

文章编号: 1001-0920(2010)11-1689-04

## 铜转炉吹炼过程熔剂加入量的模糊操作模式挖掘方法

胡志坤<sup>a</sup>, 桂卫华<sup>b</sup>, 阳春华<sup>b</sup>, 彭小奇<sup>b</sup>

(中南大学 a. 物理科学与技术学院, b. 信息科学与工程学院, 长沙 410083)

**摘要:** 将包含条件参数和需要决策的操作参数组成一个操作模式向量, 采用 Sugeno 模型构造一类基于操作模式的模糊推理系统以进行操作参数的决策. 首先利用先验知识对海量数据集进行初步分类; 然后采用一种基于模式相似度和相似矩阵的无监督聚类方法来辨识模糊操作模式决策的结构, 自动确定模式的数量和模式的值. 实例仿真表明, 利用该方法进行操作参数优化决策时简便有效, 可应用于 PS 转炉熔剂加入量的优化决策.

**关键词:** 操作模式; 模糊推理系统; 铜转炉

中图分类号: TP223

文献标识码: A

## A mining method for fuzzy operational patterns of the amount of flux in the process of copper converter

HU Zhi-kun<sup>a</sup>, GUI Wei-hua<sup>b</sup>, YANG Chun-hua<sup>b</sup>, PENG Xiao-qi<sup>b</sup>

(a. College of Physics Science and Technology, b. College of Information Science and Engineering, Central South University, Changsha 410083, China. Correspondent: HU Zhi-kun, E-mail: huzk@mail.csu.edu.cn)

**Abstract:** An operational pattern vector, which is made up of condition parameters and decision parameters, is constructed. A Sugeno fuzzy system is constructed based on vector set of operational patterns. The huge dataset of industrial data is initially classified according to professional knowledge. Similarity factor and similar matrix of operational patterns are constructed for a simply clustering process, which is used to identify the structure of the fuzzy system, and the number and initial value of patterns can be determined automatically. The numerical simulation shows that, the proposed method is simple and effective, which is applied to the determination of the flux amount of copper converting furnace in the process of matter converting, and the real result shows that the method works well.

**Key words:** Operational pattern; Fuzzy reasoning; PS converter furnace

### 1 引言

铜转炉的作用是通过铜铈和空气中的氧气进行剧烈的自热反应, 将  $\text{Cu}_2\text{S}$  氧化成金属铜, 通常分为造渣期和造铜期两个阶段. 造渣期的主要任务是通过加入石英熔剂与  $\text{FeO}$  反应, 生成熔融的  $2\text{FeOSiO}_2$ , 再利用它与铜铈的密度差异, 形成炉渣层. 其中熔剂直接决定了炉渣的最终成分, 对吹炼的状态有着非常重要的影响. 若不稳定控制渣的成分, 则会造成大量的渣进入造铜期, 导致有价金属损失, 恶化吹炼环境<sup>[1]</sup>.

目前, 铜转炉吹炼过程已积累了大量的生产数据, 这些数据实际上就是一个操作实例. 案例推理具有良好的解释性, 广泛应用于复杂工业系统操作参数的辅助决策<sup>[2-5]</sup>. 但由于熔剂数据都是数值型数据, 很难收集到能用语言表达的案例. 典型案例的具体数目

代表了知识的粒度, 一般很难确定, 需要进行灵活的变化, 聚类是一种有效的方法. 模糊  $C$  均值方法可获得若干类簇, 每个类簇为一个规则, 但它需要预先设定类别数, 还受初始点的影响, 学习效率较慢<sup>[6,7]</sup>. 山峰法及其改进的减法聚类等可以自动识别类别数, 但它无法自动对数据进行分类, 且计算量过大<sup>[8]</sup>. 模糊系统由于具有较好的语言表达能力, 非常适合于操作参数的辅助决策, 但其一般仅有一个语言变量<sup>[9]</sup>.

本文将生产系统的工艺参数组成一个“操作模式”或“操作实例”<sup>[10,11]</sup>, 提出一种快速迭代的聚类方法以辨识基于操作模式的模糊推理系统的结构, 自动确定模式的数量和数值, 设计一个基于操作模式的模糊推理系统, 并对推理系统的结论进行语言化, 从而实现铜转炉吹炼过程的熔剂加入量的优化决策.

