

关于覆盖粗糙集模型性质的一个注记

梁俊奇^{1,2} 闫淑霞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

(郑州师范学院数学系 郑州 450044)³

摘 要 对覆盖粗糙集模型的部分性质进行了推广,即通过引进一对新的算子,把并与交的上下近似集之包含关系推广到了相等关系,从而得到了更好的结果。

关键词 覆盖粗糙集,近似算子,性质,推广

中图法分类号 O159 **文献标识码** A

Note of Covering Rough Set Model Nature

LIANG Jun-qi^{1,2} YAN Shu-xia³

(Department of Mathematics, Shangqiu Normal University, Shangqiu 476000, China)¹

(School of Mathematics & Statistics, Wuhan University, Wuhan 430072, China)²

(Department of Mathematics, Zhengzhou Normal University, Zhengzhou 450044, China)³

Abstract We promoted part nature of covering rough set model, introduced a pair of new operator and extended the inclusion relation of join and intersection of the upper and lower approximation sets, then got some better results.

Keywords Covering rough sets, Approximation operators, Nature, Promotion

1 引言与预备

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^[1,2],1982年由波兰华沙理工大学数学家 Z. Pawlak^[1]首次提出,经过30余年的研究与发展,已经在理论和实际应用上均取得了长足的发展。目前,它已经在人工智能、知识与数据发现、模式识别与分类、故障检测等方面得到了成功的应用。

众所周知,在 Z. Pawlak 粗糙模型中,论域上的等价关系起着至关重要的作用,基于等价关系形成的划分,构造了论域上的上、下近似算子,用于刻画不精确概念,进而研究相应的知识约简与知识获取问题。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没有等价关系。为了推广粗糙理论及其应用的范围,根据具体问题,人们对 Z. Pawlak 粗糙模型进行了多种形式的推广,相继出现了变精度粗糙模型、程度粗糙模型、模糊粗糙模型、粗糙模糊模型、基于一般二元关系的粗糙模型、基于覆盖理论的粗糙模型^[4]等。本文在文献[3]的基础上,进一步讨论覆盖粗糙集模型性质的一个推广。

定义 1^[5] 设 U 是非空有限论域, C 是 U 的子集类,如果 C 中的元素非空且 $\bigcup C = U$, 则 C 是 U 的一个覆盖,序对 (U, C) 称覆盖近似空间。

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

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

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

式中, $\underline{C}(X)$ 为 X 在该近似空间上的下近似, $\overline{C}(X)$ 为 X 在该近似空间上的上近似。若 $\overline{C}(X) = \underline{C}(X)$, 则称 X 为 (U, C) 上可定义的,也称 X 为可定义集;否则, X 为 (U, C) 上不可定义的,或称 X 为不可定义集。

很显然,当 C 是 U 的一个划分时, $N(x)$ 就是 x 的所在等价类,它就是 Z. Pawlak 粗糙集模型。因此,覆盖粗糙集是 Z. Pawlak 粗糙集的一种推广。

2 覆盖粗糙集模型性质的一个注记

定理 1^[5] 设 (U, C) 为覆盖近似空间,对任意的 $X, Y \subseteq U$, 则覆盖上下近似满足

$$(1L) \underline{C}(\emptyset) = \emptyset, (1U) \overline{C}(\emptyset) = \emptyset;$$

