# 一般二元关系下基于粗糙隶属函数的程度粗糙集

徐伟华[1], 刘士虎[1], 张文修[2]

(1.重庆理工大学 数学与统计学院,重庆 400054; 2.西安交通大学 理学院,西安 710049)

摘 要:在分析经典粗糙集模型不足的基础上,通过引入精度系数 $k(k\in(0.5,1])$,给出了一般关系下基于粗糙隶属函数的程度粗糙集,并讨论了所给模型的相关重要性质。与经典粗糙集模型相比较,发现该模型不仅是对经典粗糙集模型的拓展,还是对基于等价关系的变精度粗糙集模型的拓展。

关 键 词:一般二元关系;程度粗糙集;粗糙隶属函数;变精度粗糙集

中图分类号:TP18 文献标识码:A 文章编号:1674-8425(2010)10-0101-08

# The Graded Rough Set based on Membership Function in General Binary-relation

XU Wei-hua, LIU Shi-hu, ZHANG Wen-xiu

(1. School of Mathematics and Physics, Chongqing University of Technology, Chongqing 400054, China;
2. School of Sciences, Xi'an Jiaotong University, X'ian 710049, China)

**Abstract** Upon scribing the limits of classical rough set model, the graded rough set model based on the rough membership function is established by introducing the precision coefficient and some important properties of it are discussed. Its key issue is that the graded rough set model can be seen not only as a expansion of classical rough set model but also as a variable precision rough set model based on general binary-relation. These results will be very helpful for the research of rough set model expansion and significant for establishing a framework of knowledge discovering in database system.

**Key words** general binary-relation; graded rough set; rough membership function; variable precision rough set

收稿日期:2010-05-20
基金项目:重庆市教委科学技术研究项目(KJ090612)
作者简介:徐伟华(1979—),男,山西浑源人,博士,副教授,主要从事模糊集、粗糙集、人工智能的数学基础研究。

1982年波兰数学家 Pawlak Z 提出的经典粗糙集理论[1-2]是一种新的处理模糊和不确定性知识的软计算工具。该理论将知识看作是对论域的划分, 把分类理解为在特定空间上的等价关系(满足自反性、对称性和传递性), 其主要思想就是在保持分类能力不变的前提下, 通过知识约简, 导出问题的决策或分类规则。由于该理论蕴含的思想独特, 方法新颖, 已越来越引起国际学术界的关注。

经典粗糙集模型的一个局限性是它所处理的分类必须是完全正确或肯定的, 亦即只考虑"包含"与"属于"问题, 没有某种程度上的"包含"与"属于"问题。另一个局限性是它所处理的对象是已知的, 且所得到的结论仅仅适用于某些特定的对象集。然而在实际应用中, 有时由于信息缺失等各种原因, 导致数据库系统中的知识划分是不以等价关系为基础, 有时有些问题需要将一些小规模的对象集合中得到的结论运用到大规模的对象集中去等, 这使得经典粗糙集模型的应用受到了极大的限制。为此, Ziarko[3]研究了等价关系下的变精度粗糙集模型问题。该模型考虑了某种程度上的包含与属于问题, 但是对知识的划分仍然以等价关系为基础。张等[4]研究了一般关系下的程度粗糙集模型, 但是对程度的大小并没有给出严格的说明, 这使得对经典粗糙集模型的推广[5-6]十分必要。可喜的是, 近年来对推广经典粗糙集的研究取得了许多成果[7-11]。

对此, 非常有必要考虑数据库系统中知识在一般二元关系下的"某种程度"上的包含与属于问题。故本文在数据库系统中以通过放宽等价关系为一般二元关系为基础, 以精度系数 $k$ 为"某种程度"的准则, 建立了基于粗糙隶属函数的程度粗糙集模型, 并讨论了在给定精度系数的前提下该模型的相关重要性质。进一步结合知识的粗糙度与属性约简问题, 给出了在精度系数 $k$ 下的粗糙度的定义以及属性的近似约简与依赖问题。通过讨论精度系数的变化得知, 当 $k=1$ 时, 该模型退化为一般关系下的粗糙集模型, 而当一般关系强化为等价关系时, 该模型就成为 Ziarko 研究的变精度粗糙集。这些结论为数据库系统中知识的划分、发现以及规则的提取奠定了一定的理论研究基础。

## 1 一般二元关系下粗糙集的基本概念

为方便论述, 首先给出一般二元关系下粗糙集的基本概念。

定义 1[4] 称三元组 $K=(U, A, F)$ 为一数据库系统, 其中: $U=(u_1, u_2, \cdots, u_n)$ 为对象集, 每个 $u_i(i\leqslant n)$ 称为一个对象; $A=(a_1, a_2, \cdots, a_m)$ 为属性集, 每个 $a_j(j\leqslant m)$ 称为一个属性; $F=\{f_l \mid f_l: U\rightarrow V_l(l\leqslant m)\}$ 为论域 $U$ 和属性集 $A$ 的关系集, 其中 $V_l$ 为属性 $a_l$ 的值域。

在数据库系统中, 如果属性集 $A$ 可以分为条件属性集 $C$ 和决策属性集 $D$, 即 $A=C\cup D$, $C\cap D=\phi$, $D\neq\phi$, 则该数据库系统称为带有决策的或决策系统。