收稿日期: 2009-09-29; 修回日期: 2009-12-11.

基金项目: 国家自然科学基金项目(60904077, 60874069); 国家863计划项目(2009AA04Z137).

作者简介: 胡志坤(1976—), 男, 湖北鄂州人, 副教授, 博士, 从事复杂有色冶金过程优化控制、故障诊断的研究;  
桂卫华(1950—), 男, 武汉人, 教授, 博士生导师, 从事复杂有色冶金过程优化控制、故障诊断等研究.

## 2 快速模糊聚类

设一个铜转炉熔剂加入量的历史数据集  $S = \{s_k\} \in R, k = 1, 2, \dots, n, n$  为记录数量. 则单个记录可表示为

$$s_k = \{t_k, b_k, x_k, y_k\}. \quad (1)$$

式中:  $s_k$  为第  $k$  个记录;  $t_k$  为记录产生的时间;  $b_k$  为边界条件, 一般不容易确定;  $x_k$  为记录  $s_k$  的描述生产过程变量, 设为条件参数,  $x_k = (x_k(1), x_k(2), \dots, x_k(p))$ ;  $y_k$  为需要决策的工艺参数, 设为操作参数向量,  $y_k = (y_k(1), y_k(2), \dots, y_k(q))$ .

边界条件在一定时间内通常是稳定的, 因此由历史数据集  $S$  可得到在一定时间内具有代表性的操作模式集  $P = \{p_1, p_2, \dots, p_m\}$ , 单个操作模式为  $P_j = \{\hat{x}_j, \hat{y}_j\}$ ,  $\hat{x}$  为条件参数向量,  $\hat{y}$  为决策参数向量,  $j = 1, 2, \dots, m$ .

定义两条数据记录的条件参数  $x_i$  和  $x_l$  的相似度为

$$r_{il} = \exp\left(-\frac{\|x_i - x_l\|^2}{2\sigma^2}\right), i, j = 1, 2, \dots, n, \quad (2)$$

式中采用高斯函数度量两条记录是否相似, 将相似度定义在  $[0, 1]$  之间, 可以形象地说明两者的相似程度. 如果  $x_i$  与  $x_l$  越相似, 则  $r_{il}$  越大. 当  $r_{il} = 1$  时, 表示  $x_i$  与  $x_l$  完全相似. 高斯函数的形状决定了当两个工况相差较大时, 相似度几乎为 0, 这为两个工况之间相似程度的评价提供了便利. 利用相似度构建如下  $n \times n$  的方阵作为关系矩阵, 来衡量数据集相互相似的程度:

$$R_{n \times n} = \{r_{il} | i, l \in [1, n]\}. \quad (3)$$

由于数据记录和操作模式中的操作参数向量有可能与条件参数向量重叠, 而其决策包含了人为因素, 应采用条件参数向量进行分类和聚类, 形成操作模式集. 基于聚类的模式挖掘过程其实质是由实例集浓缩为操作模式集的过程, 即

$$\begin{aligned} \{x_i, y_i, 1 \leq i \leq n\} &\rightarrow \{\hat{x}_j, \hat{y}_j, 1 \leq j \leq m\}, \\ m &< n. \end{aligned} \quad (4)$$

因事先不设定聚类中心, 故将原始记录都设为聚类中心, 即令  $x_i^{(0)} = x_i$ . 聚类步骤如下:

Step 1: 按照式 (2) 计算关系矩阵  $\{r_{il}\}$ , 设阈值为  $\zeta$ , 则有

$$\text{If } r_{il} < \zeta, \text{ then } r_{il} = 0. \quad (5)$$

Step 2: 按下式重新计算记录:

$$x_i^{(t)} = \sum_{l=1}^n r_{il} x_l^{t-1} / \sum_{l=1}^n r_{il}, t = 1, 2, \dots, \quad (6)$$

式中  $t$  为迭代步数.

