

文章编号: 1001-0920(2009)03-0464-04

一种新的粗糙集属性约简方法及其应用

关欣^{1,2}, 衣晓¹, 何友¹

(1. 海军航空工程学院 信息融合技术研究所, 山东 烟台 264001;

2. 国防科学技术大学 电子科学与工程学院, 长沙 410073)

摘要: 对粗糙集理论中属性约简问题进行研究, 借助于离散化获得的断点, 提出一种指导属性约简的新方法, 并提出一种改进的连续属性值离散化方法. 以雷达辐射源用途识别为例, 给出了识别实例和计算机仿真实验, 并与工程中常用的统计模式识别方法进行比较, 结果证明了该算法的正确性和有效性.

关键词: 粗糙集; 知识约简; 离散化; 决策表

中图分类号: TP18

文献标识码: A

Knowledge reduction and its applications based on rough set

GUAN Xin^{1,2}, YI Xiao¹, HE You¹

(1. Research Institute of Information Fusion, Naval Aeronautical and Astronautical University, Yantai 264001, China; 2. School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China. Correspondent: GUAN Xin, E-mail: gxtongwin@163.com)

Abstract: A knowledge reduction method based on discretization cuts is proposed. An improved attribute discretization method is also presented in this paper. Examples of recognizing the emitter purpose are selected to demonstrate this new method. Moreover, computer simulation of recognizing the emitter purpose is selected and compared with classical statistical recognition algorithm through the simulation. Experiments results demonstrate the excellent performance of this recognition method.

Key words: Rough set; Knowledge reduction; Discretization; Decision table

1 引言

粗糙集理论是处理模糊和不确定知识的一种数学工具^[1-3]. 在实际应用过程中, 知识库中的知识量非常庞大, 提高系统知识清晰度的一个重要方面是对数据库中的数据进行约简. 粗糙集理论无需提供数据集外的任何先验知识, 利用数据集上的等价关系便可对知识的不确定程度进行度量, 从而避免了对知识的主观评价. 属性约简是粗糙集理论的核心问题之一. 目前, 国内外学者已提出一些有效的属性约简方法^[4-10], 如一般约简算法、基于可辨识矩阵和逻辑运算方法、归纳属性约简算法、基于互信息的约简算法、基于不一致决策表的属性约简方法等. 属性约简方法的选取, 对于分类质量具有很大的影响.

在不同的系统中或在不同的条件下, 人们对属性约简的要求和期望是不一致的. 已有的约简方法

是对离散后的决策表进行属性约简, 一般会得到多种可能的约简结果, 然后根据问题的实际要求选取一个满意的结果. 当采用粗糙集理论进行目标识别时, 如何使识别的效果最好是人们所关心的问题^[11].

本文借助于离散化获得的断点, 提出一种指导属性约简的新方法, 并提出一种改进的连续属性值离散化方法, 该算法可获得更简化和有效的决策规则集. 最后通过对雷达辐射源识别的应用, 验证了该方法的正确性和实用价值.

2 一种改进的离散化方法

粗糙集理论只能处理离散的属性值, 对于实际应用中大量存在的连续属性值, 必须经过离散化后才能用粗糙集方法进行处理. 离散化问题可描述为:

决策表 $S = (U, R, V, f)$. 其中: $R = C \quad D$ 是

收稿日期: 2007-12-28; 修回日期: 2008-07-04.

基金项目: 国家自然科学基金项目(60572161); 全国优秀博士论文专项资金项目(200443); 泰山学者建设工程专项基金项目.

作者简介: 关欣(1978—), 女, 辽宁锦州人, 副教授, 博士后, 从事多传感器信息融合、智能计算的研究; 何友(1956—), 男, 吉林磐石人, 教授, 博士生导师, 从事多传感器信息融合、多目标跟踪等研究.

属性集合,子集 C 和 D 分别为条件属性集和决策属性集; $U = \{x_1, x_2, \dots, x_n\}$ 是有限的对象集合即论域. 设决策种类数为 $r(d)$, 属性 a 的值域 V_a 上的一个断点记为 (a, c) . 其中: $a \in R, c$ 为实数集. 在值域 $V_a = [l_a, r_a]$ 上, 任意一个断点集合 $\{(a, c_1^a), (a, c_2^a), \dots, (a, c_{k_a}^a)\}$ 定义了 V_a 上的一个分类 P_a , 即

$$P_a = \{[c_0^a, c_1^a), [c_1^a, c_2^a), \dots, [c_{k_a}^a, c_{k_a+1}^a]\}, \quad (1)$$

$$l_a = c_0^a < c_1^a < \dots < c_{k_a+1}^a = r_a, \quad (2)$$

$$V_a = [c_0^a, c_1^a) \cup [c_1^a, c_2^a) \cup \dots \cup [c_{k_a}^a, c_{k_a+1}^a]. \quad (3)$$