$$(2L) \underline{C}(U) = U, (2U) \overline{C}(U) = U;$$

$$(3L) \underline{C}(X) \subseteq X, (3U) \overline{C}(X) \supseteq X;$$

$$(4L) \underline{C}(X \cap Y) = \underline{C}(X) \cap \underline{C}(Y),$$

$$(4U) \overline{C}(X \cup Y) = \overline{C}(X) \cup \overline{C}(Y);$$

$$(5L) X \subseteq Y \Rightarrow \underline{C}(X) \subseteq \underline{C}(Y),$$

$$(5U) X \subseteq Y \Rightarrow \overline{C}(X) \subseteq \overline{C}(Y);$$

$$(6L) \underline{C}(\underline{C}(X)) = \underline{C}(X),$$

$$(6U)\bar{C}(\bar{C}(X))=\bar{C}(X);$$

$$(7)\bar{C}(\sim X)=\sim \bar{C}(X), \bar{C}(\sim X)=\sim \bar{C}(X).$$

上述性质的证明见文献[5]。下面证明比(4L)和(4U)稍弱些的相对应的性质,即

$$(4L_1)\underline{C}(XUY)\supseteq \underline{C}(X)\cup \underline{C}(Y),$$

$$(4U_1)\bar{C}(X\cap Y)\subseteq \bar{C}(X)\cap \bar{C}(Y).$$

证明: $X\subseteq XUY, Y\subseteq XUY\Rightarrow \underline{C}(X)\subseteq \underline{C}(XUY), \underline{C}(Y)\subseteq \underline{C}(XUY)\Rightarrow \underline{C}(X)\cup \underline{C}(Y)\subseteq \underline{C}(XUY), X\cap Y\subseteq X, X\cap Y\subseteq Y\Rightarrow \bar{C}(X\cap Y)\subseteq \bar{C}(X), \bar{C}(X\cap Y)\subseteq \bar{C}(Y)\Rightarrow \bar{C}(X\cap Y)\subseteq \bar{C}(X)\cap \bar{C}(Y).$

然而,由于(4L₁)和(4U₁)是包含关系,因此不能得到确切的结果或好的性质。本文试图把它们推广到等式关系。为此,首先给出两个新算子。

定义4 设(U, C)为覆盖近似空间,对于任意集合 $X\subset U, \forall x\in U$,定义两个新集合

$$h_X(x)=N(x)-X$$

$$l_X(x)=N(x)-h_X(x)$$

定义5 设(U, C)为覆盖近似空间,对任意的 $X, Y\subset U$,分别称

$$\underline{Z}_X(Y)=\bigcup\{N(x)|x\in bnc(X)\wedge l_X(x)\not\subset Y\wedge h_X(x)\subseteq Y\}$$

$$\bar{Z}_X(Y)=\bigcup\{N(x)|x\in bnc(X)\wedge l_X(x)\cap Y=\phi\wedge h_X(x)$$

$\cap Y\neq\phi\}$

为X的确定增值算子和不确定增值算子。

利用这两个新算子,可以把(4L₁)和(4U₁)进行推广,即有下面的定理成立:

定理2 设(U, C)为覆盖近似空间,对任意的 $X, Y\subset U$,则有

$$\underline{C}(XUY)=\underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)$$

$$\bar{C}(X\cap Y)=\bar{C}(X)\cap \bar{C}(Y)-\bar{Z}_X(Y)$$

为了证明这个定理,需要下面的引理:

引理1 设(U, C)为覆盖近似空间,对任意的 $X, Y\subset U$,则有

(1) 如果 $l_X(x)\not\subset Y, h_X(x)\subset Y$, 则 $x\in bnc(X)\Leftrightarrow x\in bnc(Y)$;

(2) 如果 $l_X(x)\cap Y=\phi, h_X(x)\cap Y\neq\phi$, 则 $x\in bnc(X)\Leftrightarrow x\in bnc(Y)$ 。

证明:(1)设 $l_X(x)\not\subset Y, h_X(x)\subset Y$, 因为 $x\in bnc(X)\Rightarrow l_X(x)\neq\phi\wedge h_X(x)\neq\phi$, 且

$$l_X(x)\not\subset Y\Rightarrow l_X(x)\cup h_X(x)=N(x)\not\subset \underline{C}(Y)$$

