变精度覆盖近似算子与覆盖近似算子的关系

梁俊奇[1,2] 张文君[3]
(商丘师范学院数学系 商丘 476000)[1] (武汉大学数学与统计学院 武汉 430072)[2]
(商丘技师学院 商丘 476000)[3]

摘 要 变精度覆盖粗糙集模型是在放宽了覆盖标准的前提下给出的,因而导致近似算子发生了变化,但其变化有一定的规律。在介绍了覆盖粗糙集模型和变精度覆盖粗糙集模型的概念的基础上,给出并证明了变精度覆盖粗糙近似算子与覆盖粗糙近似算子之间的关系,即定理 1、定理 2 及其推论。

关键词 覆盖粗糙集模型,变精度覆盖粗糙集模型,β覆盖近似算子,覆盖近似算子,关系

中图法分类号 TP18 **文献标识码** A

Relations between for Variable Precision Covering Approximation Operators and Covering Approximation Operators

LIANG Jun-qi[1,2] ZHANG Wen-jun[3]
(Department of Mathematics, Shangqiu Normal University, Shangqiu 476000, China)[1]
(School of Mathematics & Statistics, Wuhan University, Wuhan 430072, China)[2]
(Shangqiu Technician College, Shangqiu 476000, China)[3]

Abstract The model of variable precision coverage rough set was proposed under certain conditions of relaxing coverage standard which yields diversification of the approximation operators, and the change is in certain laws in addition. After introducing the concepts on models of coverage rough set and also of variable precision coverage rough set, we proved the relationships of approximation operators between them. That is, we achieved the results in theorem 1, theorem 2 and related corollaries.

Keywords Covering rough set model, Variable precision covering rough set model, β covering approximation operators, Covering approximation operators, Relations

1 引言

粗糙集(rough sets)理论是一种新的处理模糊性和不确定性知识的数学工具,1982 年由波兰数学家 Z. Pawlak[1,2] 首次提出以来,经过二十余年的研究,已经在理论和实际应用上取得了长足的发展,特别是由于 20 世纪 80 年代末和 90 年代初在知识发现、决策分析等领域的成功应用而受到了国际上的广泛关注。目前,它已经在人工智能、知识发现、模式识别与分类、故障检测等方面得到了普遍应用。

粗糙集理论将分类与知识联系在一起,根据已知知识自身的不可分辨关系,通过一对近似算子,对某一给定的概念进行近似表示,它是一种数据驱动的方法,本质上不需要任何关于数据和相应问题以外的先验知识,因此特别适合应用于知识发现(KDD)与数据挖掘(DM)领域。Z. Pawlak 粗糙集模型的一个局限性就是它所处理的分类必须是完全正确或肯定的,因而它的分类是精确的,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”。Z. Pawlak 粗糙集模型的另一个局限性是它所处理的对象是已知的,且从模型中得到的结论仅适合这些对象。但在实际应用中,往往需要把小规模对象集中得到的结论应用到大规模对象集上去。另外,有些实际问题的分类也不一定要求完全精确。为了克服这些局限性,Ziarko 提出了变精度粗糙集模型[3,4],它的基本思想是在 Z. Pawlak 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。显然,当 $\beta=0$ 时,变精度粗糙集模型就退化为 Z. Pawlak 粗糙集模型。这种推广的模型有利于从看似不相关的数据中发现潜在的相关数据。同样的思路,我们也可以把覆盖粗糙集模型推广为变精度覆盖粗糙集模型。

本文主要讨论变精度覆盖粗糙集模型与覆盖粗糙集模型的近似算子之间的关系。

2 覆盖粗糙集模型

定义 1[5,6] 设 U 是有限非空论域,C 是 U 的一个子集族,即 $C\subseteq P(U)$,如果 C 中任意一个子集非空且 $\bigcup C=U$,则 C 是 U 的一个覆盖,称 (U,C) 为一个覆盖近似空间。

到稿日期:2010-04-27 返修日期:2010-10-06 本文受河南省自然科学研究基金资助项目(094300510062),河南省教育厅自然科学基金项目(2008B120006)资助。

