

中图分类号: TN911.7 TP391.4 TG156 文献标志码: A 文章编号: 1006-8961(2011)07-1213-06

论文索引信息: 汤勃, 孔建益, 王兴东, 侯宇, 陈黎. 粗糙集理论的带钢表面缺陷图像的识别与分类 [J]. 中国图象图形学报, 2011, 16(7): 1213-1218

粗糙集理论的带钢表面缺陷图像的识别与分类

汤勃¹⁾, 孔建益¹⁾, 王兴东¹⁾, 侯宇¹⁾, 陈黎²⁾

¹⁾(武汉科技大学机械自动化学院, 武汉 430081) ²⁾(武汉科技大学计算机科学与技术学院, 武汉 430081)

摘要: 针对带钢表面的划伤、黑斑、翘皮、辊印、褶皱和压印6种典型缺陷, 提取样本图像的灰度、纹理和几何形状特征等20维特征向量; 给出粗糙集理论的关键技术, 基于粗糙集理论构造带钢表面缺陷图像识别的决策表, 对决策表进行属性约简, 并直接从训练样本图像中导出决策规则; 应用所获取的规则对带钢表面缺陷测试样本图像进行分类, 并同BP算法进行对比, 验证了基于粗糙集理论的分类识别算法的有效性。

关键词: 粗糙集理论; 带钢表面缺陷; 识别; 分类

Recognition and classification for steel strip surface defect images based on rough set theory

Tang Bo¹⁾, Kong Jianyi¹⁾, Wang Xingdong¹⁾, Hou Yu¹⁾, Chen Li²⁾

¹⁾(School of Mechanical and Automation Engineering, Wuhan University of Science and Technology, Wuhan 430081 China)

²⁾(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081 China)

Abstract: The 20-dimensional feature vectors of intensity, texture and geometry characteristics for six kinds of steel strip surface typical defects images are extracted. The key technology of Rough Set theory is described. The decision table of the steel strip surface images recognition is created, the reduction for decision table is carried out, and the decision rules are obtained from the training sample images directly. The test samples of the steel surface defect images have been classified with application of decision rules, and then compare with the BP neural network algorithm. The recognition and classification of steel strip surface typical defects images based on rough set theory is effective.

Keywords: rough set theory; steel strip surface defects; recognition; classification

0 引言

带钢在制造过程中, 难免会产生划伤、黑斑、翘皮、辊印、褶皱和压印等表面缺陷。这些缺陷影响产品的外观和性能, 所以对其进行检测、分类以便及时发现并加以控制是必要的。目前, 基于机器视觉的带钢表面质量检测方法是研究的热点^[1-2]。基于机

器视觉的带钢表面质量检测, 对 CCD 摄像机采集带钢图像, 通过图像处理和分析, 提取缺陷图像特征, 进行缺陷的自动分类。图 1 是带钢表面的 6 种典型缺陷图像, 依次为划伤、黑斑、褶皱、辊印、翘皮和压印。

粗糙集 (RS) 理论是波兰科学家 Pawlak 提出的, 是一种新型的处理模糊和不确定知识的数学工具, 其揭示潜在的规律, 能有效地分析和处理不精

收稿日期: 2010-05-13; 修回日期: 2010-06-30

基金项目: 高等学校博士学科点专项科研项目 (20104219110001); 武汉市科技攻关项目 (200910321100); 武汉科技大学青年科技骨干培育计划项目 (2009xz24)。

第一作者简介: 汤勃 (1973—), 男, 机械设计及理论专业博士研究生, 主要研究方向为智能设计与控制。E-mail: tang1017@163.com。

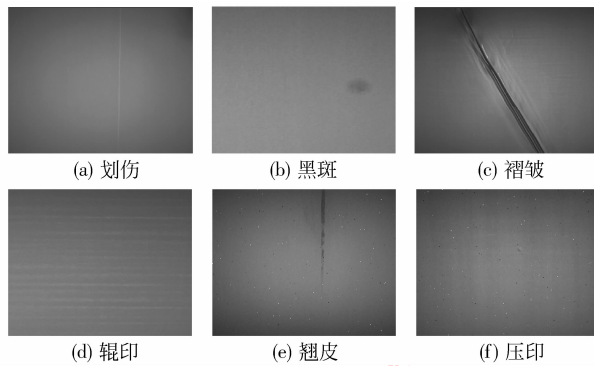


图 1 几种带钢常见缺陷

Fig. 1 Several common defects of steel strip

确、不一致和不完整等各种不完备信息,并能从中发现隐含的知识和揭示潜在的规律^[3-4]。它提供了一套比较完备的从小样本数据中寻找规律的系统方法,用其可找到描述正常模型的最小预测规则集,这不仅有利于提高检测速度,而且可应用于系统的实时检测。

