U, R_S(x)\neq\emptyset$，$P_S(R_S(x),U)=0$，从而任意的 $x\in U$，$\underline{R}_\beta(U)(x)=\overline{R}_\beta(U)(x)=1$，故 $\underline{R}_\beta(U)=\overline{R}_\beta(U)=U$。而 $P_S([x]_R,\emptyset)=1$，$\underline{R}_\beta(\emptyset)(x)=\overline{R}_\beta(\emptyset)(x)=0$，故 $\underline{R}_\beta(\emptyset)=\overline{R}_\beta(\emptyset)=\emptyset$。

(2)由 $X\subseteq Y$，则 $\mu_X(x)\leqslant\mu_Y(x)$ 且 $\forall x\in U, P_S(R_S(x),Y)\leqslant$

$P_S(R_S(x),X)$，故 $\forall x\in U$，有 $\underline{R}_\beta(X)(x)\leqslant\underline{R}_\beta(Y)(x)$，因此 $\underline{R}_\beta(X)\subseteq\underline{R}_\beta(Y)$。同理，$\overline{R}_\beta(X)\subseteq\overline{R}_\beta(Y)$。

(3)由(2)知 $\underline{R}_\beta(X\cap Y)\subseteq\underline{R}_\beta(X)$，$\underline{R}_\beta(X\cap Y)\subseteq\underline{R}_\beta(Y)$，从而有 $\underline{R}_\beta(X\cap Y)\subseteq\underline{R}_\beta(X)\cap\underline{R}_\beta(Y)$。

其余各式证明类似。

(4)由上、下近似的定义可直接推出。

(5)当 $\alpha\geqslant\beta$ 且 $P_S(R_S(x),X)\leqslant\beta$ 时，有 $P_S(R_S(x),X)\leqslant\alpha$，则 $\underline{R}_\alpha(X)(x)\geqslant\underline{R}_\beta(X)(x)$，因此，$\underline{R}_\alpha(X)\supseteq\underline{R}_\beta(X)$。同理 $\overline{R}_\alpha(X)\subseteq\overline{R}_\beta(X)$。

(6)$\overline{R}_\beta(\sim X)(x)=\sup\{1-\mu_X(y)\mid y\in M\}$，$\sim\overline{R}_\beta(\sim X)(x)=1-\sup\{1-\mu_X(y)\mid y\in M\}=\inf\{\mu_X(y)\mid y\in M\}=\underline{R}_\beta(X)$。

同理 $\overline{R}_\beta(X)=\sim\underline{R}_\beta(\sim X)$。

5 与其它粗糙集模型的关系

(1)当 $A\in P(U)$ 时与广义 Ziarko's 变精度粗糙集模型等价

当 $A,B\in P(U)$ 时，$P_S(A,B)=P(A,B)$，$M=R_S(x)\cap A$。$\forall x\in U$，$P_S(R_S(x),A)\leqslant\beta$，则必有 $P_S(R_S(x),A)<1-\beta$。

在广义 Ziarko's 变精度粗糙集模型中，$\forall x\in U$，$x\in\underline{R}_\beta(A)$，则 $\underline{R}_\beta(A)(x)=1$；而在本文模型中 $\forall x\in U$，由于 $M=R_S(x)\cap A$，$\forall y\in M$，都有 $\mu_A(y)=1$，$\underline{R}_\beta(A)(x)=\inf\{\mu_A(y)\mid y\in M\}=1$。反之，在本文模型中 $\forall x\in U$，如果存在关系式 $\underline{R}_\beta(A)(x)=\inf\{\mu_A(y)\mid y\in M\}=1$，则 $P_S(R_S(x),A)\leqslant\beta$，也即 $P(R_S(x),A)\leqslant\beta$，故在广义 Ziarko's 变精度粗糙集模型中 $x\in\underline{R}_\beta(A)$。这样两种模型的下近似等价。

同理，两种模型的上近似也等价。

这说明广义 Ziarko's 变精度粗糙集模型是广义变精度粗糙模糊集模型的特例。

(2)当 $\beta=0$ 时与广义粗糙模糊集模型等价

当 $\beta=0$ 时，由 $P_S(R_S(x),A)\leqslant0$ 知 $|\mathrm{supp}(R_S(x)\cap A)|=|R_S(x)|$。