设三元组 $K=(U, A, F)$ 为一数据库系统, $B\subseteq A$, $R_B=\{(u, v)\in U\times U \mid uRv, \forall a_i\in B\}$ 为论域 $U$ 上关于属性子集 $B$ 的一个二元关系。若记 $[u_i]_{R_B}=\{u_j \mid (u_i, u_j)\in R_B\}$, $U/R_B=\{[u_i]_{R_B} \mid \forall u_i\in U\}$, 则称 $[u_i]_{R_B}$ 为 $u_i$ 关于 $R_B$ 的邻域, $U/R_B$ 为该数据库系统中论域 $U$ 关于 $R_B$ 的一个分类。

本文若未特别申明, 所说的关系均指一般二元关系。

定义 2[4] 设三元有序组 $K=(U, A, F)$ 为一数据库系统, $X\subseteq U$, $B\subseteq A$, $R_B$ 是任意给定的二元关系, 定义 $X$ 关于该数据库系统 $K$ 的下近似和上近似分别为:

$$\underline{R_B}(X)=\{u\in U \mid [u]_{R_B}\subseteq X\}$$

$$\overline{R_B}(X)=\{u\in U \mid [u]_{R_B}\cap X\neq\phi\}$$

定义 3[4] 设三元有序组 $K=(U, A, F)$ 为一数据库系统, $B\subseteq A$, $X\in U$, $R_B$ 为给定的二元关系。

1) 称 $X$ 为精确集, 当且仅 $\underline{R_B}(X)=\overline{R_B}(X)$。

2) 称 $X$ 为粗糙集, 当且仅当 $\underline{R_B}(X)\neq\overline{R_B}(X)$。

定义 4[4] 设三元组 $K=(U, A, F)$ 为一数据

库系统, $B\subseteq A$, $R_B$ 是给定的二元关系, 粗糙集 $X$ 关于该数据库系统的上近似与下近似分别为 $\underline{R_B}(X)$ 和 $\overline{R_B}(X)$, 称

$$\theta_{R_B}(X)=1-\frac{|\underline{R_B}(X)|}{|\overline{R_B}(X)|}$$

为 $X$ 关于该数据库系统的粗糙度。

例 1 表 1 是一数据库系统。若取 $B=\{a_1, a_2, a_3, a_4, a_5\}$, 即可产生该数据库系统上的一个优势关系 $R_B$, 即

$$[u_i]^{\succeq}_{R_B}=\{u_j\in U\mid f_l(u_j)\geqslant f_l(u_i)(\forall a_l\in B)\}$$

由上述定义有

$$[u_1]^{\succeq}_{R_B}=\{u_1, u_2, u_5, u_6\},\quad [u_2]^{\succeq}_{R_B}=\{u_2, u_6\}$$
$$[u_3]^{\succeq}_{R_B}=\{u_3, u_5, u_6\},\quad [u_4]^{\succeq}_{R_B}=\{u_4\}$$
$$[u_5]^{\succeq}_{R_B}=\{u_5\},\quad [u_6]^{\succeq}_{R_B}=\{u_6\}$$

取 $X=\{u_1, u_3, u_5, u_6\}$时, 其下、上近似分别为

$$\underline{R^{\succeq}_B}(X)=\{u_5, u_6\}$$
$$\overline{R^{\succeq}_B}(X)=\{u_1, u_3, u_5, u_6\}$$

由于 $\underline{R^{\succeq}_B}(X)\neq\overline{R^{\succeq}_B}(X)$, 故 $X$ 为粗糙集, 其粗糙度 $\theta^{\succeq}_{R_B}(X)=1-\frac{2}{5}=\frac{3}{5}$。

表 1 数据库系统

| $U$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|---|---|---|---|---|---|
| $u_1$ | 1 | 2 | 1 | 1 | 1 |
| $u_2$ | 3 | 2 | 2 | 4 | 3 |
| $u_3$ | 1 | 1 | 2 | 2 | 1 |
| $u_4$ | 2 | 1 | 3 | 1 | 2 |
| $u_5$ | 3 | 3 | 2 | 3 | 4 |
| $u_6$ | 3 | 2 | 3 | 4 | 4 |

命题 1[4] 设三元组 $K=(U, A, F)$为一数据库系统, $X\subseteq U$, $R_B$ 是给定的二元关系。

1) $\forall X\subseteq U$, 均有 $0\leqslant\theta_{R_B}(X)\leqslant 1$成立。

2) $X$ 为精确集, 当且仅当 $\theta_{R_B}(X)=0$。

3) $X$ 为粗糙集, 当且仅当 $\theta_{R_B}(X)>0$。

定义 5[4] 设三元组 $K=(U, A, F)$为一数据库系统, $X\subseteq U$, $B\subseteq A$, $R_B$ 是给定的二元关系。$X$ 的正域 $pos_{R_B}(X)$, 负域 $neg_{R_B}(X)$, 边界域 $bn_{R_B}(X)$分别定义为:

$$pos_{R_B}(X)=\underline{R_B}(X)$$
$$neg_{R_B}(X)=U-\overline{R_B}(X)$$
$$bn_{R_B}(X)=\overline{R_B}(X)-\underline{R_B}(X)$$

