

粗糙集理论在阳离子反浮选控制中的应用

张 勇^{1,2}, 王 峰², 潘学军¹, 王 伟¹

(1. 大连理工大学 信息与控制研究中心, 辽宁 大连, 116023;
2. 鞍山科技大学 电子信息与工程学院, 辽宁 鞍山, 114002)

摘要: 针对阳离子反浮选生产过程被控对象复杂、数学模型不确定以及控制精度要求高等特点, 提出了一种基于粗糙集理论的智能控制方法. 通过对控制过程中的一些有代表性的状态以及操作人员在这些状态下采取的控制策略所构成的决策表进行约简, 归纳出一系列控制规则. 运用该方法进行准确度测试和命中率试验, 结果表明控制模型的给出值与现场实际数值基本吻合, 命中指标满足综合指标要求. 在阳离子反浮选控制中运用粗糙集理论时, 便于现场技术人员和操作工人的操作, 能合理添加药剂量, 避免了传统工艺由人为加药造成的浪费.

关键词: 浮选过程控制; 粗糙控制; 粗糙集; 信息系统

中图分类号: TP242.3 **文献标识码:** A **文章编号:** 1005-9792(2003)04-0368-05

粗糙控制是在 Z. Pawlak 提出的粗糙集理论的基础上发展起来的^[1], 是一种根据观测数据获得控制策略的方法. 通过对控制过程中的一些有代表性的状态以及操作人员在这些状态下所采取的控制策略所构成的决策表进行约简, 从而归纳出一系列控制规则. 该理论在数据的决策与分析、模式识别、机器学习与知识发现等方面的广泛应用, 才引起了世界各国学者的关注. 粗糙集理论能有效地处理下列问题: 不确定或不精确知识的表达; 经验学习并从经验中获取知识; 不一致信息的分析; 根据不确定、不完整的知识进行推理; 在保留信息的前提下进行数据化简; 近似模式分类; 识别并评估数据之间的依赖关系^[1-4]. 目前, 粗糙集理论已成为信息科学中最为活跃的研究方向之一. 同时, 该理论还在医学、化学、材料学、地理学、管理科学和金融等其他学科得到了成功的应用.

浮选法是依据各种矿物表面性质的差异, 从矿浆中借助气泡的浮力选分矿物的方法, 其中对浮选药剂的控制是重要的环节. 而浮选过程具有时变、强耦合、非线性和不确定性等特征, 是一个物理化学综合反应过程, 受到矿浆体积流量、矿浆浓度、目标输出品位、设备运行状况、浮选药剂成分状况以及浮选过程中的人工操作等诸因素直接或间接的影响, 很

难建立精确的数学模型, 使得控制变得复杂^[5]. 为此, 作者运用粗糙集理论, 提出一种用于阳离子反浮选过程的智能控制方法.

1 粗糙集基本理论

1.1 粗糙集理论基本概念

1.1.1 粗糙集

定义 1 令 $X \subseteq U$, 且 R 为 U 上的一个等价关系, 当 X 能表达成某些 R 基本范畴的并时, 称 X 是 R 可定义的; 否则称 X 是 R 不可定义的. R 的可定义集也称作 R 的精确集, 而 R 的不可定义集也称作 R 的非精确集或 R 的粗糙集^[1,6].

1.1.2 信息系统

一个信息系统 S , 可以表示为 $S = \{U, R, V\}$, 其中 U 是对象的有限集 $U = \{x_1, x_2, \dots, x_n\}$, R 中的属性可进一步划分为互不相交的 2 个属性集 C 和 D , 分别称为条件属性集和决策属性集, $R = C \cup D$ 是属性有限集, $V = \bigcup_{p \in R} V_p$ 是属性值集合, V_p 是属性 p 的值域.

令 $P \subset R$, $x_i \in U$, $x_j \in U$, 定义一个不可分辨二元关系 $IND(B)$, 即

收稿日期: 2003-04-15

基金项目: 教育部科研技术重点研究项目(01053)资助

作者简介: 张 勇(1963-), 男, 辽宁鞍山人, 鞍山科技大学副教授, 大连理工大学博士研究生, 从事工业生产过程控制、智能控制等研究.

$$\text{IND}(B) = \{(x_i, x_j) \in U \times U: \\ \forall p \in P, p(x_i) = p(x_j)\}.$$

对于 $\forall p \in P$, 当且仅当 $p(x_i) = p(x_j)$ 时, 称 x_i 与 x_j 依据 P 在 S 中的属性集是不可分辨的. $\text{IND}(B)$ 称为对于 $\forall P \subset R$ 关于 U 的等价关系. $\text{IND}(B)$ 的等价类称为 P 在 S 中的基本集.