目前,粗糙集理论已经在数据的决策和分析、模式识别、机器学习等方面有着成功的应用,已成为信息科学最活跃的研究领域之一。本文基于 RS 理论对带钢表面缺陷图像进行识别。

1 粗糙集理论基本概念^[5]

定义 1 设 U 和 R 分别为论域和等价关系集合, U/R 表示 R 将 U 分成的所有等价类构成的等价类簇,亦称为知识,知识库可表示为 $K = (U, R)$ 。

定义 2 一个知识表达系统 S 定义为

$$S = \langle U, C, D, V, f \rangle \quad (1)$$

式中: U 是对象的非空有限集合; $C \cup D = R$ 是属性集合,子集 C 和 D 分别称为条件属性和决策属性; $V = \bigcup_{a \in A} V_a$ 是属性值的集合, V_a 表示了属性 $a \in A$ 的取值范围; $f: U \times A \rightarrow V$ 是一个信息函数,它指定 U 中的每一对象 x 的属性值。

定义 3 给定论域 U , 等价关系 R 将 U 划分为互不相交的基本等价类 U/R 。设 X 是论域 U 上的一个集合,若 X 能表示成等价类组成的并集时,则称 X 在 U 上是 R 可定义的;否则 X 为 R 是不可定义的。 R 可定义集即为 R 精确集; R 不可定义集则称为 R 非精确集或是 RS。

定义 4 设 $X, Y \in U, R$ 是定义在 U 上的等价

关系,则集合 X 关于 R 的下近似集和上近似集分别定义为

$$\underline{R}(X) = \bigcup \{ Y \in U/R : Y \subseteq X \} \quad (2)$$

$$\overline{R}(X) = \bigcup \{ Y \in U/R : Y \cap X \neq \emptyset \} \quad (3)$$

定义 5 $\underline{R}(X)$ 是根据现有知识判断确定 X 的对象组成的最大集合称为正域,即 $POS(X) = \underline{R}(X)$; X 的 R 边界定义为 $BN_R(X) = \overline{R}(X) - \underline{R}(X)$ 。从形式上看,上近似就是正域和边界的并集。

定义 6 令 R 是一簇等价关系,给定知识库 $K = (U, R), r \in R$, 如果

$$ind(R) = ind(R - \{r\}) \quad (4)$$

则称 r 为 R 中冗余的,否则 r 为 R 中非冗余的。即论域中对象的属性集合 R , 去除其冗余属性 r 后,余下的属性集 $R - r$ 仍旧保持原有的等价关系。

定义 7 设 $Q \subseteq P$, 如果 Q 是独立的,且 $ind(Q) = ind(P)$, 则称 Q 为 P 的一个约简 (reduct), 记为 $red(P)$ 。一簇等价关系 P 可能为多个约简,全部约简的交集定义为 P 的核 (core), 即

$$core(P) = \bigcap red(P) \quad (5)$$

定义 8 决策表中决策属性 D 依赖条件属性 C 的依赖度定义为

$$\gamma(C, D) = \frac{|POS(C, D)|}{|U|} \quad (6)$$

式中, $|POS(C, D)|$ 表示正域元素的个数, $|U|$ 表示整个对象集合的个数。

定义 9 设 C 和 D 分别为条件属性集和决策属性集, $C, D \in A, a \in C$, 属性 a 关于 D 的重要度定义为

$$SGF(a, C, D) = \gamma(C, D) - \gamma(C - \{a\}, D) \quad (7)$$

式中: $\gamma(C - \{a\}, D)$ 表示缺少属性 a 后,条件属性与决策属性的依赖程度; $SGF(a, C, D)$ 表示 C 中缺少属性 a 后,导致不能被准确分类的对象在系统中所占的比例。