定义 6[4] 设三元组 $K=(U, A, F)$为一数据库系统, $X\subseteq U$, $B$, $C\subseteq A$, 给定关系 $R_B$, $R_C$。$R_C$ 的 $R_B$ 正域 $pos_{R_B}(R_C)$, 负域 $neg_{R_B}(R_C)$分别定义为:

$$pos_{R_B}(R_C)=\bigcup_{X\in U/R_C}pos_{R_B}(X)$$

$$neg_{R_B}(R_C)=U-pos_{R_B}(R_C)=\bigcup_{X\in U/R_C}neg_{R_B}(X)$$

定义 7[4] 设三元组 $K=(U, A, F)$为一数据库系统, $B\subseteq A$, $b\in B$。若 $U/R_B=U/R_{B-\{b\}}$, 则称 $b\in B$是不必要的, 否则称 $b\in B$ 是必要的。

定义 8[4] 设三元组 $K=(U, A, F)$为一数据库系统, $B\subseteq A$, 若对于每个 $b\in B$ 是必要的, 则称 $R_B$ 是独立的, 否则称 $R_B$ 是依赖的, 或不独立的。

定义 9[4] 设三元组 $K=(U, A, F)$为一数据库系统, $B\subseteq A$, 若进一步满足:

1) $R_B$ 是独立的;

2) $U/R_B=U/R_A$;

则称 $B$ 是 $A$ 的一个约简, 记为 $red(R_B)$。

定义 10[4] 设三元组 $K=(U, A, F)$为一数据库系统, 将 $R_A$ 中所有必要属性组成的集合称之为 $R_A$ 的核, 记为 $core(R_A)$, 即

$$core(R_A)=\bigcap_{B\subseteq A}red(R_B)$$

关于粗糙集理论中更为详细的概念以及定义请参阅文献[4]。

## 2 一般二元关系下基于粗糙隶属函数的程度粗糙集

以上给出的一般二元关系下的粗糙集模型, 完全是通过对象 $u$ 的邻域 $[u]_{R_B}$ 与论域 $U$ 中的集合 $X$ 的简单的定性关系来定义近似算子的, 即只考虑“包含”与“属于”这种情况, 但这样定义有一个缺陷, 那就是没有考虑到对象 $u$ 的邻域 $[u]_{R_B}$ 与 $X$ 重叠部分的定量信息, 亦即不存在某种程度上的“包含”与“属于”。若要依据 $[u]_{R_B}$ 与 $X$ 重叠的多少来刻画集合 $X$ 时, 这一模型就显得无能为力了。为了解决这个问题, 通过引入“精度系数”这

一概念给出了一般二元关系下基于粗糙隶属函数的程度粗糙集模型。

定义 11[12] 设三元组 $K=(U,A,F)$为一数据库系统, $X\in U$, $B\subseteq A$, $R_B$ 为给定的二元关系。记对象 $u$在关系 $R_B$ 下关于集合 $X$ 的粗糙隶属函数(度)为

$$\mu_X^{R_B}(u)=\frac{|[u]_{R_B}\cap X|}{|[u]_{R_B}|}$$

从上述定义可以看出, 粗糙隶属函数(度)的确定是通过数据库系统所给出的数据信息计算得知的, 与模糊数学中隶属函数的确定相比减少了主观性, 增加了客观性。

定义 12 设三元组 $K=(U,A,F)$为一数据库系统, $B\subseteq A$, $X\subseteq U$, $k\in(0.5,1]$。定义 $X$ 依粗糙隶属函数 $\mu_X^{R_B}(u)$的 $k-$下近似和 $k-$上近似分别为:

$$\underline{R_B^k}(X)=\{u\in U\mid\mu_X^{R_B}(u)\geqslant k\}$$

$$\overline{R_B^k}(X)=\{u\in U\mid\mu_X^{R_B}(u)>1-k\}$$

进一步, 称 $X$ 是 $k-$精确集, 当且仅当 $\underline{R_B^k}(X)=\overline{R_B^k}(X)$, 否则 $X$ 是 $k-$粗糙集。

由上述定义知, 当对象 $u$ 在关系 $R_B$ 下关于集合 $X$ 的粗糙隶属函数(度)不小于 $k$时, 对象 $u$ 就属于$\underline{R_B^k}(X)$; 当对象 $u$ 在关系 $R_B$ 下关于集合 $X$ 的粗糙隶属函数(度)大于 $1-k$时, 对象 $u$ 就属于$\overline{R_B^k}(X)$。

与 Pawlak Z 经典粗糙集相比, $k-$下近似算子和 $k-$上近似算子具有下列重要性质。

定理 1 设 $K=(U,A,F)$为一数据库系统, $X\subseteq U$, $B\subseteq A$, $R_B$ 为给定的二元关系, 对于 $k\in(0.5,1]$, 下述结论成立:

$(L_1)$ $\underline{R_B^k}(X)=\sim\overline{R_B^k}(\sim X)$

$(U_1)$ $\overline{R_B^k}(X)=\sim\underline{R_B^k}(\sim X)$

$(L_2)$ $\underline{R_B^k}(U)=U$

$(U_2)$ $\overline{R_B^k}(\Phi)=\Phi$

$(L_3)$ $\mu_X^{R_B}(\underline{R_B^k}(X))\geqslant k$

$(U_3)$ $\underline{R_B^k}(X)\subseteq\overline{R_B^k}(X)$

$(L_4)$ $X\subseteq Y\Rightarrow\underline{R_B^k}(X)\subseteq\underline{R_B^k}(Y)$