对于一个信息系统 $S = \{U, R, V\}$, 且 $B \subset R$ 是一个等价关系(不可分辨的关系), 有序对 $RS = (U, B)$ 称为近似空间. 对于 U 中任何元素 x_i 在关系 B 中的等价类记为 $[x_i]_B$. 由于 B 的等价类描述该对象的最小不可分辨组, 故它被称为 R 的基本集. RS 中任何基本集的有限并称为 RS 中的可定义集合.

令 $X \subset U$, RS 中关于 X 的下近似定义为 $B(X) = \{x_i \in U | [x_i]_B \subset X\}$. 上近似定义为 $\bar{B}(X) = \{x_i \in U | [x_i]_B \cap X \neq \emptyset\}$. $B(X)$ 是所有包含于 X 的基本集的并. 对于 $\forall x_i \in B(X)$, 必然有 $x_i \in X$. $B(X)$ 是所有与 X 的交不为空的基本集的并. 对于 $\forall x_i \in B(X)$, 只能说 x_i 可能属于 X .

1.1.3 核与属性约简

核与约简是粗糙集的2个基本概念. 约简是信息系统的不可缺少的组成部分, 它可以识别原信息系统可识别的所有对象. 核是所有约简的共有部分.

对于信息系统 $S = \{U, R, V\}$, 其中 $R = C \cup D$ 且 $B \subset C$, 定义 $\text{IND}(D)$ 中的正域 B 为 $\text{POS}_B(D) = \cup\{B(X) : X \in \text{IND}(D)\}$. 正域 $\text{POS}_B(D)$ 包含 U 中所有对象, 仅仅基于 $\text{IND}(B)$ 的分类信息, 就能将这些对象划分为 $\text{IND}(D)$ 等价类.

定义2 对于属性 $p \in B$, 如果 $\text{POS}_B(D) = \text{POS}_{B-\{p\}}(D)$, 则称属性 p 关于 D 是冗余的. 否则, 称属性 p 关于 D 是必不可少的.

在一个信息系统中, 如果某一个属性是冗余的, 那么可以从信息系统中把它删除且不会改变原系统的依赖关系; 若一个属性是必不可少的, 由于它携带着信息系统中对象的重要信息, 如果不想改变原系统的依赖关系, 那么就保留该属性.

定义3 条件属性集 C 中关于决策属性 D 的所有必不可少的属性集合被称为 C 的核. $\text{CORE}(C, D) = \{a \in C : \text{POS}_C(D) = \text{POS}_{C-\{a\}}(D)\}$.

定义4 对于 $B \subset C$, 如果 B 关于 D 是独立的且 $\text{POS}_C(D) = \text{POS}_B(D)$, 则 B 称为 S 中的约简.

定义5 信息系统 S 中 P 的可辨识矩阵 $M(P) = \{m_{i,j}\}_{n \times n}$ 定义为

$$\{m_{i,j}\} = \{q \in P : p(x_i, q) \neq p(x_j, q)\}, \\ \text{for } i, j \text{ 均为 } 1, 2, \dots, n.$$

其中: $m_{i,j}$ 是可识别对象 x_i 和 x_j 的所有属性集. 由于 $M(P)$ 是对称的, 所以只需要计算当 $1 \leq j \leq i \leq n$ 时的 $m_{i,j}$.

若检测 $P' \subset P$ 是否为 P 的约简, 使用可辨识矩阵只需检测在 $M(P)$ 中对于任意非空项 m 是否有 $P' \cap m \neq \emptyset$, 这是很容易实现的. 这就意味着约简是属性的最小子集, 该子集能够识别整个属性集上所有可识别的对象.

1.2 粗糙集的约简方法

1.2.1 基于可辨识矩阵和逻辑运算的属性约简算法

步骤如下^[7].