2 带钢表面缺陷图像的识别与分类

2.1 算法流程

基于粗糙集理论进行带钢表面缺陷图像的规则获取和分类识别算法流程如图 2 所示。

2.2 基于粗糙集理论建立决策表

由定义 2, 知识表达系统可表示为 $S = \langle U, C, D, V, f \rangle$, 由这样的“属性-值”对就构成了一

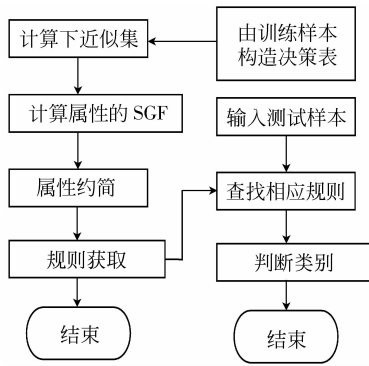


图 2 规则获取和分类识别算法流程

Fig. 2 The algorithm flow of rules access and classification

张 2 维表,称之为信息表。具有条件属性和决策属性的知识表达系统称为决策系统。

带钢表面缺陷的分类其实也是一个模式分类问题。根据粗糙集理论,将表面缺陷分类描述为一个系统决策问题,所以可以根据其条件属性和决策属性构造一个决策表。

提取带钢表面划伤、黑斑、翘皮、辊印、褶皱和压印 6 类典型缺陷(决策属性)的灰度特征、纹理特征和几何形状特征共 20 维特征值(条件属性)。具体地,决策表的条件属性设为

$A = \{A1 = \text{缺陷区面积}, A2 = \text{分散度}, A3 = \text{长宽比}, A4 = \text{欧拉数}, A5 = \text{灰度最大值}, A6 = \text{灰度最小值}, A7 = \text{灰度平均值}, A8 = \text{灰度的值域}, A9 = \text{灰度的方差}, A10 = \text{灰度的熵}, A11 = \text{背景灰度的值域}, A12 = \text{背景纹理的能量}, A13 = \text{缺陷区纹理的对比度}, A14 = \text{缺陷区纹理的能量}, A15 = \text{缺陷区纹理的同质性}, A16 = \text{一阶不变矩}, A17 = \text{二阶不变矩}, A18 = \text{三阶不变矩}, A19 = \text{五阶不变矩}, A20 = \text{六阶不变矩}\}$ 。

决策属性设为 $\text{Dec}[D] = \{1, 2, 3, 4, 5, 6\}$ 分别表示划伤、黑斑、翘皮、辊印、褶皱和压印 6 类典型缺陷。

此处采用 6 类典型缺陷共 164 个样本作为训练样本。将其属性值形成一张的 2 维表格,每一行描述一个对象,每一列描述对象的一种属性。于是形成了一个 21×164 的决策表,列向量的前 20 个属性表征带钢表面缺陷的 20 维特征值,第 21 个属性是决策属性,表征缺陷的类别。164 维向量表示 164 个训练样本。由于篇幅所限,表 1 只给出了带钢表面典型缺陷图像训练样本决策表的一部分。

表 1 缺陷图像训练样本决策表

Tab. 1 Decision table of raining samples for defect images

样本	条件属性								决策属性
	A1	A2	A3	...	A18	A19	A20		
1	485	492	0.06	...	22	50	26		1
2	468	471	0.06	...	18	47	23		1
				⋮					
29	35	39	0.57	...	28	59	36		2
30	43	48	0.56	...	26	54	33		2
				⋮					
61	943	834	0.13	...	22	49	27		3
62	999	891	0.11	...	22	48	27		3
				⋮					
94	491	272	0.17	...	24	48	29		4
95	532	275	0.09	...	24	49	29		4
				⋮					
114	4 001	1 486	0.50	...	23	47	30		5
115	1 948	76	0.50	...	23	46	30		5
				⋮					
163	61	20	0.33	...	26	52	32		6
164	45	18	0.29	...	27	55	34		6

2.3 决策表连续属性的离散化

当决策表 $S = \langle U, C, D, V, f \rangle$ 的属性值繁多时,通过将对象的属性值向量和决策表的每行进行匹配来发现新对象的方法是不可取的,因此对带有实值属性的决策表,应该采用一些离散化策略,以在尽量减少决策表信息损失的前提下,得到简化和浓缩的决策表,以使用粗糙集理论分析,获得决策所需要的知识。

给定决策表 $S = \langle U, C, D, V, f \rangle$,实例 $x \in U$ 在属性 $a \in C \cup D$ 上的取值记为 $a(x)$,若 $\exists a \in A \wedge \exists x \in U(a(x)) \in \mathbf{R}$,则 S 是连续的;否则 S 是离散的。对 $\exists a \in A, \mathbf{R}$ 是实数集合,且 $V_a = [l_a, r_a] \subseteq \mathbf{R}$ 是属性 a 的值域,将 V_a 划分为若干区间,即 $P_a = \{[V_0^a, V_1^a], [V_1^a, V_2^a), \dots, [V_k^a, V_{k+1}^a)\}$ 是 V_a 的划分。对于 $[V_i^a, V_{i+1}^a]$,若 $V_i^a \leq C_i^a \leq V_{i+1}^a$,则 C_i^a 可作为区间 i 的断点来有效分辨 V_i^a 和 V_{i+1}^a 。于是, S 被断点的集合 $\{C_a \mid a \in A\}$ 离散化。因此, $P = \bigcup_{a \in \mathbf{R}} P_a$ 定义了一个新的决策表:

$S^P = \langle U, C, D, V^P, f^P(x_a) \rangle \Leftrightarrow f(x_a) \in [V_i^a, V_{i+1}^a]$ 即经过离散后,原来的信息系统被一个新的信息系统所代替。

连续属性离散化的算法较多,例如布尔逻辑算法^[6]、贪心算法^[7]、遗传算法^[8]、模糊聚类算法^[9]、

基于决策表属性重要性算法^[10]和基于熵的离散化算法^[11]。

基于熵的离散化可以减少数据量,同时被离散化后的信息丢失少,由于使用的是类信息,所以更有可能将区间边界定义在准确位置,有利于提高分类的准确性。

设 $X \subseteq U$, 决策属性为 j 的概率 p_j 时定义其信息熵

$$H(X) = - \sum_j p_j \log_2 p_j \quad (8)$$

将连续属性的取值范围先划分为若干区间,再选择两个毗邻区间合并,使合并前后的熵差最小;然后重置划分点,保存合并后的熵。重复这一合并步骤,当达到熵损失和区间数之间的最佳平衡状态时也就得到了合理的划分点数和位置。

基于熵的离散化方法将表 1 所示的决策表离散化,例如属性 A_1 离散区间有 7 个,即 $(-\infty, 36)$, $[36, 75.5)$, $[75.5, 95)$, $[95, 449)$, $[449, 553)$, $[553, 1945)$, $[1945, +\infty)$ 每个区间相应的离散值为 0,1,2,3,4,5,6。测试样本决策表的离散化后的结果见表 2。

表 2 训练样本决策表的离散化

Tab.2 Discrete decision table for training samples

样本	条件属性							决策属性
	A1	A2	A3	...	A18	A19	A20	
1	4	2	0	...	1	2	1	1
2	4	2	0	...	1	1	0	1
				⋮				
29	0	0	1	...	2	2	2	2
30	1	0	1	...	2	2	2	2
				⋮				
61	5	3	0	...	1	1	1	3
62	5	3	0	...	1	1	1	3
				⋮				
94	4	2	0	...	2	1	2	4
95	4	2	0	...	2	1	2	4
				⋮				
114	6	3	1	...	2	1	2	5
115	6	3	1	...	2	1	2	5
				⋮				
163	1	0	0	...	2	2	2	6
164	1	0	0	...	2	2	2	6

2.4 决策表的约简和规则的生成

决策表离散化后就进行约简。从表 2 就可以看出,在决策表中,各个条件属性之间往往存在着某些程度上的依赖或关联,并非所有条件属性都是必要

的或同等重要的,甚至有的属性是冗余的。约简的目的就是要在不丢失信息的前提下,去除其中不相关或不重要的属性,这样并不影响原有的表达效果。

属性约简算法的步骤为:

- 1) 计算条件属性和决策属性的等价集;
- 2) 计算决策属性各等价集的下近似集;
- 3) 计算各属性的重要度 $SGF(a, C, D)$, 如果不为 0, 则属性 a 不可约简, 否则属性 a 可约简;
- 4) 对约简后的决策表进行一致性检查, 如果决策表有一致性的属性, 则该属性可约简, 否则该属性不可约简;
- 5) 删除决策表中的可约简属性, 最终得到简化后的决策表。

据此, 针对 164 个训练样本得到的条件属性集约简取 100 条, 节录如下:

- Reduct #1 = {A7, A8, A14, A17, A19};
- Reduct #2 = {A6, A11, A14, A17, A19};
- Reduct #3 = {A3, A7, A11, A14, A17, A19, A20};
-
- Reduct #51 = {A6, A10, A12, A14, A16, A17, A20};
- Reduct #52 = {A6, A7, A10, A12, A14, A16, A20};
-
- Reduct #98 = {A1, A6, A10, A14, A17};
- Reduct #99 = {A1, A7, A8, A14, A17};
- Reduct #100 = {A2, A7, A8, A14, A17}。