$(U_4)$ $X\subseteq Y\Rightarrow\overline{R_B^k}(X)\subseteq\overline{R_B^k}(Y)$

$(L_5)$ $\underline{R_B^k}(X\cup Y)\supseteq\underline{R_B^k}(X)\cup\underline{R_B^k}(Y)$

$(U_5)$ $\overline{R_B^k}(X\cap Y)\subseteq\overline{R_B^k}(X)\cap\overline{R_B^k}(Y)$

$(L_6)$ $\underline{R_B^k}(X\cap Y)\subseteq\underline{R_B^k}(X)\cap\underline{R_B^k}(Y)$

$(U_6)$ $\overline{R_B^k}(X\cup Y)\supseteq\overline{R_B^k}(X)\cup\overline{R_B^k}(Y)$

$(L_7)$ $l\leqslant k\Rightarrow\underline{R_B^k}(X)\subseteq\underline{R_B^l}(X)$

$(U_7)$ $l\leqslant k\Rightarrow\overline{R_B^l}(X)\subseteq\overline{R_B^k}(X)$

证明

$(L_1)$: 由于 $u\in\overline{R_B^k}(\sim X)\Leftrightarrow\frac{|[u]_{R_B}\cap\sim X|}{|[u]_{R_B}|}>k\Leftrightarrow\frac{|[u]_{R_B}|-|[u]_{R_B}\cap X|}{|[u]_{R_B}|}>k\Leftrightarrow u\in\sim\underline{R_B^k}(X)$, 因此$\underline{R_B^k}(X)=\sim\overline{R_B^k}(\sim X)$。

$(U_1)$: 由$(L_1)$知, $\underline{R_B^k}(\sim X)=\sim\overline{R_B^k}(X)$成立, 故$\overline{R_B^k}(X)=\sim\underline{R_B^k}(\sim X)$。

$(L_2)$, $(U_2)$, $(L_3)$, $(U_3)$可由定义 12直接得证。

$(L_4)$: 对于$\forall u\in U$, 由$X\subseteq Y$可知

$$\frac{|[u]_{R_B}|-|[u]_{R_B}\cap X|}{|[u]_{R_B}|}\geqslant\frac{|[u]_{R_B}|-|[u]_{R_B}\cap Y|}{|[u]_{R_B}|}$$

于是, 当 $u\in\underline{R_B^k}(X)$时, 有

$$\frac{|[u]_{R_B}|-|[u]_{R_B}\cap X|}{|[u]_{R_B}|}\leqslant k$$

故

$$\frac{|[u]_{R_B}|-|[u]_{R_B}\cap Y|}{|[u]_{R_B}|}\leqslant k$$

即 $u\in\underline{R_B^k}(Y)$, 因此, $\underline{R_B^k}(X)\subseteq\underline{R_B^k}(Y)$。

$(U_4)$可由$(L_4)$及$(L_1)$和$(U_1)$得证; $(L_5)$, $(U_5)$, $(L_6)$, $(U_6)$可由$(L_4)$和$(U_4)$直接得证; $(L_7)$, $(U_7)$可由定义 12直接得证。

定义 13 设三元组 $K=(U,A,F)$为一数据库系统, $X\subseteq U$, $B\subseteq A$, $R_B$ 为给定的二元关系, 对于 $k\in(0.5,1]$, 集合 $X$ 关于数据库系统 $K$ 的 $k-$正域 $pos_{R_B}^k(X)$, $k-$负域 $neg_{R_B}^k(X)$, $k-$边界 $bn_{R_B}^k(X)$分别定义为:

$$pos_{R_B}^k(X)=\{u\in U\mid\mu_X^{R_B}(u)\geqslant k\}$$

$$neg_{R_B}^k(X)=\{u\in U\mid\mu_X^{R_B}(u)\leqslant 1-k\}$$

$$bn_{R_B}^k(X)=\{u\in U\mid 1-k<\mu_X^{R_B}(u)<k\}$$

定理 2 设三元组 $K=(U, A, F)$为一数据库系统, $X\subseteq U$, $R_B$ 是任意给定的二元关系, $k\in(0.5, 1]$, 则下述结论成立。

1) $X$ 是 $k$–精确集, 当且仅当 $bn_{R_B}^{k}(X)=\Phi$。

2) $X$ 是 $k$–粗糙集, 当且仅当 $bn_{R_B}^{k}(X)\neq\Phi$。

证明 由定义 12以及定义 13直接可得。

推论 1 设三元组 $K=(U, A, F)$为一数据库系统, $X\subseteq U$, $R_B$ 是任意给定的二元关系, 对于 $k_1$, $k_2$, $k\in(0.5, 1]$, 有

1) 若 $k_1<k$, 则当 $X$ 是 $k$–精确集时亦为 $k_1$–精确集。

2) 若 $k_2>k$, 则当 $X$ 是 $k$–粗糙集时亦为 $k_2$–粗糙集。

若集合 $X$ 对任意的精度系数 $k\in(0.5, 1]$是粗糙的, 则称集合 $X$ 为绝对(强)粗糙集; 若集合 $X$ 不是绝对(强)粗糙集, 则称集合 $X$ 为相对(弱)粗糙集。对每一个相对(弱)粗糙集 $X$, 存在一个对应的 $k_0\in(0.5, 1]$, 使得集合 $X$ 为 $k_0$–精确集。