- a. 计算可辨识矩阵 $M(P)$;
- b. 对于可辨识矩阵中的所有取值为非空集合的元素 $m_{i,j}$ ($m_{i,j} \neq \emptyset$, $m_{i,j} \neq \emptyset$), 建立相应的析取逻辑表达式 L_{ij} , $L_{ij} = \bigvee_{a_i \in m_{ij}} a_i$;
- c. 将所有的析取逻辑表达式 L_{ij} 进行合取运算, 得到一个合取范式 L , 即 $L_{ij} = \bigwedge_{m_{ij} = \emptyset, m_{ij} \neq \emptyset} L_{ij}$;
- d. 将合取范式 L 转换为析取范式的形式 L' , 即 $L' = \bigvee_i L_i$;

e. 输出属性约简结果. 析取范式中的每个合取项对应一个属性约简的结果, 每个合取项中所包含的属性组成约简后的条件属性集合. 其中 \wedge 和 \vee 是合取和析取运算符.

1.2.2 基于数据分析方法属性值约简算法

基于数据分析方法提出如下的属性值约简算法^[8]:

- a. $i = 1; i \in [1, m]$;
- b. $\forall a_j \in C, A_j = [x_i]_{a_j}, \wedge (a_j)_{x_i} \rightarrow d_{x_i}$;
- c. $M_i = \bigcap_j A_j, b_k \in M_i$;
- d. $N_i = [x_i]_D$;
- e. 若 $M_i \subseteq N_i$ 则转 f., 否则转 b.;
- f. $\wedge (b_k)_{x_i} \rightarrow d_{x_i}'; i = i + 1, i \leq m$ 转 g., 否则转 g.;
- g. 结束.

2 粗糙集理论在阳离子反浮选过程控制中的应用

粗糙集理论的基本思想是把数据集看作决策表, 对决策表进行属性约简和属性值约简, 删除冗余的属性和属性值, 最后由约简后的决策表得出决策规则. 其实质是用条件属性 C 划分的等价类表示用

决策属性 D 划分的等价类. 粗糙集理论的应用过程也就是先用训练数据得到判别规则, 然后用判别规则进行分类, 进而对控制对象进行控制. 在阳离子反浮选过程控制中应用粗糙控制时, 采取如下方案.

2.1 数据准备

在数据准备阶段, 主要是对反浮选过程中观测量的数据采集做一些准备工作. 该阶段是要对大量的历史记录进行搜集并整理, 并且力求数据的准确、可靠.

2.2 数据的预处理

数据的预处理主要包括对数据的理解、属性的选择、连续属性的离散化、数据中的噪声及丢失值处理等等. 数据的预处理的目的是提供可以进行数据挖掘和知识发现的数据.

2.2.1 属性的选择

经过对流程的分析, 选取如下属性: 给矿品位, 精矿品位, 给矿浓度, 给矿粒度, 给矿流量, 药剂流量. 其中药剂流量是可控对象, 故选为决策属性, 其他 5 个属性选取为条件属性.

2.2.2 连续属性离散化

对于获取的连续属性值, 应用领域知识进行离散化处理. 具体过程如下:

a. 给矿品位属性值表示范围(%): 1) < 60 ; 2) $[60, 62)$; 3) $[62, 64)$; 4) $[64, 66)$; 5) ≥ 66 .

b. 精矿品位属性值表示范围(%): 1) < 68 ; 2) $[68.5, 69)$; 3) $[62, 64)$; 4) $[69, 69.5)$; 5) ≥ 69.5 .

c. 给矿浓度属性值表示范围(%): 1) < 30 ; 2) $[30, 32)$; 3) $[32, 34)$; 4) $[34, 36)$; 5) ≥ 36 .

d. 给矿流量属性值表示范围($m^3 \cdot h^{-1}$): 1) < 290 ; 2) $[290, 310)$; 3) $[310, 330)$; 4) ≥ 330 .

e. 给矿粒度属性值表示范围(%): 1) < 85 ; 2) $[85, 90)$; 3) $[90, 95)$; 4) ≥ 95 .

f. 药剂流量属性值表示范围(L): 1) < 7 ; 2) $[7, 8.5)$; 3) $[8, 5, 10)$; 4) $[10, 11.5)$; 5) ≥ 11.5 .

2.2.3 属性字段的定义

为方便决策表建立及属性、属性值的约简, 定义属性字段如下: 给矿品位(a)、精矿品位(b)、给矿浓度(c)、给矿粒度(d)、给矿流量(e)、药剂流量(f).

2.2.4 噪声及丢失值处理

数据库中噪声和缺值是一种常见的现象, 对测量值中存在的噪声和缺值的处理采用如下步骤: 删除噪声数据, 以正常值代替缺失值.

2.2.5 建立初始决策表

根据定义的字段和经过噪声及缺值处理后的离散值建立初始决策表, 选取部分有代表性的数据建

表, 如表 1 所示.