梁俊奇(1958—),男,博士生,教授,主要研究方向为智能计算与不确定性信息处理。

定义 2[5,6] 设(U,C)为一个覆盖近似空间，对任意 $x\in U$，称 $N(x)=\cap\{K\in C|x\in K\}$为 x 的邻域。

定义 3[5,6] 设(U,C)为一个覆盖近似空间，对任意 $X\subseteq U$，X 关于(U,C)的覆盖下近似算子和覆盖上近似算子分别定义为

$\underline{C}(X)=\{x\in U|N(x)\subseteq X\}$

$\overline{C}(X)=\{x\in U|N(x)\cap X\neq\phi\}$

很显然，当 C 是 U 的一个划分时，$N(x)$就是 x 所在等价类，它就是 Z. Pawlak 粗糙集模型。

3 变精度覆盖粗糙集模型

定义 4[6] 令

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

式中，$|\cdot|$表示集合·的基数，则称它为集合 X 关于集合 Y 的相对错误分类率。

定义 5[6] 令 $0\leqslant\beta<0.5$，若 $c(X,Y)\leqslant\beta$，则称 Y 多数包含 X，记做 $Y\overset{\beta}{\supseteq}X$。"多数包含"的要求是 X 与 Y 的公共元素的数目大于 X 中元素数目的 50%。

定义 6[6] 设(U,C)为一个覆盖近似空间，对任意 $X\subseteq U$，$0\leqslant\beta<0.5$，X 关于(U,C)的 β 覆盖下近似算子和 β 覆盖上近似算子分别定义为

$\underline{C}_\beta(X)=\{x\in U|c(N(x),X)\leqslant\beta\}$

$\overline{C}_\beta(X)=\{x\in U|c(N(x),X)\leqslant 1-\beta\}$

β 覆盖边界域为

$bnc_{\beta}(X)=\{x\in U|\beta<c(N(x),X)<1-\beta\}$

β 覆盖负域为

$negc_{\beta}(X)=\{x\in U|c(N(x),X)\geqslant 1-\beta\}$

直观上说，β 覆盖下近似可理解为将 U 中的对象以它的邻域关于 X 不大于 β 的错误分类率分于 X 的集合。β 覆盖上近似是由那些以它的邻域关于 X 不大于 β 的错误分类率不能分类于 $\sim X$ 的 U 中的对象所构成的集合。

4 变精度覆盖近似算子与覆盖近似算子的关系

定理 1 设(U,C)为一个覆盖近似空间，对任意 $X\subseteq U$，$0\leqslant\beta_1\leqslant\beta_2<0.5$，则有下列近似算子之间的关系成立：

(1)$\underline{C}(X)\subseteq\underline{C}_{\beta_1}(X)\subseteq\underline{C}_{\beta_2}(X)\subseteq\overline{C}_{\beta_2}(X)\subseteq\overline{C}_{\beta_1}(X)\subseteq\overline{C}(X)$

(2)$bnc_{\beta_2}(X)\subseteq bnc_{\beta_1}\subseteq bn_C(X)$，$negc_{\beta_2}(X)\supseteq negc_{\beta_1}(X)\supseteq neg_C(X)$

式中，$bn_C(x)$，$neg_C(x)$是覆盖粗糙集模型的边界域和负域。

证明：(1) $\forall x\in\underline{C}(X)\Rightarrow N(x)\subseteq X$

$\Rightarrow c(N(x),X)=0\Rightarrow c(N(x),X)\leqslant\beta_1$

$x\in\underline{C}_{\beta_1}(X)\Rightarrow c(N(x),X)\leqslant\beta_1\Rightarrow c(N(x),X)\leqslant\beta_2$

$\Rightarrow x\in\underline{C}_{\beta_2}(X)\Rightarrow x\in\overline{C}_{\beta_2}(X)$

$\Rightarrow c(N(x),X)<1-\beta_2\Rightarrow c(N(x),X)<1-\beta_1\Rightarrow x\in\overline{C}_{\beta_1}(X)$