式 (5) 和 (6) 保证了当两个向量之间的相似度不

大时, 不会影响迭代过程对记录的影响; 而相似的向量则会逐渐趋于相同的值.

Step 3: 如果所有记录都满足  $x_i^{(t)} = x_i^{(t-1)}$ , 则转向 Step 4; 否则转 Step 1.

Step 4: 最终获得的  $Z\{x_i^{(t)}\}$  会收敛到几个子集, 每个子集的所有向量均相等, 即

$$\begin{aligned} \{x_i^{(t)}\} &= \\ \underbrace{\{x_1^{(t)}, x_1^{(t)}, \dots, x_1^{(t)}, x_2^{(t)}, x_2^{(t)}, \dots, \dots\}}_{c_1} &\underbrace{\{x_2^{(t)}, x_2^{(t)}, \dots, x_2^{(t)}, x_2^{(t)}, \dots, \dots\}}_{c_2}, \\ \underbrace{\{x_m^{(t)}, x_m^{(t)}, \dots, x_m^{(t)}\}}_{c_m} &\Rightarrow \\ \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m\} &\in R^{n \times p}, \end{aligned} \quad (7)$$

式中  $c_j$  为第  $j$  类的个数. 式 (7) 的记录顺序可能不是完全按照记录集的顺序.

这样, 通过 Step 1~Step 4 便可确定数据集到模式集中条件集的聚类类别个数  $m$ , 类中心  $\{x_1^{(t)}, x_2^{(t)}, \dots, x_m^{(t)}\}$  和各条记录所在的类  $j (1 \leq j \leq m)$ , 并通过原来记录所对应的收敛向量确定原记录集所在的类别. 最后迭代的  $\{x_1^{(t)}, x_2^{(t)}, \dots, x_m^{(t)}\}$  即为最终获得的类中心  $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m\}$ .

例 1 考虑一个  $(x_1, x_2)$  的聚类问题,  $0 \leq x_1, x_2 \leq 1$ . 采用本文聚类方法进行迭代, 初始值、第 1~第 6 次迭代的取值如图 1 所示.

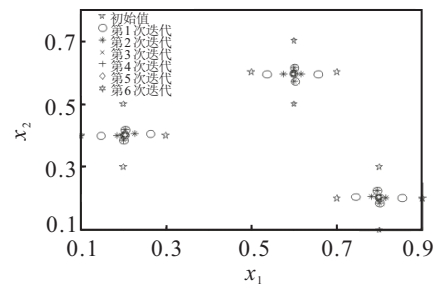


图 1 聚类的迭代过程

从图 1 可以看出, 在没有指定类别个数的前提下, 用本文聚类方法, 第 4 次迭代基本接近于类中心, 第 6 次迭代便收敛到 3 个聚类中心  $(0.5987 \ 0.5942), (0.2035 \ 0.4022), (0.7978 \ 0.2035)$ .

同样的数据, 采用 FCM 算法进行预置 3 个聚类中心的聚类, 第 7 步才收敛到聚类中心  $\{(0.6005, 0.6009), (0.8006, 0.1991), (0.1989, 0.3998)\}$ . 进行预置 4 个聚类中心的聚类, 则聚类到第 10 步才基本完成收敛. 显然, 该数据集的性质决定了预置 4 个聚类中心得到的结果肯定是错误的.

本文算法与 FCM 算法不同的是, 这种算法不需要设置误差阈值, 只要最后收敛的聚类中心不再改变, 即可认为收敛. 因此, 本算法具有良好的收敛性和快速性.

### 3 基于操作模式的模糊推理系统

设  $x^* = \{x^*(1), x^*(2), \dots, x^*(p)\}$  为表征当前工况的系统条件参数向量, 利用高斯函数定义当前工况的输出向量与第  $j$  个模式的相似度为

$$\lambda_j = \exp\left(-\frac{\sum_{i=1}^p \|x^* - \hat{x}_j\|^2}{2\sigma_j^2}\right) = \exp\left(-\frac{\sum_{i=1}^p (x^*(i) - \hat{x}_j(i))^2}{2\sigma_j^2}\right). \quad (8)$$

式中:  $i = 1, 2, \dots, p$  为变量的属性;  $j = 1, 2, \dots, m$  为模式的个数;  $\sigma$  为高斯函数的宽度;  $\lambda_j \in [0, 1]$ ;  $x^*$  为当前工况的条件参数向量.