$$h_X(x)\subset Y\Rightarrow l_X(x)\cup h_X(x)=N(x)\not\subset neg_C(Y)$$

$$\Rightarrow x\in bnc(Y);$$

反之,

$$x\in bnc(Y)\Rightarrow l_Y(x)\neq\phi\wedge h_Y(x)\neq\phi$$

且

$$l_X(x)\not\subset Y\Rightarrow l_Y(x)\cup h_Y(x)=N(x)\not\subset neg_C(X)$$

$$h_X(x)\subset Y\Rightarrow l_Y(x)\cup h_Y(x)=N(x)\not\subset \bar{C}(X)$$

$$\Rightarrow x\in bnc(X)$$

(2) 类似(1)的证明方法可证。

引理2 设(U, C)为覆盖近似空间,对任意的 $X, Y\subset U$,则有

$$\underline{Z}_X(Y)=\bigcup\{N(x)|x\in bnc(Y)\wedge h_X(x)\subseteq l_Y(x)\}$$

$$\bar{Z}_X(Y)=\bigcup\{N(x)|x\in bnc(Y)\wedge l_X(x)\subseteq h_Y(x)\}$$

证明:任意的 $N(x)\in \underline{Z}_X(Y)$, 有 $x\in bnc(X), l_X(x)\not\subset Y, h_X(x)\subseteq Y$, 由引理1(1)知, $x\in bnc(Y)$; 若 $h_X(x)\not\subset l_Y(x)$, 则存在 $y\in h_X(x)$, 但 $y\notin l_Y(x)$, 这与 $h_X(x)\subseteq Y$ 矛盾, 所以 $N(x)\in \bigcup\{N(x)|x\in bnc(Y)\wedge h_X(x)\subseteq l_Y(x)\}$, 反之, 任意的 $C_x\in \bigcup\{N(x)|x\in bnc(Y)\wedge h_X(x)\subseteq l_Y(x)\}$, 有 $x\in bnc(Y)$, 且 $h_X(x)\subseteq l_Y(x)$, 进而有 $h_X(x)\subseteq Y$ 。进一步证明 $l_X(x)\not\subset Y$ 。

事实上,若 $l_X(x)\subseteq Y$, 则 $l_X(x)\cup h_X(x)=N(x)\subseteq Y$, 这与 $x\in bnc(Y)$ 矛盾, 所以 $N(x)\in \underline{Z}_X(Y)$ 。

综合上述,有

$$\underline{Z}_X(Y)=\bigcup\{N(x)|x\in bnc(Y)\wedge h_X(x)\subseteq l_Y(x)\}$$

同理可证

$$\bar{Z}_X(Y)=\bigcup\{N(x)|x\in bnc(Y)\wedge l_X(x)\subseteq h_Y(x)\}$$

下面给出定理2的证明。

(1)任意的 $x\in \underline{C}(XUY)$, 存在 $N(x)$, 使 $N(x)\subseteq XUY$, 所以 $N(x)\subseteq X$ 或 $N(x)\subseteq Y$ 或当 $x\in bnc(Y), h_X(x)\subseteq l_Y(x)$, 于是 $N(x)\subseteq \underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)$, 这就得到 $\underline{C}(XUY)\subseteq \underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)$; 反之, 任意的 $x\in \underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)$, 存在 $N(x)$, 若 $N(x)\subseteq \underline{C}(X)$ 或 $N(x)\subseteq \underline{C}(Y)$, 则 $N(x)\subseteq XUY$, 于是 $N(x)\subseteq \underline{C}(XUY)$; 若 $N(x)\subseteq \underline{Z}_X(Y)$, 则 $l_X(x)\subseteq X, h_X(x)\subseteq Y$, 所以 $N(x)=l_X(x)\cup h_X(x)\subseteq XUY$, 于是 $N(x)\subseteq \underline{C}(XUY)$, 这就得到 $\underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)\subseteq \underline{C}(XUY)$ 。

综合上述,有 $\underline{C}(XUY)=\underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)$ 。

(2)任意的 $x\in \bar{C}(X\cap Y)$, 则存在 $N(x)$ 且

$$N(x)\cap (X\cap Y)\neq\phi\Rightarrow N(x)\cap X\neq\phi, N(x)\cap Y\neq\phi, l_X(x)\cap Y\neq\phi\Rightarrow N(x)\not\subset \bar{Z}_X(Y)$$

因此,

$$\bar{C}(X\cap Y)=\bar{C}(X)\cap \bar{C}(Y)-\bar{Z}_X(Y).$$

3 主要结论的实例验证

例 设 $U=\{x_1, x_2, \dots, x_8\}$ 为有限论域, $C=\{C_1, C_2, C_3, C_4\}$ 是U的一个覆盖, 其中 $C_1=\{x_1, x_4, x_8\}, C_2=\{x_2, x_4, x_5, x_7\}, C_3=\{x_3\}, C_4=\{x_6\}$ 。