每条约简大括号内的数字表示该约简所包含的条件属性有哪些。以 Reduct #1 = {A7, A8, A14, A17, A19} 为例, A7, A8, A14, A17, A19 分别表示该约简只需用带钢表面缺陷的 20 维特征值中的第 7、8、14、17、19 共 5 个特征值。对本约简而言, 其他 15 个特征值是冗余的。

规则提取其实就是决策规则的约简结果, 主要对条件属性值的约简。在决策表中, 每一行对应一条决策规则, 首先计算决策规则的条件属性的核值, 其次求它的条件属性值的约简。如果约简表中出现重复行, 则将其删除, 因为其表示同一决策规则。本文最终获取的规则有 11 条:

- 规则 1 (A5 = 2) & (A11 = 2) ⇒ (Dec = 1);
- 规则 2 (A1 = 3) & (A7 = 2) ⇒ (Dec = 1);
- 规则 3 (A2 = 0) & (A11 = 0) ⇒ (Dec = 2);

- 规则 4 $(A3 = 0) \& (A7 = 1) \& (A9 = 1) \& (A14 = 0) \& (A18 = 2) \Rightarrow (Dec = 2)$;
- 规则 5 $(A6 = 2) \& (A17 = 2) \Rightarrow (Dec = 2)$;
- 规则 6 $(A4 = 4) \& (A17 = 1) \Rightarrow (Dec = 3)$;
- 规则 7 $(A7 = 1) \& (A8 = 3) \Rightarrow (Dec = 3)$;
- 规则 8 $(A3 = 0) \& (A9 = 0) \& (A14 = 1) \Rightarrow (Dec = 4)$;
- 规则 9 $(A2 = 3) \& (A17 = 0) \Rightarrow (Dec = 4)$;
- 规则 10 $(A10 = 2) \& (A17 = 2) \Rightarrow (Dec = 5)$;
- 规则 11 $(A6 \text{ in } \{5, 4\}) \& (A11 = 2) \& (A18 = 2) \Rightarrow (Dec = 6)$ 。

2.5 实例验证

在 150 个测试样本中随机抽取 2#、29#和 72# 3 个样本,验证的情况如下:

1)2#样本数据为

168 165 0.02 2 152 107 135 45 10 5 101 0.54 0.32 0.50 0.84 4 8.21 21 52 25;

离散后数据为

$\{(A1 = 3), (A2 = 1), (A3 = 0), (A4 = 1), (A5 = 2), (A6 = 6), (A7 = 2), (A8 = 1), (A9 = 1), (A10 = 1), (A11 = 3), (A12 = 1), (A13 = 0), (A14 = 1), (A15 = 0), (A16 = 0), (A17 = 0), (A18 = 1), (A19 = 2), (A20 = 1), (Dec = 1)\}$;

适用规则为规则 2;判断的正确性为“正确”。

2)29#样本数据为

23 25 0.56 6 66 45 53 21 6 4 42 0.49 0.2 0.37

0.9 5.6 12 26 52 32

离散后数据为

$\{(A1 = 0), (A2 = 0), (A3 = 1), (A4 = 2), (A5 = 0), (A6 = 1), (A7 = 0), (A8 = 0), (A9 = 0), (A10 = 0), (A11 = 0), (A12 = 0), (A13 = 0), (A14 = 0), (A15 = 0), (A16 = 2), (A17 = 2), (A18 = 2), (A19 = 2), (A20 = 2), (Dec = 2)\}$;

适用规则为规则 3;判断的正确性为“正确”。

3)72#样本数据为

957 915 0.1 168 135 58 96 77 11 5 93 0.52 0.03 0.54 0.98 4 9 20 49 25

离散后数据为

$\{(A1 = 5), (A2 = 3), (A3 = 0), (A4 = 4), (A5 = 1), (A6 = 2), (A7 = 1), (A8 = 3), (A9 = 1), (A10 = 1), (A11 = 2), (A12 = 1), (A13 = 0), (A14 = 1), (A15 = 0), (A16 = 1), (A17 = 1), (A18 = 1), (A19 = 1), (A20 = 1), (Dec = 3)\}$;