显然, 由 $k$–正域, $k$–负域, $k$–边界的定义, 下述定理成立。

定理 3 设 $K=(U, A, F)$为一数据库系统, $R_B$ 任意给定的二元关系。对于任意 $X\subseteq U$, $k\in(0.5, 1]$, 有

1) $pos_{R_B}^{k}(X)=neg_{R_B}^{k}(\sim X)$

2) $bn_{R_B}^{k}(X)=bn_{R_B}^{k}(\sim X)$

3) 若 $bn_{R_B}^{k}(X)=\Phi$, 则

$$pos_{R_B}^{k}(X)\cup neg_{R_B}^{k}(X)=U$$

其中 $\sim X=U-X$。

证明 由定义 13直接得证。

定理 4 设 $K=(U, A, F)$为一数据库系统, $B\subseteq A$, $X\subseteq U$, $R_B$ 为给定的二元关系, $k\in(0.5, 1]$, 下述结论成立:

(1a) $\overline{R_B}(X)\subseteq\overline{R_B^{k}}(X)$

(1b) $\overline{R_B^{k}}(X)\subseteq\overline{R_B}(X)$

(2a) $bn_{R_B}^{k}(X)\subseteq bn_{R_B}(X)$

(2b) $neg_{R_B}(X)\subseteq neg_{R_B}^{k}(X)$

证明

(1a) 对于任意的 $u\in\underline{R_B}(X)$, 成立 $[u]_{R_B}\subseteq X$, 由定义 11知 $\mu_X^{R_B}(u)=1\geqslant k$, 故 $u\in\underline{R_B^{k}}(X)$, 即 $\underline{R_B}(X)\subseteq\underline{R_B^{k}}(X)$。

(1b) 由定理 1($U_1$)及 (1a)直接得证。

(2a) 由定义 5, 定义 13及定理 4(1a), (1b)直接得证。

(2b) 同(2a)。

推论 2 设 $K=(U, A, F)$为一数据库系统, $B\subseteq A$, $X\subseteq U$, $R_B$ 为给定的二元关系。当 $k=1$时, 有

(1a) $\underline{R_B}(X)=\underline{R_B^{k}}(X)$

(1b) $\overline{R_B^{k}}(X)=\overline{R_B}(X)$

(2a) $bn_{R_B}^{k}(X)=bn_{R_B}(X)$

(2b) $neg_{R_B}(X)=neg_{R_B}^{k}(X)$

由上可知随着精度系数 $k$的不断变大, $X$ 的正域与负域将缩小, 而边界域将扩大; 反之, 随着精度系数 $k$的不断减小, $X$ 的正域与负域将扩大, 边界域将缩小。特别的, 当 $k=1$时一般关系下基于粗糙隶属函数(度)的程度粗糙集模型就退化为一般关系下的粗糙集模型, 而精度系数 $k$趋于 0.5时, 即 $k\overrightarrow{\ }0.5$, 有以下结论成立:

定义 14 设三元组 $K=(U, A, F)$为一数据库系统, $X\in U$, $B\subseteq A$, $R_B$ 为给定的二元关系。当精度系数 $k$趋于 0.5时, 即 $k\overrightarrow{\ }0.5$, 记集合 $X$ 在关系 $R_B$ 下的绝对边界为

$$bn_{R_B}^{0.5}(X)=\{u\in U\mid\mu_X^{R_B}(u)=\frac{1}{2}\}$$

由上述定义直接可得以下结论成立:

命题 2 设 $K=(U, A, F)$为一数据库系统, $B\subseteq A$, $X\subseteq U$, $R_B$ 为给定的二元关系, 当 $k\overrightarrow{\ }0.5$时, 则有

(1a) $\underline{R_B^{0.5}}(X)=\bigcup^{k}\underline{R_B^{k}}(X)$

(1b) $\overline{R_B^{0.5}}(X)=\bigcap^{k}\overline{R_B^{k}}(X)$

(2a) $bn_{R_B}^{0.5}(X)=\bigcap^{k}bn_{R_B}^{k}(X)$

(2b) $neg_{R_B}^{0.5}(X)=\bigcup^{k}neg_{R_B}^{k}(X)$

例 2(续例 1) 取 $B=\{a_1, a_2, a_3, a_4, a_5\}$, $X=\{u_1, u_3, u_5, u_6\}$, $Y=\{u_2, u_4\}$。取 $k=0.6$时, 可得

(1a) $\underline{R_B^{0.6}}(X)=\{u_1, u_3, u_5, u_6\}$

$$\overline{R_B^{0.6}}(X)=\{u_1,u_2,u_3,u_5,u_6\}$$

(2a) $$\underline{\overline{R_B^{0.6}}}(Y)=\{u_4\}$$

$$\overline{\overline{R_B^{0.6}}}(Y)=\{u_2,u_4\}$$

而由定义 3知:

$$pos_B^{0.6}(X)=\{u_1,u_3,u_5,u_6\}\quad neg_B^{0.6}(X)=\{u_4\},\quad bn_B^{0.6}(X)=\{u_2\}$$

$$neg_B^{0.6}(Y)=\{u_1,u_3,u_5,u_6\}\quad pos_B^{0.6}(Y)=\{u_4\},\quad bn_B^{0.6}(Y)=\{u_2\}$$