因此,任意的 $P = \prod_{a \in R} P_a$ 定义了一个新的决策表

$$S^P = (U, R, V^P, f^P),$$

$$f^P(x_a) = i \Leftrightarrow x_a \in [c_i^a, c_{i+1}^a),$$

$$x \in U, i \in \{0, 1, \dots, k_a\}. \quad (4)$$

即经过离散化后,原来的信息系统被一个新的信息系统所代替.

下面给出一种改进的离散化算法:对每个属性 $a \in C$ 进行下面的过程:

步骤 1: 根据属性 $a(x)$ 的值,由小到大排列实例 $x \in U$.

步骤 2: 从上到下扫描,从属性 a 的最小值开始,将重新排列的相邻条件属性值相同的作为一个整体,依次与其他相邻实例进行比较.如果样本属性值与决策值不同,则取两属性值的平均值为一断点.

已有的离散化方法是用最小数目的断点集使属性空间划分的数目最少.实际上,提高分类能力才是最重要的.在高识别率的前提下,运算的时间和空间复杂度大一些是可以接受的.采用监督型局部离散化方法,能使每个条件属性都对决策规则发挥作用.考虑到噪声的影响,这种离散化方法的实际应用效果要好于全局离散化方法.

与 Naive-Scaler 算法^[4] 相比,新算法将属性值相同的作为一个整体来考虑,从而克服了实例顺序的影响.

3 一种新的属性约简方法

考虑到噪声的影响,为使每个条件属性都对决策规则发挥作用,用监督型局部离散化过程中产生的断点指导属性约简,就能直接得到最好的约简结果.具体约简步骤如下:

步骤 1: 逐个计算出每个条件属性各断点两侧属性值之差占该断点值的比例.

步骤 2: 取出每个条件属性得到的最小比例值.

步骤 3: 以步骤 2 所有最小比例值中的最大值所在的条件属性作为属性约简的结果,判断它是否

能实现对决策规则的不可分辨关系表达.如果不能,则继续取步骤 2 剩余数据中最大值所在的条件属性,并与第一次取得的条件属性一起构成属性约简的结果.重复以上过程,直到所得的属性约简结果能表达决策规则的不可分辨关系.

步骤 4: 将步骤 3 得到的属性约简结果进行一般性的属性约简.对于作用相同的条件属性,根据实际情况保留步骤 2 中最大比例值所在的条件属性.

这里的断点是采用监督型局部离散化方法得到的,如用第 2 节改进的固定值离散化方法得到的断点.通过以上过程,便可得到多种可能的约简结果中最好的一种.

4 实验测试与结果分析

4.1 实例分析

下面给出雷达辐射源用途识别的应用实例,用于说明本文提出的属性约简算法的有效性.设雷达特征矢量由 3 个特征参数构成,即射频频率(RF),脉冲重复频率(PRF)和脉冲宽度(PW).从已知雷达知识库中提取 3 类不同用途的雷达,已知样本特征参数如表 1 所示.

表 1 提取的雷达辐射源实例

序号	RF/ MHz	PRF/ Hz	PW/ μ s	类别
1	2774.3	428	3	1
2	1280	300	4	1
3	1313	301	4.1	1
4	1251	601	1.7	1
5	9214	429	0.8	1
6	2746	2873	0.6	1
7	2985	325	2.6	1
8	3109	375	2.1	1
9	2695	335	1.1	2
10	160	190	3.7	2
11	2700	375	1.7	2
12	2700	330	0.8	2
13	1999.9	1600	3.5	2
14	3400	2500	0.5	3
15	2970	1250	0.8	3
16	9000	1750	0.25	3
17	3700	2250	0.37	3

步骤 1:对原始数据进行离散化处理,这里采用第 2 节给出的算法进行离散化.为方便处理,作以下标识: $U = \{x_1, x_2, \dots, x_{17}\}$ —序号, d —类别, a —RF, b —PRF, c —PW.

经过计算得到的断点集如表 2 所示.根据断点集对表 1 离散化的结果如表 3 所示.