所以，

$$\underline{R}_0(A)(x)=\begin{cases}\inf\{\mu_A(y):y\in R_S(x)\},\\ \quad\text{if } P_S(R_S(x),A)\leqslant0,x\in U\\ \qquad 0, \qquad\qquad \text{otherwise}\end{cases}$$

由定义 4 可知，它与广义粗糙模糊集模型的下近似等价。

同理由定义 5 可知，$\overline{R}_0(A)(x)$ 与广义粗糙模糊集模型的上近似等价。

这也说明广义粗糙模糊集模型是广义变精度粗糙模糊集模型的特例。

(3)当 $A\in P(U)$ 且 $\beta=0$ 时与广义 Pawlak's 粗糙集模型等价

由上述关系(1)和(2)可以直接证明。

6 广义变精度粗糙模糊集模型的另一种定义形式

为了克服式(3)、式(4)中的缺点②，给出广义变精度粗糙模糊集模型的另一种定义方法：

$$\underline{R}_\beta(A)(x)=\begin{cases}\dfrac{\mathrm{power}(R_S(x)\cap A)}{|\mathrm{supp}(R_S(x)\cap A)|},\\ \text{if } P_S(R_S(x),A)\leqslant\beta,x\in U\\ \qquad 0,\quad \text{otherwise}\end{cases}\tag{8}$$

$$\overline{R}_\beta(A)(x)=\begin{cases}\dfrac{\mathrm{power}(R_S(x)\cap A)}{|\mathrm{supp}(R_S(x)\cap A)|},\\ \text{if } P_S(R_S(x),A)<1-\beta,x\in U\\ \qquad 0,\qquad \text{otherwise}\end{cases}\tag{9}$$

其中，$\mathrm{power}(A)=\sum_{i=1}^{n}\mu_A(x_i)$。

(1)当 $A\in P(U)$ 时与广义 Ziarko's 变精度粗糙集模型等价

当 $A,B\in P(U)$ 时，$P_S(A,B)=P(A,B)$，$\mathrm{supp}\,(R_S(x)\cap A)=R_S(x)\cap A$ 及 $|\mathrm{supp}(R_S(x)\cap A)|=\mathrm{power}(R_S(x)\cap A)$。$\forall x\in U$，$P_S(R_S(x),A)\leqslant\beta$，则必有 $P_S(R_S(x),A)<1-\beta$。

在广义 Ziarko's 变精度粗糙集模型中，$\forall x\in U$，$x\in\underline{R}^\beta(A)$，则 $\underline{R}^\beta(A)(x)=1$；而在本文模型中 $\forall x\in U$，由于 $\mathrm{supp}(R_S(x)\cap A)=R_S(x)\cap A$ 及 $|\mathrm{supp}(R_S(x)\cap A)|=\mathrm{power}(R_S(x)\cap A)$，$\forall y\in R_S(x)\cap A$，都有 $\mu_A(y)=1$，$\underline{R}^\beta(A)(x)=1$。反之，在本文模型中 $\forall x\in U$，如果存在关系式 $\underline{R}^\beta(A)(x)=1$，则 $P_S(R_S(x),A)\leqslant\beta$，也即 $P(R_S(x),A)\leqslant\beta$，故在广义 Ziarko's 变精度粗糙集模型中 $x\in\underline{R}^\beta(A)$。这样两种模型的下近似等价。

同理，两种模型的上近似也等价。

这说明广义 Ziarko's 变精度粗糙集模型也是本节定义模型的特例。

(2)当 $A\in P(U)$ 且 $\beta=0$ 时与广义 Pawlak's 粗糙集模型等价

与上一部分的分析类似。

(3)该定义中的上、下近似算子是由后继集合中的成员函数加权平均来实现的，它解决了由单值确定不稳定的缺点，但它与广义粗糙模糊集模型之间具有较大的差异，它不是广义粗糙模糊集模型的一般形式。在数据处理过程中，当强调数据间的公平性时可以采用本模型；当强调数据的个性特色时要采用式(6)和式(7)的模型。