适用规则为规则 6、规则 7;判断的正确性为“正确”。

可见用上述所获取的规则对这 3 个样本的分类结果是正确的。

2.6 分类识别结果

利用上述训练样本表生成的规则对带钢表面 6 类缺陷图像的 150 个样本进行测试,分类结果如表 3 所示。

表 3 RS 算法与 BP 算法识别结果比较

Tab.3 The recognition results compared RS algorithm with BP algorithm

缺陷类型	测试样本数	识别正确数		识别正确率/%	
		RS 算法	BP 算法	RS 算法	BP 算法
划伤	24	22	22	91.67	91.67
黑斑	32	32	31	100	96.88
翘皮	30	30	23	100	76.67
辊印	14	13	14	92.86	100
褶皱	20	20	20	100	100
压印	30	29	29	96.67	96.67

从上表可以看出,利用粗糙集理论对带钢表面缺陷图像进行分类识别的正确率都在 90% 以上,有 3 种缺陷图像的正确识别率达到了 100%,而且粗糙集理论算法的正确识别率整体上高于 BP 神经网络算法。

3 结 论

1)给出粗糙集理论的关键技术,基于粗糙集理论研究了带钢表面缺陷图像识别的规则获取和分类

流程;

2) 获取了带钢表面 6 种典型缺陷图像的 314 个样本(其中 164 个样本作为训练样本,另外 150 个样本作为测试样本),提取了带钢表面缺陷图像的灰度、纹理和几何形状特征等 20 维特征向量;

3) 基于粗糙集理论对带钢表面缺陷图像进行分类,正确识别率很高,而且整体上优于 BP 神经网络算法。同时亦表明所提取的带钢表面缺陷图像的特征向量是有效的。

参考文献 (References)

- [1] Xu Ke, Xu Jinwu, Liu Shouli, et al. Research on automatic surface inspection system for cold rolled strips [J]. Journal of Iron and Steel, 2000, 35(10): 63-64. [徐科, 徐金梧, 鹿守理, 等. 冷轧带钢表面自动监测系统的研究 [J]. 钢铁, 2000, 35(10): 63-64.]
- [2] He Yonghui, Wang Kangjian, Shi Guifen. High-speed hole inspection system for steel strips based on machine vision [J]. Journal of Applied Optics, 2007, 28(3): 345-347. [何永辉, 王康健, 石桂芬. 基于机器视觉的高速带钢孔洞检测系统 [J]. 应用光学, 2007, 28(3): 345-347.]
- [3] Pawlak Z. Rough Set Theoretical Aspects of Reasoning about Date [M]. Warsaw, Poland: Kluwer Academic, 1991.
- [4] Wang Guoyin. Rough Set Theory and Knowledge Acquisition [M]. Xi ' an: Xi ' an Jiaotong University Press, 2001: 23-27. [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: 23-27.]
- [5] Qiu Yuxia. The Research and Application for Evolutionary Computation and Rough Set [M]. Beijing: Metallurgical Industry Press, 2009: 105-126. [邱玉霞. 进化计算与粗糙集研究及应用 [M]. 北京: 冶金工业出版社, 2009: 105-126.]
- [6] Nguyen H S, Skowron A. Quantization of real values attributes, rough set and boolean reasoning approaches [C] // Proceedings of the 2nd Joint Annual Conference on Information Science. Wrightsville Beach. North Carolina: Duke University, 1995: 34-37.
- [7] Nguyen S H, Nguyen H S. Some efficient algorithms for rough set methods [C] // Proceedings of Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems. Granada, Spain: University of Granada, 1996: 1451-1456.
- [8] Richeldi M, Rossotto M. Class-driven statistical discretization of continuous attributes [C] // Proceedings of European Conference on Machine Learning, Lecture Notes in Artificial Intelligence. Berlin: Springer Verlag, 1995: 335-338.
- [9] Yu Daren, Hu Qinghua, Bao Wen. Combing rough set methodology and fuzzy clustering for knowledge discovery from quantitative data [J]. Journal of Proceedings of the CSEE, 2004, 24(6): 205-210. [于达仁, 胡清华, 鲍文. 融合粗糙集和模糊聚类的连续数据知识发现 [J]. 中国电机工程学报, 2004, 24(6): 205-210.]
- [10] Zhang Jianjun, Zhang Jingbo. A novel rough set theory-based decision table discretization algorithm [J]. Journal of Xidian University, 2004, 13(3): 469-472. [张建军, 张静波. 一种新的基于粗糙集理论的决策表离散化算法 [J]. 西安电子科技大学学报: 自然科学版, 2004, 13(3): 469-472.]
- [11] Chmielewski M R, Grzymala-Busse J W. Global discretization of attributes as preprocessing for machine learning [C] // Proceedings of the Third International Workshop on Rough Sets and Soft Computing. San Jose, California: San Jose State University, 1994: 294-301.