根据背景知识将  $\lambda_j(x^*)$  分为若干等分, 用语言变量 (设为  $B_i, 1 \leq i \leq 5$ ) 表示为 S, MS, M, ML, L, 其含义为不近、偏近、较近、接近、很接近. 采用三角型隶属度函数来定义式 (8) 的相似度  $\lambda_j(x^*, \hat{x}_j)$  属于语言变量  $B_i$  的隶属度, 即

$$\mu(\lambda_j) = \begin{cases} \frac{\lambda_j - b_{i-1}}{c_i - c_{i-1}}, & b_{i-1} \leq \lambda_j \leq b_i; \\ \frac{\lambda_j - x}{b_{i+1} - b_i}, & b_i \leq \lambda_j \leq b_{i+1}. \end{cases} \quad (9)$$

式中  $2 \leq i \leq 4$ . 由于采用三角型隶属函数, 均有至多两个模糊区域与之对应, 设为  $\mu^{(1)}(\lambda_j(x^*, x_j))$  和  $\mu^{(2)}(\lambda_j(x^*, x_j))$ . 又设

$$\mu^{(1)}(\lambda_j(x^*, x_j)) > \mu^{(2)}(\lambda_j(x^*, x_j)), \quad (10)$$

显然有

$$\mu^{(1)}(\lambda_j(x^*, x_j)) + \mu^{(2)}(\lambda_j(x^*, x_j)) = 1. \quad (11)$$

根据背景知识, 采用同样的方法对操作模式的决策分量  $\hat{y}$  进行语言化, 得到两个隶属度  $\mu^{(1)}(\hat{y})$  和  $\mu^{(2)}(\hat{y})$ .

考虑一类多输入、单输出的模糊操作模式, 基于该操作模式集表示与推理如下:

$$\text{Pattern } j: \text{ if } x^* \text{ is } \lambda_j(x^*, \hat{x}_j) = \max_{1 \leq k \leq m} \{\lambda_k\}, \text{ then } y^* = \hat{y}_k, \text{ CF} = r. \quad (12)$$

式中:  $m$  为模式的个数,  $y^*$  为当前工况需要决策的操作参数,  $r(0 \leq r \leq 1)$  为由该模式进行操作参数决策的可信度,  $\lambda_j(x^*, \hat{x}_j)$  等于  $x^*$  与操作模式集中最接近的第  $j$  个模式  $\hat{x}_j$  的条件参数向量的相似度.

式 (12) 可用如下两个语言变量的模糊规则系统表示:

$$\begin{aligned} \text{Pattern } j: & \text{ if } x^* \text{ is } \hat{x}_i \text{ } B_i^{(l)} \hat{x}_j, \\ & \text{ then } y^* = \hat{y}_k, \\ & \text{ CF} = r = \mu^{(l)}(\lambda_j(x^*, \hat{x}_j)), \quad l = 1, 2, \end{aligned} \quad (13)$$

式中  $B_i^{(l)}$  表示“最接近”. 用两个语言变量的意义在于,

操作人员可根据实际工况进行判断和调整, 以便适应于特殊工况.

### 4 应用实例

由于生产数据集非常庞大, 仅依靠先验知识进行分类, 其分类的子集仍然非常大. 这里采用两级分类的方法进行操作模式的提取: 第 1 级采用先验知识进行预分类; 第 2 级采用聚类的方法进行无教师监督学习的分类, 即采用本文聚类方法进行分类器的学习, 然后将这些类别形成模糊规则. 具体过程如图 2 所示.