因此

$$pos_{R_B}^{k}(X)=neg_{R_B}^{k}(\sim X)\quad bn_{R_B}^{k}(X)=bn_{R_B}^{k}(\sim X)$$

$$\underline{R_B}(X)\subseteq\underline{R_B^{0.6}}(X),\ \overline{R_B^{0.6}}(X)\subseteq\overline{R_B}(X)$$

$$bn_{R_B}^{0.6}(X)\subseteq bn_{R_B}(X),\ neg_{R_B}(X)\subseteq neg_{R_B}^{0.6}(X)$$

若取 $k=1$时, 经计算有以下成立:

$$\underline{R_B}(X)=\underline{R_B^{1}}(X),\ \overline{R_B^{1}}(X)=\overline{R_B}(X)\quad bn_{R_B}^{1}(X)=bn_{R_B}(X),\ neg_{R_B}(X)=neg_{R_B}^{1}(X)$$

## 3 知识的 $k$-粗糙度

集合的不精确性(粗糙性)是由于边界域的存在而引起的, 集合的边界域越大, 其精确(粗糙)性就越低(高), 为了更准确地表达这一点, 本文引入近似粗糙度的概念.

定义 15 设 $K=(U,A,F)$为一数据库系统, $B\subseteq A$, $R_B$是任意给定的二元关系, $k\in(0.5,1]$, 集合 $X$ 基于粗糙隶属函数的上、下近似算子分别为 $\overline{R_B^k}(X)$和$\underline{R_B^k}(X)$, 则称

$$\theta_{R_B}^{k}(X)=1-\frac{|\underline{\overline{R_B^k}}(X)|}{|\overline{R_B^k}(X)|}$$

为 $X$ 关于数据库系统 $K$ 的 $k$-粗糙度。

由上知, $k$-粗糙度 $\theta_{R_B}^{k}(X)$反映了我们对集合 $X$ 认知的不了解程度。

定理 5 设 $K=(U,A,F)$为一数据库系统, $B\subseteq A$, $X\subseteq U$, 若 $R_B$ 为任意给定的一般二元关系, $k\in(0.5,1]$, 则 $\theta_{R_B}^{k}(X)\leqslant\theta_{R_B}(X)$。

证明 由定理 4知

$|\underline{R_B}(X)|\leqslant|\underline{R_B^k}(X)|$且 $|\overline{R_B}(X)|\geqslant|\overline{R_B^k}(X)|$,

故有

$$\frac{|\underline{\overline{R_B}}(X)|}{|\overline{R_B}(X)|}\leqslant\frac{|\underline{\overline{R_B^k}}(X)|}{|\overline{R_B^k}(X)|}$$

根据定义 15, 则有 $\theta_{R_B}^{k}(X)\leqslant\theta_{R_B}(X)$。

推论 3 设三元组 $K=(U,A,F)$为一数据库系统, $X\subseteq U$, $R_B$ 是任意给定的二元关系, $k\in(0.5,1]$, 则有:

1) $\forall X\in U$, 均有 $0\leqslant\theta_{R_B}^{k}(X)\leqslant1$。

2) $X$ 为 $k$-精确集当且仅当 $\theta_{R_B}^{k}(X)=0$。

3) $X$ 为 $k$-粗糙集当且仅当 $\theta_{R_B}^{k}(X)>0$。

定理 6 设 $K=(U,A,F)$为一数据库系统, $B$, $C\subseteq A$ 且满足 $C\subseteq B$, $k\in(0.5,1]$, 则对任意给定的 $X\subseteq U$, $\theta_{R_B}^{k}(X)\leqslant\theta_{R_C}^{k}(X)$成立。

证明 由文献[9]知, 若 $C\subseteq B$, 则 $R_B\subseteq R_C$, 即对任意的 $u\in U$, $[u]_{R_B}\subseteq[u]_{R_C}$成立。根据定义 12知$\underline{R_C^k}(X)\subseteq\underline{R_B^k}(X)$及$\overline{R_B^k}(X)\subseteq\overline{R_C^k}(X)$成立, 结合定义 15, $\theta_{R_B}^{k}(X)\leqslant\theta_{R_C}^{k}(X)$成立。

定理 7 设 $K=(U,A,F)$为一数据库系统, $B\subseteq A$, $X\subseteq U$, $R_B$ 是任意给定的二元关系。对任意的 $0.5\leqslant l\leqslant k\leqslant1$, $\theta_{R_B}^{l}(X)\leqslant\theta_{R_B}^{k}(X)$成立。

证明 由定理 1($U_7$), ($L_7$)及定义 15可证之。

例 3(续例 1) 取 $B=\{a_1,a_2,a_3,a_4\}$, $C=\{a_1,a_3,a_4\}$, $X=\{u_3,u_5,u_6\}$。

取 $k=0.68$时, 经计算知

$$\theta_{R_B}^{0.68}(X)=\frac{1}{2},\ \theta_{R_C}^{0.68}(X)=\frac{2}{3}$$

显然有 $\theta_{R_B}^{0.68}(X)\leqslant\theta_{R_B}^{0.68}(X)$ 成立。

取 $k=0.76$时, 经计算知

$$\theta_{R_B}^{0.76}(X)=\frac{2}{3},\ \theta_{R_C}^{0.76}(X)=\frac{5}{6}$$

显然有 $\theta_{R_B}^{0.76}(X)\leqslant\theta_{R_B}^{0.76}(X)$ 成立, 而且有

$$\theta_{R_B}^{0.68}(X)\leqslant\theta_{R_B}^{0.76}(X),\ \theta_{R_C}^{0.68}(X)\leqslant\theta_{R_C}^{0.76}(X)$$