先取 $X=\{x_1, x_4\}, Y=\{x_2, x_5, x_8\}$, 则

$$\underline{C}(XUY)=\{x_1, x_4, x_8\}, \underline{C}(X)=\phi$$

$$\underline{C}(Y)=\phi$$

$$\bar{C}(Y)=\{x_1, x_2, x_4, x_5, x_7, x_8\}$$

$$bnc(Y)=\{x_1, x_2, x_4, x_5, x_7, x_8\}$$

经过计算,得知

$$h_X(x_1)=h_X(x_4)=h_X(x_8)=\{x_8\}$$

$$h_X(x_2)=h_X(x_4)=h_X(x_5)=h_X(x_7)=\{x_2, x_5, x_7\}$$

$$l_Y(x_1)=l_Y(x_4)=l_Y(x_8)=\{x_1, x_4, x_8\}$$

$$l_Y(x_2)=l_Y(x_4)=l_Y(x_5)=l_Y(x_7)=\{x_2, x_5\}$$

由引理2知, $\underline{Z}_X(Y)=\{x_1, x_4, x_8\}$, 从而得知

$$\underline{C}(XUY)=\underline{C}(X)\cup \underline{C}(Y)\cup \underline{Z}_X(Y)$$

再取 $X=\{x_1, x_4\}, Y=\{x_3, x_4\}$, 则

$$\bar{C}(X\cap Y)=\{x_1, x_2, x_4, x_5, x_7, x_8\}$$

$$\underline{C}(X)=\phi$$

$$\bar{C}(X)=\{x_1, x_2, x_4, x_5, x_7, x_8\}$$

$$\underline{C}(Y)=\phi$$

$$\bar{C}(Y)=\{x_1, x_2, x_3, x_4, x_5, x_7, x_8\}$$

$$bn_C(Y) = \{x_1, x_2, x_3, x_4, x_5, x_7, x_8\}$$

经过计算得知

$$h_Y(x_1) = h_Y(x_8) = \{x_1, x_8\}$$

$$h_Y(x_3) = \phi \quad h_Y(x_4) = \phi$$

$$h_Y(x_2) = h_Y(x_5) = h_Y(x_7) = \{x_2, x_5, x_7\}$$

$$l_X(x_1) = \{x_1, x_4\}, l_X(x_2) = \{x_4\}$$

$$l_X(x_3) = \phi \quad l_X(x_4) = \phi$$

$$l_Y(x_5) = \{x_4\} \quad l_Y(x_7) = \{x_4\}$$

$$l_Y(x_8) = \{x_4\}$$

由引理 2 知, $\bar{Z}_X(Y) = \{x_3\}$, 从而得知

$$\bar{C}(X \cap Y) = \bar{C}(X) \cap \bar{C}(Y) - \bar{Z}_X(Y).$$

参考文献

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

- [2] Pawlak Z. Rough sets: Theoretical Aspects of Reasoning about Data[J]. Kluwer Academic Publishers, Boston, 1991
- [3] Zhang Hua-guang, Liang Hong-li, et al. Two new operators in rough set theory with applications to fuzzy sets[J]. Information Sciences, 2004, 166: 147-165
- [4] 徐忠印, 王勤. 覆盖粗糙集模型的性质[J]. 河南师范大学学报: 自然科学版, 2005, 33(1): 130-132
- [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
- [7] 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
- [8] 苗夺谦, 王国胤, 等. 粒计算: 过去、现在与展望[M]. 北京: 科学出版社, 2007

(上接第 223 页)

SSLMK 的分类正确率略好于 MI-Kernel, 我们认为这是由于数据集中的数据的分布规律性并不突出, 导致半监督学习的效果不明显, 特别是 boots 数据集的图像中同时包含主题(boots)和非主题(如模特、背景、装饰物等)都会对算法的分类正确率有较大的影响, 且同一主题的图像之间的分布未能呈现明显的规律性, 因此本文的算法所计算的矩阵 M 不能很好地反映真实的数据分布。