表 1 初始决策表

论域(U)	条件属性(C)					决策属性(D)
	a	b	c	d	e	f
1	4	5	4	1	4	2
2	3	4	4	1	4	2
3	4	4	4	2	3	2
4	3	5	4	3	3	3
5	3	5	4	3	3	3
6	3	3	3	1	4	3
7	3	5	4	1	4	2
8	4	3	3	1	4	2
9	2	2	3	3	2	4
10	2	2	4	3	3	4
11	3	3	4	3	3	4
12	2	2	4	2	3	3
13	4	3	4	3	2	4
14	3	3	2	1	4	4

2.3 属性约简

运用基于辨识矩阵和逻辑运算的属性约简算法可以得出 $L' = a \wedge b \wedge c \wedge d$, 即 $IND(a, b, c, d, e) = IND(a, b, c, d)$. 所以, e 为可约简属性, 可从表 1 中去掉.

2.4 属性值约简

运用基于数据分析方法属性值约简算法可以得到属性值约简后的决策表, 去掉决策表中冗余属性值可以得到包含所有约简决策规则表(如表 2 所示).

表 2 去掉冗余属性值的约简决策表

论域(U)	条件属性(C)				决策属性(D)
	a	b	c	d	f
1	4	5	*	*	2
$1^{(1)}, 8^{(1)}$	4	*	*	1	2
$1^{(2)}, 7$	*	5	*	1	2
$1^{(3)}, 2^{(1)}, 7^{(1)}$	*	*	4	1	2
2, 3	*	4	*	*	2
$3^{(1)}$	4	*	*	2	2
$4, 5^{(1)}$	3	5	*	*	3
$4^{(1)}$	3	*	*	3	3
$4^{(2)}$	*	5	*	3	3
5	*	*	*	4	3
6	3	*	3	*	3
8	4	*	3	*	2
9	2	*	3	*	4
$9^{(1)}, 10$	2	*	*	3	4
$9^{(2)}$	*	2	3	*	4
$9^{(3)}, 10^{(1)}$	*	2	*	3	4
$9^{(4)}$	*	*	3	3	4
$11, 13^{(1)}$	*	3	4	*	4
$11^{(1)}, 13^{(2)}$	*	3	*	3	4
$11^{(2)}, 13^{(3)}$	*	*	4	3	4
12	2	*	*	2	3
$12^{(1)}$	*	2	*	2	3
13	4	*	*	3	4

14 * * 2 * 4

2.5 规则的产生

由表2可以得到最终控制规则如下:

a. $a_4(b_5 \vee c_3 \vee d_1 \vee d_2) \vee (b_5 \vee c_4) d_1 \vee b_4 \rightarrow f_2.$

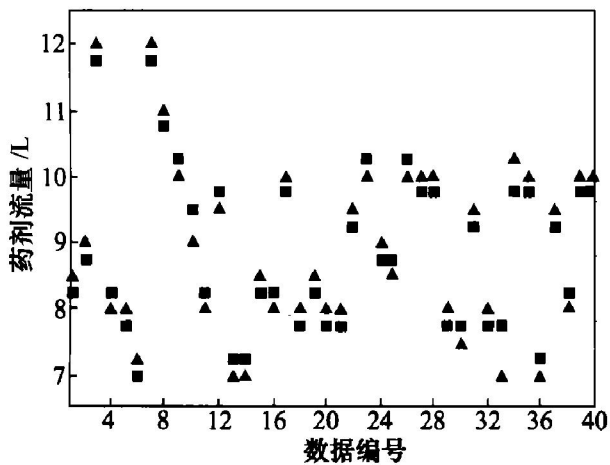
b. $a_3(b_5 \vee c_3 \vee d_3) \vee (a_2 \vee b_2) d_2 \vee b_5 d_3 \vee d_4 \rightarrow f_3.$

c. $(a_2 \vee a_4 \vee b_2 \vee b_3 \vee c_3 \vee c_4) d_3 \vee (a_2 \vee b_2) c_3 \vee b_3 c_4 \vee c_2 \rightarrow f_4.$

3 仿真与应用

3.1 准确度测试

利用现场数据进行仿真,将模型提供的药剂流量给定值与目前人工操作条件下的药剂流量给定值进行比较,称为准确度测试.由于浮选控制模型给出的值是离散的,代表一个区间值.所以,在模型给出系统输出后,还必须将离散值转化为连续值.这里在仿真时利用取区间数中点的方法所得值作为实际的输出.药剂流量的准确度仿真结果如图1所示,模型理论值与现场实测值的差值分布如图2所示.