7 广义变精度粗糙模糊集模型中的近似约简

属性约简是粗糙集模型理论中最重要的概念之一，所谓目标信息系统[7,14]的一个约简是指保持和决策属性相同或接近的最小条件属性子集。在粗糙集模型中有基于正域的约简[14]、基于包含度的约简[7]、基于信息熵的约简[15]和基于覆盖的约简方法[13]；在变精度粗糙集模型中基于等价关系下的约简方法有 β 约简[6]、β 上(下)近似(分布)约简[7]、基于结构的约简[8]等。目前还没有相关文献对粗糙模糊集模型和模糊粗糙集模型中的约简问题进行探讨。本文从近似质量角度出发，给出广义变精度粗糙模糊集模型中的约简定义及约简算法，并用实例说明了算法的有效性。

(1)近似质量

设论域是 U，属性集为 C，R 是由 C 决定的 U 上一个任意二元关系，$A\in F(U)$，则

$$\gamma_R=\frac{\mathrm{power}(pos_R(U))}{|U|}\tag{10}$$

被称为由 R 决定的 A 的近似质量，其中 $pos_R(U)=\underline{R}_\beta(A\cap R_S(x))(x\in U)$，$\mathrm{power}(A)=\sum_{i=1}^{n}\mu_A(x_i)$。近似质量反映了下近似中的元素占全体论域元素的百分比。

(2)广义变精度粗糙模糊集模型中近似约简的定义

定义 7(约简) 目标信息系统中删除一个属性元素得到一个新的目标信息系统,如果新目标信息系统的近似质量与原目标信息系统的近似质量相同或接近,则新目标信息系统是原目标信息系统的一个约简。

从定义中可以看出:在约简中删除的是对一个目标信息系统中近似质量没有影响或影响不大的一个元素;一个目标信息系统可能存在多个约简,但每个约简可能都不是最终的约简。

(3)广义变精度粗糙模糊集模型中的近似约简算法

算法:近似约简

输入:一个目标信息系统。

输出:目标信息系统的约简。

步骤 1 初始化。求原目标信息系统的近似质量 γ_R;

步骤 2 求原目标信息系统中删除一个属性元素的关系 R'、由 R'确定的划分 $U/R'=\{C_1,C_2,\cdots,C_t\}$和新目标信息系统的近似质量 $\gamma_R{}'$;

步骤 3 比较 γ_R 和 $\gamma_R{}'$,如果相同或接近则新目标信息系统是原目标信息系统的一个约简;

步骤 4 依次按步骤 2、步骤 3 对目标信息系统中的各个属性进行试探,看所得到的新目标信息系统是否是原目标信息系统的约简。

从上面的算法可以看出:原目标信息系统的约简可能不止一个;该算法是对目标信息系统中的各个属性进行试探,每次可能都会得到原目标信息系统的一个约简,但这个约简又可能不是最终的约简;另外,步骤 3 中的接近程度要根据具体的应用要求来确定达到什么值时才算是得到目标信息系统的约简。

8 应用举例

用下面的例子来分析一个目标信息系统。如表 1 所列,其中 $U=\{x_1,x_2,x_3,x_4,x_5\}$为论域,$(c_1,c_2,c_3)$为条件属性,$(d)$为决策属性。设 $A=0.6/x_1+0.4/x_2+0/x_3+0.5/x_4+0.7/x_5$ 是论域 U 上的模糊集。

表 1 目标信息系统

U	c1	c2	c3	d
x_1	A_2	B_2	C_2	D_1
x_2	A_1	B_2	C_2	D_2
x_3	A_3	B_1	C_1	D_2
x_4	A_1	B_1	C_1	D_1
x_5	A_1	B_1	C_2	D_1