步骤 2:消去重复的属性,得到简化后的决策

表 2 计算所得断点集

<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
705.5	245	0.55	2977.5	925.5	1.9
1656.45	327.5	0.7	3254.5	1425	3.25
2723	355	0.95	9107	1675	3.85
2872.15	401.5	1.4	2686.5	4.05	

表 3 已知样本离散后的决策表

<i>U</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>U</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
1	4	5	6	1	10	1	1	7	2
2	2	2	8	1	11	3	4	5	2
3	2	2	9	1	12	3	3	3	2
4	2	5	5	1	13	3	7	7	2
5	1	5	3	1	14	7	8	1	3
6	4	1	2	1	15	5	6	3	3
7	6	2	6	1	16	7	8	1	3
8	6	4	6	1	17	7	8	1	3
9	3	3	4	2					

表 4 简化后的决策表

<i>U</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>U</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
1	4	5	6	1	9	3	3	4	2
2	2	2	8	1	10	1	1	7	2
3	2	2	9	1	11	3	4	5	2
4	2	5	5	1	12	3	3	3	2
5	1	5	3	1	13	3	7	7	2
6	4	1	2	1	14(16,17)	7	8	1	3
7	6	2	6	1	15	5	6	3	3
8	6	4	6	1					

表,如表 4 所示.

下面利用第 3 节提出的新的属性约简方法进行属性约简.

计算条件属性 *a*, 断点两侧属性值之差占断点的比例分别为

$$1.5464, 0.4145, 0.0169, 0.0681, \\ 0.0050, 0.0894, 0.0235.$$

最小值为 0.0050.

计算条件属性 *b*, 断点两侧属性值之差占断点的比例分别为

$$0.4490, 0.0153, 0.1127, 0.1320, \\ 0.7012, 0.2456, 0.0896, 0.1388.$$

最小值为 0.0153.

计算条件属性 *c*, 断点两侧属性值之差占断点的比例分别为

$$0.1818, 0.2857, 0.3158, 0.4286, \\ 0.2105, 0.1538, 0.0779, 0.0247.$$

最小值为 0.0247.

取条件属性 *c* 为属性约简结果, 它不能表示决策的不可分辨关系. 取条件属性 *b* 和 *c* 为属性约简的结果, 它能表示决策的不可分辨关系, 且不能再进行约简. 因此, 条件属性 *b* 和 *c* 组成的集合为所需的最好属性约简结果.

按属性分类求核值表, 并通过核值得出可能性决策表, 如表 5 所示.

表 5 可能性决策表

<i>U</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>U</i>	<i>b</i>	<i>c</i>	<i>d</i>
1	5	*	1	10	8	7	2
2	2	*	1	11	4	5	2
2'	*	8	1	12	3	*	2
3	*	9	1	13	7	*	2
6	*	2	1	14	8	*	3
7	*	6	1	14'	*	1	3
9	*	4	2	15	6	*	3

通过表 5 便可推出可能性决策

$$b_2 \quad b_5 \quad c_2 \quad c_6 \quad c_8 \quad c_9 \quad d_1, \\ b_3 \quad b_7 \quad b_4 \quad c_5 \quad c_4 \quad c_7 \quad d_2, \\ b_6 \quad b_8 \quad c_1 \quad d_3.$$

4.2 仿真实验

用表 1 提取的雷达辐射源实例, 构成特征参数为固定值的雷达模板库. 将表 1 中各样本特征矢量加上测量误差, 构成待识别目标特征矢量. 这里假定测量误差服从零均值的高斯分布. 选取两种不同的噪声环境进行仿真实验: 环境 1 的测量误差标准差为相应已知特征参数的 2%, 环境 2 的测量误差标准差为相应已知特征参数的 5%. 离散化过程采用第 2 节给出的离散化算法, 属性约简采用第 3 节给出的属性约简新方法.

在上述条件下经过 17000 次实验, 表 6 给出了基于粗糙集的辐射源识别算法得到的仿真结果, 并与统计模式识别方法进行比较.

表 6 基于 FS 的辐射源识别算法正确识别率 %

噪声	一般属性约简方法	断点指导属性约简方法	统计模式识别方法
环境 1	94.5 或 81.5 或 76.6	94.5	85.5
环境 2	83.7 或 67.6 或 59.6	83.7	71.3

由仿真结果可以看出, 就识别效果而言, 这种新的属性约简方法可直接得到最好的约简结果, 证明该方法有效而可行的.

5 结 论

本文研究粗糙集的属性约简问题及其在雷达辐

射源识别中的应用. 现有的约简方法都是对离散化后的决策表进行属性约简, 一般会得到多种约简结果, 然后根据问题的实际要求选取一个满意的结果. 本文通过监督型局部离散化过程中产生的断点, 提出一种指导属性约简的新方法, 它可直接得到合理的约简结果. 对所提出的模型给出了雷达辐射源用途识别的实例计算和仿真对比实验. 实验结果表明, 该方法是有效而可行的, 可应用于利用粗糙集理论进行目标识别的实际问题.