结束语 本文讨论了利用多示例学习算法解决图像主题识别的问题, 其利用多示例学习中的核学习方法, 把每一个图像看作多示例学习框架中的包, 进而定义包之间的核函数, 把包看作一个 KNN 图或 ϵ -Ball 图, 由图核的定义得出顶点核和边核, 并利用图的光滑性理论推导出半监督多示例核, 能够利用未标记包所蕴含的数据分布规律改善学习器的性能。这是一个把半监督核应用到多示例学习问题的尝试, 能够有效解决标记样本获取代价大的问题, 并能通过已有未标记样本最大限度提升学习器的泛化能力, 是一种完全的半监督学习方法。

在真实的数码照片数据集和 ALOI 对象数据集上进行算法的性能测试, 实验结果表明本文算法对于图像的主题识别问题是有效的, 相比以往的算法有较高的正确率和泛化能力, 且引入未标记的示例能有效提高学习器的性能和泛化能力, 可以推广到图像主题识别的其它方面。

算法的一个核心问题是如何在所定义的边核和点核中结合反映数据流形光滑程度的度量, 我们采用的方法是使用图的拉普拉斯矩阵, 但由于处理的是多示例学习问题, 每个包中有一定数量的实例, 而本文的拉普拉斯的计算方法是针对单示例学习问题的, 使用包中心作为代表示例的可行性也值得研究, 这将是后续研究工作的重点之一。

参考文献

- [1] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple-instance problem with axis-parallel rectangles[J]. Artificial Intelligence, 1997, 89(1/2): 31-71
- [2] Zhou Zhi-hua. Multi-Instance Learning: A Survey[R]. CS, Nanjing University, 2004

- [3] Jia Yang-qing, Zhang Chang-shui. Instance-level Semisupervised Multiple Instance Learning[C]// Proceedings of the 23rd Intl. Conf. AAAI, 2008: 640-645
- [4] Rahmani R, Goldman S A. MISSL: Multiple-instance semi-supervised learning[C]// Proceedings of the 23rd Intl. Conf. ICML, 2006: 705-712
- [5] Maron O, Lozano-Pérez T. A framework for multiple-instance learning[J]. Neural Information Processing Systems, 1998
- [6] Gartner T, Flach P A, Kowalczyk A, et al. Multi-instance kernels[C]// Proceedings of the 19th Intl. Conf. ICML, 2002: 179-186
- [7] Sindhwani V, Niyogi P, Belkin M. Beyond the point cloud: from transductive to semi-supervised learning[C]// Proceedings of the 22nd Intl. Conf. ICML, 2005: 824-831
- [8] Zhou Zhi-hua, Sun Yu-yin, Li Yu-feng. Multi-instance learning by treating instances as non-I. I. D. samples[C]// Proceedings of the 26th Intl. Conf. ICML, 2009: 1249-1256
- [9] Zhu X, Ghahramani Z, Lafferty J. Semisupervised learning using gaussian fields and harmonic functions[C]// Proceedings of the 20th Intl. Conf. ICML, 2003
- [10] Zhou Zhi-hua, Xu Jun-ming. On the relation between multi-instance learning and semi-supervised learning[C]// Proceedings of the 24th Intl. Conf. ICML, 2007: 1167-1174
- [11] Niyogi P. Manifold regularization and semi-supervised learning: Some theoretical analysis[R]. Department of Computer Science, University of Chicago, 2008
- [12] Chen Y, Wang J Z. Image categorization by learning and reasoning with regions[J]. Journal of Machine Learning Research, 2004, 5: 913-939
- [13] Geusebroek J M, Burghouts G J, Smeulders A W M. The Amsterdam library of object images[J]. International Journal of Computer Vision, 2005, 61: 103-112
- [14] Sindhwani V, Rosenberg D S. An RKHS for multi-view learning and manifold co-regularization[C]// Proceedings of the 25th Intl. Conf. ICML, 2008
- [15] 刘叶青, 刘三阳, 谷明涛. 一种多项式光滑的半监督支持向量机分类算法[J]. 计算机科学, 2009, 36(7): 179-181