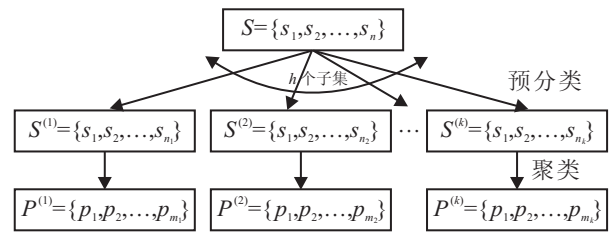


图 2 两级操作模式挖掘过程

每个数据子集通过聚类的方法得到一个操作模式集. 具体步骤如下:

Step 1: 以渣中 Cu 和 SiO<sub>2</sub>/Fe 先验知识进行预分类, 其中满足“Cu < 5%, SiO<sub>2</sub>/Fe = 0.50 ~ 0.53”的实例为 248 个样本.

Step 2: 采用上述聚类方法, 通过调节  $\zeta$  和  $\sigma$ , 得到初始的聚类中心  $\{x_1^{(t)}, x_2^{(t)}, \dots, x_m^{(t)}\}$ .

Step 3: 为了更加精确, 以此为初始值并采用最近邻方法进行聚类, 得到最终的聚类中心  $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m\}$ .

Step 4: 采用式 (12), (13) 对这些聚类中心进行模糊化, 得到所有的操作模式.

限于篇幅, 表 1 只列出了基于前 12 个模式的模糊推理系统结构.

图 3 给出了对历史数据集的实际操作值与采用模糊规则得出的决策结果比较. 从图 3 可以看出, 采用本方法能够较好地与数据集的操作实例相吻合.

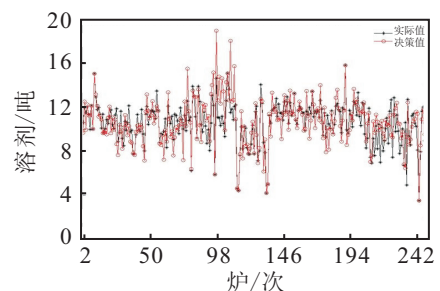


图 3 实际操作值与决策值的比较

石英石熔剂的作用是与 Fe 元素结合生成熔融的 Fe<sub>2</sub>O<sub>3</sub> · SiO<sub>2</sub>, 从而减少渣中 Cu 的含量, 降低有价

表 1 基于操作模式的模糊推理系统实例

序号	铜硫	Cu	S	Fe	SiO <sub>2</sub>	溶剂	l = 1		l = 2	
							$\hat{y}^{(1)}$	$\mu^{(1)}(\hat{y})$	$\hat{y}^{(2)}$	$\mu^{(2)}(\hat{y})$
1	120.0	57.13	22.52	14.43	0.2989	11.26	ML	0.915	L	0.085
2	125.0	59.09	23.30	13.71	0.3667	9.67	CE	0.621	S <sub>1</sub>	0.379
3	119.0	62.32	22.29	13.72	0.2945	8.29	S <sub>1</sub>	0.849	S <sub>2</sub>	0.152
4	129.0	61.20	22.45	11.86	0.3798	8.66	S <sub>1</sub>	0.918	S <sub>2</sub>	0.082
5	128.5	56.05	19.93	15.20	0.3113	14.03	B <sub>2</sub>	1	B <sub>1</sub>	0
6	119.5	58.64	22.84	15.36	0.3298	11.07	B <sub>1</sub>	0.967	B <sub>2</sub>	0.033
7	121.0	62.72	22.33	12.00	0.2914	10.16	CE	0.8	B <sub>1</sub>	0.2
8	122.0	62.54	22.58	10.97	0.2445	7.64	S <sub>1</sub>	0.745	S <sub>2</sub>	0.256
9	122.0	60.47	21.77	14.27	0.4097	13.24	B <sub>2</sub>	0.736	B <sub>1</sub>	0.264
10	119.5	58.64	22.84	15.36	0.3298	11.07	B <sub>1</sub>	0.974	CE	0.026
11	124.0	59.34	23.12	13.74	0.5950	9.96	CE	0.93	S <sub>1</sub>	0.07
12	118.0	58.84	22.99	14.19	0.3497	9.91	CE	0.923	S <sub>1</sub>	0.077