参考文献(References)

- [1] Pawlak Z. Rough sets [J]. *Int J of Information and Computer Science*, 1982, (11): 341-356.
- [2] Pawlak Z. Rough set theory and its application to data analysis [J]. *Int J of Cybernetics and Systems*, 1998, 29: 661-688.
- [3] 孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式 [J]. *电子学报*, 2000, 28(8): 117-119.
(Sun Q, Ye X Q, Gu W K. A new combination rules of evidence theory [J]. *Acta Electronica Sinica*, 2000, 28(8): 117-119.)
- [4] 王国胤. 粗糙集理论与知识获取 [M]. 西安: 西安交通大学出版社, 2001.
(Wang G Y. Rough set theory and knowledge acquisition [M]. Xi'an: Xi'an Jiaotong University Press, 2001.)
- [5] 常犁云, 王国胤, 吴渝. 一种基于 Rough Set 理论的属性约简及规则提取方法 [J]. *软件学报*, 1999, 10(11): 1206-1211.
(Chang L Y, Wang G Y, Wu Y. An approach for attribute reduction and rule generation based on rough set theory [J]. *J of Software*, 1999, 10(11): 1206-1211.)
- [6] 张文修, 米据生, 吴伟志. 不协调目标信息系统的知识约简 [J]. *计算机学报*, 2003, 26(1): 12-18.
(Zhang W X, Mi J S, Wu W Z. Knowledge reduction in inconsistent information systems [J]. *J of Computers*, 2003, 26(1): 12-18.)
- [7] 刘丽艳, 王海涌, 郑丽英. 基于粗集理论的决策规则约简算法的研究与应用 [J]. *兰州交通大学学报*, 2004, 23(6): 78-80.
(Liu L Y, Wang H Y, Zheng L Y. Research and application of algorithm for decision rule reduction based on rough set theory [J]. *J of Lanzhou Jiaotong University*, 2004, 23(6): 78-80.)
- [8] Yan Li, Simon C K Shiu, Sankar K Pal. Combining feature reduction and case selection in building CBR classifiers [J]. *IEEE Trans on Knowledge and Data Engineering*, 2006, 18(3): 415-429.
- [9] 李凡, 刘启和, 叶茂, 等. 不一致决策表的知识约简方法研究 [J]. *控制与决策*, 2006, 21(8): 857-862.
(Li F, Liu Q H, Ye M, et al. Approaches to knowledge reductions in inconsistent decision tables [J]. *Control and Decision*, 2006, 21(8): 857-862.)
- [10] Han J C, Hu X H, Lin T Y. A new computation model for rough set theory based on database systems [C]. *DAWAK '2003*. Berlin: Springer-Verlag, 2003: 381-390.
- [11] 关欣, 何友. 智能化雷达对抗情报处理技术研究 [J]. *海军航空工程学院学报*, 2005, 20(1): 101-106.
(Guan X, He Y. Research on technology of intellectualized radar countermeasures intelligence processing [J]. *J of Naval Aeronautical Engineering Institute*, 2005, 20(1): 101-106.)
- [12] Gao D X, Sun Z Q, Du T R. Dynamic surface control for hypersonic aircraft using fuzzy logic system [C]. *Proc of the IEEE Int Conf on Automation and Logistic*. Ji'nan, 2007: 2314-2319.
- [13] Kokotovic P V. The joy of feedback: Nonlinear and adaptive [J]. *IEEE Control System Magazine*, 1992, 12(7): 7-17.
- [14] Shin D H, Kim Y D. Reconfigurable flight control system design using adaptive neural networks [J]. *IEEE Trans on Control Systems Technology*, 2004, 12(1): 87-100.
- [15] Kim S H, Kim Y S, Song C. A robust adaptive nonlinear control approach to missile autopilot design [J]. *Control Engineering Practice*, 2004, 12(2): 149-154.
- [16] Jagannathan S, Lewis F L. Robust back-stepping control of a class of nonlinear systems using fuzzy logic [J]. *Information Sciences*, 2000, 123(3/4): 223-240.
- [17] Zhang T, Ge S S, Hang C C. Adaptive neural network control for strict-feedback nonlinear systems using backstepping design [J]. *Automatica*, 2000, 36(12): 1835-1846.
- [18] Mareels I M, Penfold H B, Evans R J. Controlling nonlinear time-varying systems via Euler approximations [J]. *Automatica*, 1992, 28(4): 681-696.

(上接第 463 页)