$\Rightarrow c(N(x),X)<1-\beta_1$

$\Rightarrow c(N(x),X)<1\Rightarrow N(x)\cap X\neq\phi\Rightarrow x\in\overline{C}(X)$

即结论成立。

(2) $\forall x\in bnc_{\beta_2}(X)$

$\Rightarrow\beta_2<c(N(x),X)<1-\beta_2$

$\Rightarrow\beta_1<c(N(x),X)<1-\beta_1$

$\Rightarrow x\in bnc_{\beta_1}(X)$

$\Rightarrow 0<c(N(x),X)<1\Rightarrow N(x)\not\subset X,N(x)\cap X\neq\phi$

$\Rightarrow x\in bn_C(X)$

即第一式成立。另一式类似可证。

注 1：定理 1 告诉我们，(1)β 覆盖下近似算子相对分类误差 β 具有单调递减性，β 覆盖上近似算子相对分类误差 β 具有单调递增性。(2)β 覆盖边界域随着分类误差 β 的减小而增大，β 覆盖负域却随着分类误差 β 的减小而减小。

注 2：定理 1 的逆命题也是成立的。

定理 2 设(U,C)为一个覆盖近似空间，对于任意 $X\subseteq U$，$0\leqslant\beta<0.5$，则

$\lim\limits_{\beta\to 0.5^-}\underline{C}_\beta(X)=\underline{C}_{0.5}(X)=\{x\in U|c(N(x),X)<0.5\}$

$\lim\limits_{\beta\to 0.5^-}\overline{C}_\beta(X)=\overline{C}_{0.5}(X)=\{x\in U|c(N(x),X)\leqslant 0.5\}$

$\lim\limits_{\beta\to 0.5^-}bnc_{\beta}(X)=bnc_{0.5}(X)=\{x\in U|c(N(x),X)=0.5\}$

$\lim\limits_{\beta\to 0.5^-}negc_{\beta}(X)=negc_{0.5}(X)=\{x\in U|c(N(x),X)>0.5\}$

定理 2 的证明是显然的。

推论 1 设(U,C)为一个覆盖近似空间，对任意 $X\subseteq U$，$0\leqslant\beta<0.5$，则

$\lim\limits_{\beta\to 0}\underline{C}_\beta(X)=\underline{C}_0(X)=\underline{C}(X)=\{x\in U|N(x)\subseteq X\}$

$\lim\limits_{\beta\to 0}\overline{C}_\beta(X)=\overline{C}_0(X)=\overline{C}(X)=\{x\in U|N(x)\cap X\neq\phi\}$ $\lim\limits_{\beta\to 0}bnc_{\beta}(X)=bnc_0(X)=bn_c(x)$

$\lim\limits_{\beta\to 0}negc_{\beta}(X)=negc_0(X)=neg_C(x)$

推论 2 设(U,C)为一个覆盖近似空间，对于任意 $X\subseteq U$，$0\leqslant\beta<0.5$，则

$\underline{C}_{0.5}(X)=\bigcup\limits_\beta\underline{C}_\beta(X)\quad\overline{C}_{0.5}(X)=\bigcap\limits_\beta\overline{C}_\beta(X)$

$bnc_{0.5}(X)=\bigcap\limits_\beta bnc_{\beta}(X)\quad negc_{0.5}(X)=\bigcup\limits_\beta negc_{\beta}(X)$

参 考 文 献

[1] Pawlak Z. Roughsets[J]. International Journal of Computer and Information Science, 1982, 11: 341-356

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

[3] Ziarko W. Variable precision Rough sets model[J]. Joumal of Computer System Science, 1993, 46(1): 39-59

[4] 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，2001

[5] 杨勇，朱晓钟，李廉. 覆盖粗糙集的公理化[J]. 计算机科学，2009, 36(5): 181-182

[6] 刘瑞新，孙士保，秦克云. 变精度覆盖粗糙集[J]. 计算机工程与应用，2008, 44(12): 47-50