金属的损失. 如果石英石熔剂过多, 则会造成浪费, 且可能会与 Cu 元素结合, 造成有价金属的损失; 如果石英石熔剂过少, 则会恶化吹炼环境, 造成大量难溶的 Fe<sub>3</sub>O<sub>4</sub> 生成, 并导致大量的 Cu 元素进入渣中, 同时也会造成渣中 Cu 含量的增加. 将本文方法应用于某冶炼厂铜转炉熔剂加入量的优化决策. 首先从历史数据中挖掘出熔剂加入量的优化操作模式, 然后基于操作模式的模糊系统进行优化决策. 由于所有熔剂都能有规律地加入, 而且能够比较准确地反映吹炼过程的条件, 导致每炉吹炼后的炉渣渣型非常好. 在 1 个月的试用期间, 渣含铜的平均值为 4.64, 比前 2 个月降低了 0.5%, 从而降低了有价金属的损失.

## 5 结 论

本文提出了一种基于相似度和相似矩阵的快速聚类方法, 可从大量数据集中挖掘出操作模式. 该方法的优点是不需要复杂的迭代运算, 而且具有很强的非线性处理能力. 另外还设计了基于操作模式集的模糊推理系统, 该模糊系统能够充分利用操作模式的信息, 且具有较强的非线性处理能力和解释能力.

## 参考文献(References)

- [1] 任鸿九, 王立川. 有色金属提取冶金手册·铜镍[M]. 北京: 冶金工业出版社, 2000.  
(Ren H J, Wang L C. Nonferrous metal extractive metallurgy handbook[M]. Beijing: Metallurgy Industry Press, 2000.)
- [2] Yang Chunhua, Gui Weihua, Kong Lingshuang, et al. A two-stage intelligent optimization system for the raw slurry preparing process of alumina sintering production[J]. Engineering Application of Artificial Intelligence, 2008, 22(3): 796-805.
- [3] 柴天佑, 丁进良, 王宏, 等. 复杂工业过程运行的混合智能优化控制方法[J]. 自动化学报, 2008, 34(5): 505-515.

- (Chai T Y, Ding J L, Wang H, et al. Hybrid intelligent optimal control method for operation of complex industrial process[J]. Acta Automatica Sinica, 2008, 34(5): 505-515.)
- [4] 耿增显, 柴天佑. 基于案例推理的浮选过程智能优化设定[J]. 东北大学学报, 2008, 29(6): 761-764.  
(Geng Z X, Chai T Y. Intelligently optimal index setting for flotation process by CBR[J]. J of Northeastern University, 2008, 29(6): 761-764.)
- [5] Chan F T S. Application of a hybrid case-based reasoning approach in electroplating industry[J]. Expert System with Application, 2005, 29(1): 121-130.
- [6] Chen Ning, Chen An, Zhou Long-xiang. Efficient algorithms for mining fuzzy rules in large relational databases[J]. J of Software, 2001, 12(7): 949-959.
- [7] Kim E, Park M, Ji S, et al. A new approach to fuzzy modeling[J]. IEEE Trans on Fuzzy Systems, 1997, 5(3): 328-337.
- [8] Yager R R, Filev D P. Approximate clustering via the mountain method[J]. IEEE Trans on System, Man and Cybernetics, 1994, 24(8): 1279-1284.
- [9] Sushmita M, Sankar K P, Pabitra M. Data mining in soft computing framework: A survey[J]. IEEE Trans on neural networks, 2002, 13(1): 3-14.
- [10] 胡志坤, 桂卫华, 彭小奇, 等. 铜转炉生产操作模式智能优化[J]. 控制理论与应用, 2005, 22(2): 243-247.  
(Hu Z K, Gui W H, Peng X Q, et al. Intelligent optimization of optimal operational pattern in the process of copper converting furnace[J]. Control Theory & Application, 2005, 22(2): 243-247.)
- [11] 彭晓波, 桂卫华, 胡志坤, 等. 铜闪速熔炼过程操作模式智能优化[J]. 控制与决策, 2008, 23(3): 297-301.  
(Peng X B, Gui W H, Hu Z K, et al. Operational pattern intelligent optimization for copper flash smelting process[J]. Control and Decision, 2008, 23(3): 297-301.)