## 4 程度粗糙集模型的近似依赖与属性约简

定义 16 设三元组 $K=(U,A,F)$为一数据库系统, $X\subseteq U$, $B$, $C\subseteq A$, $k\in(0.5,1]$, 则 $R_C$ 关于 $R_B$ 的 $k$-正域 $pos_{R_B}^{k}(R_C)$, $k$-负域 $neg_{R_B}^{k}(R_C)$分别定

义为

$$pos^{k}_{R_B}(R_C) = \bigcup_{X\in U/R_C} pos^{k}_{R_B}(X)$$

$$neg^{k}_{R_B}(R_C) = U - pos^{k}_{R_B}(R_C) = \bigcup_{X\in U/R_C} neg^{k}_{R_B}(X)$$

定义 17 设三元组 $K=(U, C\cup D, F)$为一数据库系统, $P\subseteq C$ 为条件属性集, $Q\subseteq D$ 为决策属性集, $U/R_P$ 为条件类, $U/R_Q$为决策类, $k\in(0.5, 1]$。定义 $Q$ 关于 $P$的 $k-$依赖为

$$\gamma(R_P, R_Q, k) = \frac{|pos^{k}_{R_P}(R_Q)|}{|U|}$$

近似依赖性是对执行具有精度系数 $k$ 的对象分类能力的评价, 它推广了粗糙依赖性的思想, 但是, 近似依赖不等同于粗糙依赖, 不能解释为属性的函数或部分依赖, 由于它的性质比函数依赖的性质更弱, 如传递性不成立。当精度系数 $k=1$ 时, 一般关系下的近似依赖性就退化为粗糙依赖性。进一步, 若关系为等价的, 则成为基于等价关系的近似依赖性。

推论 4 设三元组 $K=(U, C\cup D, F)$为一数据库系统, $P\subseteq C$ 为条件属性集, $Q\subseteq D$ 为决策属性集, $R_P$, $R_Q$ 为给定的二元关系, $U/R_P$ 为条件类, $U/R_Q$为决策类。$Q$ 关于 $P$ 的 $k-$依赖 $\gamma(R_P, R_Q, k)$是包含度。

令 $P$ 和 $Q$ 分别为条件属性集和决策属性集, 对于 $k\in(0.5, 1]$, 属性子集 $P'\subseteq P$ 关于 $D$ 的 $k-$重要性定义为

$$\sigma^{k}_{PQ}(P') = \gamma(R_P, R_Q, k) - \gamma(R_{P-P'}, R_Q, k)$$

特别的, 对于 $k\in(0.5, 1]$, 若 $P'=\{a\}$时, 属性 $a\in P$关于 $D$ 的 $k-$重要性为

$$\sigma^{k}_{PQ}(a) = \gamma(R_P, R_Q, k) - \gamma(R_{P-\{a\}}, R_Q, k)$$

定义 18 设三元组 $K=(U, A, F)$为一数据库系统, $P, Q\subseteq A$ 为条件、决策属性集, $p\in P$, $R_P$, $R_Q$ 为给定的二元关系。对于 $k\in(0.5, 1]$, 若 $\gamma(R_P, R_Q, k)=\gamma(red(R_{P-\{p\}}, R_Q, k), R_Q, k)$, 则称 $p\in P$ 是不必要的, 否则是必要的。

推论 5 设三元组 $K=(U, C\cup D, F)$为一数据库系统, $P\subseteq C$ 为条件属性集, $Q\subseteq D$ 为决策属性集, $R_P$, $R_Q$ 为给定的二元关系。上述定义的重要度 $\sigma^{k}_{PQ}(P')$是包含度。

定义 19 设三元组 $K=(U, A, F)$为一数据库系统, $P, Q\subseteq A$ 为条件、决策属性集, $p\in P$, $R_P$, $R_Q$ 为给定的二元关系, 对于 $k\in(0.5, 1]$, 则条件属性集 $P$ 关于决策属性集 $Q$ 的 $k-$近似约简定义为条件属性集 $P$的一个最小属性子集 $red(R_P, R_Q, k)$, 其进一步满足:

1) $\gamma(R_P, R_Q, k)=\gamma(red(R_P, R_Q, k), R_Q, k)$

2) $\gamma(R_P, R_Q, k)\neq\gamma(red(R_{P-\{p\}}, R_Q, k), R_Q, k)$

综上所述, 引入精度系 $k(k\in(0.5, 1])$后, 不但扩充了经典粗糙集理论, 而且更好地体现了数据分析中的数据相关性。

例 4(续例 1) 取 $k=0.75$, $B=\{a_1, a_4\}$, 即可产生该数据库系统上的优势关系 $R_B$, 即 $R_B=\{(u_i, u_j)\in U\times U \mid f_l(u_j)\geqslant f_l(u_i)\ (\forall a_l\in B)\}$. 同理, 取 $C=\{a_1, a_2, a_3\}$和 $D=\{a_5\}$时, 分别产生该数据库系统上的优势关系 $R_C$, $R_D$, 其对论域的划分如下:

$$R_B=\{\{a_1, a_2, a_3, a_4, a_5, a_6\}, \{a_2, a_6\}, \{a_2, a_3, a_5, a_6\}, \{a_3, a_4, a_5, a_6\}, \{a_3, a_5, a_6\}, \{a_2, a_6\}\}$$

$$R_C=\{\{a_1, a_2, a_5, a_6\}, \{a_2, a_6\}, \{a_2, a_3, a_4, a_5, a_6\}, \{a_4, a_6\}, \{a_5\}, \{a_6\}\}$$

$$R_D=\{\{a_1, a_2, a_3, a_4, a_5, a_6\}, \{a_3, a_5, a_6\}, \{a_1, a_2, a_3, a_4, a_5, a_6\}, \{a_3, a_4, a_5, a_6\}, \{a_5, a_6\}, \{a_5, a_6\}\}$$