根据前面的约简算法可知目标信息系统的近似质量 $\gamma_R=0.44$;当去掉属性 c_3 时,新的近似质量 $\gamma_R{}'=0.4$;而当去掉属性 c_1 或 c_2 时,新的近似质量 $\gamma_R{}'=0.44$。这说明去掉属性 c_3 时会严重影响新目标信息系统的近似质量,而去掉属性 c_1 或 c_2 时不会影响新系统的近似质量,这足以表明属性 c_1 或 c_2 在系统中是可以约简的,故原目标信息系统的约简为(c_1,c_3)或(c_2,c_3),即(c_1,c_3)或(c_2,c_3)可以和条件属性集(c_1,c_2,c_3)产生同样的决策能力。

结束语 广义变精度粗糙集模型在处理含有不完备信息的目标系统中是非常有用的,粗糙模糊集模型是在等价关系下对模糊数据进行处理的,把这两种方法结合起来处理目标信息系统中含有不完备的数据具有广阔的应用前景。本文就是基于这一目的对广义变精度粗糙模糊模型进行了详尽的探讨并给出了它的应用理论与方法。但是,当该模型中的一般关系变成论域上的模糊关系时,如何进行数据处理将是我们下一步要研究的内容。

参考文献

[1] Pawlak Z,Skowron A. Rudiments of rough sets[J]. Information Sciences,2007,177:3-27

[2] Pawlak Z,Skowron A. Rough sets:Some extensions[J]. Information Sciences,2007,177:28-40

[3] Pawlak Z,Skowron A. Rough Sets and Boolean Reasoning[J]. Information Sciences,2007,177:41-73

[4] Ziarko W. Variable precision rough set model [J]. Journal of computer system science,1993,46(1):39-59

[5] 陶志,许宝栋,汪定伟,等. 基于变精度粗糙集理论的粗糙规则挖掘算法[J]. 信息与控制,2004,33(1):18-22

[6] Beynon M. Reducts within the variable precision rough sets model:A further investigation[J]. European journal of operational research,2001,134:592-605

[7] 张文修,梁怡,吴伟志,等. 信息系统与知识发现[M]. 北京:科学出版社,2003:56-67

[8] Inuiguchi M. Structure-Based Approaches to Attribute Reduction in Variable Precision Rough Set Models[C]//Proceeding of 2005 IEEE International Conference on Granular Computing. Beijing China,IEEE GrC2005,2005:34-39

[9] Mieszkowicz-Rolka A,Rolka L. Remarks on approximation quality in variable precision fuzzy rough sets model[C]//Rough Sets and Current Trends in Computing:4th International Conference,RSCTC 2004. Heidelberg:Springer-Verlag,2004:402-411

[10] 巩增泰,孙秉珍,邵亚斌,等. 一般关系下的变精度粗糙集模型[J]. 兰州大学学报:自然科学版,2005,41(6):110-114

[11] 张亚军,王艳平. 基于覆盖的变精度粗糙集模型[J]. 辽宁工学院学报,2006,26(4):274-276

[12] Zhu W,Wang F Y. Reduction and axiomization of covering generalized rough sets[J]. Information Sciences,2003,152:217-230

[13] Mieszkowicz-Rolka A,Rolka L. Fuzziness in Information Systems[J]. Electronic Notes in Theoretical Computer Science,2003,82(4):1-10

[14] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社,2003:41-131

[15] 王国胤,于洪,杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报,2002,25(7):759-766

(上接第 99 页)

[8] Kaya O S. A glance at Peer to Peer systems[R]. TR-CTIT-05-21,University of Twente,2005,1381-3625

[9] 王勇,云晓春,李奕飞. 对等网络拓扑测量与特征分析[J]. 软件学报,2008,19(4):981-992

[10] Foster I,Iamnitchi A. On death,taxes,and the convergence of peer-to-peer and grid computing [C]//The 2nd Int'l Workshop on Peer-to-Peer Systems (IPTPS'03). Berkeley,USA,2003

[11] 陈贵海,须成忠,沈海英,等. 一种新的常数度数的 p2p 覆盖网络[J]. 计算机学报,2005,28(7):1084-1095

[12] 刘方爱,刘志勇. 环、mesh 嵌入 RP(k)网络[J]. 中国科学 E 辑,2004,34(8):919-929