■—现场值; ▲—模型值

图1 药剂流量准确度仿真图

从图1和图2可以看出,控制模型的理论值与现场实际测量值基本吻合,绝对误差不超过±0.5,这种误差是由于数据的离散化引起的,在正常误差范围之内.

3.2 命中率试验

命中率试验就是要把粗糙控制模型投入到浮选系统中进行试运行,通过考察浮选输出精矿品位来判断浮选系统的命中率,从而进一步检验粗糙控制

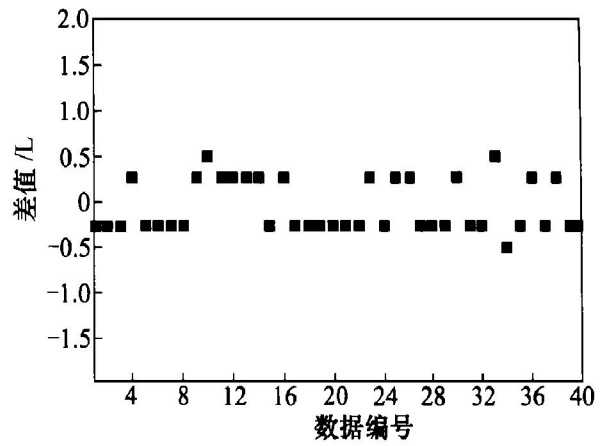


图2 模型值与现场值的差值分布
模型的准确度和可靠性.从试验数据记录中抽取1000组记录数据,浮选命中率如表3所示.可见,在浮选过程应用粗糙控制,达到了合理添加药剂的目的,浮选过程控制满足综合指标要求.

表3 命中率统计表

命中指标	记录数	所占比例/%
精矿品位 < 68%	87	8.7
68.0% ≤ 精矿品位 < 68.5%	285	28.5
68.5% ≤ 精矿品位 < 69.0%	334	33.4
69.0% ≤ 精矿品位 < 69.5%	246	24.6
精矿品位 ≥ 69.5%	48	4.8

4 结语

a. 在阳离子反浮选控制中运用粗糙集理论,便于现场技术人员和操作工人的操作,能合理添加药剂剂量,从而避免了传统工艺上由人为加药造成的浪费,给企业带来很大的经济效益.

b. 该方法还可以广泛应用于其他复杂问题的分析与控制.

参考文献:

[1] Pawlak Z. Rough sets[J]. International Journal of Information and Computer Science, 1982, 11: 341-356.
 [2] Szladow A, Ziarko Z. Rough sets: working with imperfect data [J]. AI Expert, 1993, 7: 36-41.
 [3] Slowinski R. Rough set approach to decision analysis[J]. AI Expert, 1995, 3: 19-25.
 [4] Pawlak Z. Rough sets[J]. Communications of ACM, 1995, 38 (11): 89-95.
 [5] 卢寿慈. 矿物浮选原理[M]. 北京: 冶金工业出版社, 1998.

- [6] Pawlak Z. Rough set: theoretical aspects of reasoning about data [M]. Boston: Kluwer Academic Publisher, 1991. 9-30. 出版社, 2001.
- [7] 王国胤. Rough 集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001.
- [8] 刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001.

Rough set theory and its application in cation anti-flotation control

ZHANG Yong^{1,2}, WANG Feng², PAN Xuejun¹, WANG Wei¹

(1. Dalian University of Technology, Dalian 116023, China;

2. Anshan University of Science & Technology, Anshan 114002, China)

Abstract: Intelligent control methods based on the rough set theory are proposed for the cation anti-flotation process due to its complexity, model uncertainty and the high requirement of its control quality. Using rough set theory to reduce the decision tables that are made up of some representative of state and the corresponding control strategy in the control process, a series of control rules can be obtained. Accurate rate test and process running test are done by using the proposed method. The test results show that control model value and field value are basically inosculated and the results satisfy the target request of floatation process control. Consequently rough set theory has an important application value in cation anti-flotation control, and facilitates the operations of the technical workers and operators and implements the reasonable addition of medicine. It avoids the waste caused by the manual addition traditionally. So it can bring large economic benefit. The proposed method has a general meaning, therefore it can be used in analyses and control of other complicated processes.

Key words: flotation process control; rough control; rough set; information system