则 $R_B$ 关于 $R_C$ 的 0.75-正域为

$$pos^{0.75}_{R_B}(R_C) = \bigcup_{X\in U/R_C} pos^{k}_{R_B}(X) = \{a_1, a_2, a_3, a_4, a_5, a_6\}$$

$R_B$ 关于 $R_C$ 的 0.75-负域为

$$pos^{0.75}_{R_B}(R_C) = U - pos^{k}_{R_B}(R_C) = \Phi$$

决策属性集 $D$ 关于条件属性集 $C$ 的 0.75-依赖为

$$\gamma(R_C, R_D, 0.75) = \frac{|pos^{k}_{R_P}(R_Q)|}{|U|} = 1$$

取 $C'=\{a_1\}\subseteq C$, 则

$$\sigma^{0.75}_{CD}(C') = \gamma(R_C, R_D, 0.75) - \gamma(R_{a_2 a_3}, R_D, 0.75) = 1 - \frac{|pos^{0.75}_{\{a_2 a_3\}}(R_D)|}{|U|} = 0$$

即属性 $a_1$ 是不必要的。

## 5 结束语

Pawlak Z 所建立的粗糙集模型,因其思想新颖,处理问题的方法独特,受到了普遍关注并在许多领域得到了成功应用,但是它是以等价关系为基础产生知识的划分,这使得在处理不以等价关系为基础产生知识的数据库系统中的问题时受到了一定的局限。为此,在一般二元关系的基础上,通过引入精度系数 $k(k\in(0.5,1])$,给出了一般关系下基于粗糙隶属函数的程度粗糙集,并进一步讨论了所给模型的相关重要性质。通过与经典粗糙集模型相比较,发现该模型不仅是对经典粗糙集模型的拓展,还是对基于等价关系的变精度粗糙集模型的拓展。这些结论对粗糙集模型的拓展研究以及数据库系统中知识发现的研究工作奠定了一定的理论基础。

## 参考文献:


[1] Pawlak Z. Rough sets [J]. International Journal of computer and information sciences, 1982, 11(5): 341-356.

[2] Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Date [M]. Boston: Kluwer Academic Publishers, 1991.

[3] Ziarko W. Variable precision rough model [J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59.

[4] Zhang W X, Leuang Y, Wu W Z. Information system and knowledge discovery [M]. Beijing: science Press, 2003.

[5] Bonikowski Z, Bryniarski E, Wybraniec U. Extensions and intentions in rough set theory [J]. Inform Sci, 1998, 107: 149-167.

[6] Shi En-Wei. Some properties of the indiscernibility relation in rough set [J]. Chinese Science Bulletin, 1990, 35(4): 338-341.

[7] Liang J Y, Shi Z Z. The information entropy, rough entropy and knowledge granulation in rough set theory [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based System, 2004, 12(1): 37-46.

[8] Leuang Y, Wu W Z, Zhang W X. Knowledge acquisition in incomplement information system: a rough set approach [J]. European Journal of operational Research, 2006, 168(1): 164-180.

[9] Shao M W, Zhang W X. Dominance relation and rules in an incomplete ordered information system [J]. International Journal of Intelligent Systems, 2005, 20: 13-27.

[10] Xu W H, Zhang W X. Measuring roughness of generalized rough sets introduced by a covering [J]. Fuzzy Sets and Systems, 2007, 158: 2443-2455.

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

[12] Pawlak Z. Rough sets approach to multi-attribute decision analysis [J]. European Journal of operational Research, 1994, 72: 443-459.


(责任编辑 刘 舸)

徐伟华,男,1979年5月生于山西浑源,西安交通大学博士,副教授,硕士生导师。2007年毕业于西安交通大学理学院,师从张文修教授,获应用数学专业理学博士学位。现任职于重庆理工大学数理学院应用数学系。任职以来,多次指导学生参加全国大学生数学建模竞赛和全国大学生数学竞赛,并分别获重庆市一等奖和优秀奖等。2009年1月应香港中文大学邀请进行为期3个月的学术访问研究,并得到合作单位的好评。2010年4月进入西安交通大学管理学院,进行为期2年的博士后研究工作。主要研究方向为粗糙集、模糊集、人工智能的数学基础、数据挖掘、知识发现、遥感图像处理等,在《Fuzzy Sets and Systems》《Applied Soft Computing》《Information Sciences》《Journal of Applied Mathematics and Computing》《International Journal of Business Intelligence and Data Mining》《工程数学学报》《模糊系统与数学》《计算机工程》等国内外重要期刊发表学术论文近40篇,其中被三大索引检索十余篇(次)。