

文章编号:1000-582X(2006)10-0058-03

烧结矿 FeO 含量模糊聚类算法的研究与应用*

谢志江¹, 佟莹¹, 欧阳奇^{1,2}, 权国政¹

(重庆大学1. 机械工程学院; 2. 材料科学与工程学院, 重庆 400030)

摘要: 烧结机速、料层厚度和点火温度等烧结工艺参数对烧结机尾断面辐射图像特征分布有着重要的影响, 而断面辐射图像特征与烧结矿化学成分存在复杂的非线性映射关系. 根据方差分析理论, 运用混合统计量来确定烧结机尾辐射图像特征的最佳分类数, 并应用划分熵来验证分类数的正确性. 现场实验表明: 该算法正确, 能在线运行, 准确预报烧结矿 FeO 含量区间.

关键词: 聚类; 辐射图像; 烧结矿 FeO 预报; 聚类熵

中图分类号: TP206.1; TF046

文献标识码: A

目前, 在很多烧结现场, 仍然是由看火工通过经验和检测的过程参数判断烧结终点. 也有在烧结机尾安装工业摄像机, 使看火工能在烧结控制室根据烧结矿断面图像进行判断, 这种方法仅是看火工观看方法的改变, 其本质仍然是看火工的工作. 从大量的分析数据来看, 看火工看尾矿断面的经验和技巧对 FeO 含量及烧结过程的判断是接近实际情况的. 但是, 由于看火工素质的不同、经验的差异等主观因素影响, 会造成判断结果很不稳定. 国内外在烧结矿 FeO 预报方面进行了较深入的研究, 张石^[1]等应用烧结机尾断面图像特征值进行 FeO 回归分析, 但是误差大. 蒋大军^[2]等根据烧结操作参数建立 FeO 含量的 BP 网络模型, 该模型命中率高, 但是收敛速度慢, 很难在线运行. 本文利用机器视觉技术对烧结机尾断面图像进行特征提取, 运用聚类模式识别理论, 以混和 F 统计量来确定样本特征量最佳分类数, 并应用模糊划分熵验证分类数, 从而设计模糊聚类器, 对烧结矿 FeO 含量进行在线预报, 以指导烧结生产, 提高入炉烧结矿的质量.

1 提取烧结机尾断面图像特征

烧结机尾断面图像通过图像采集系统获取并转化为 32 bit 真彩色位图格式, 将图像经过滤波、增强、恢复、去噪、分割等预处理后, 去除背景图像, 提取火焰图像数字特征. 在模糊聚类分析中, 提取样本特征参数直接影响聚类结果的准确性. 根据烧结原理和看火工人的经验, 并参照模糊识别和图像处理的特征提取方法,

从图像中提取以下特征信息构成烧结断面辐射图像聚类分析的样本向量^[3].

1) 平均光强(w_1): 若 $p(i)$ 为第 i 级灰度的概率, 其中 $0 \leq i \leq L$, L 为图像灰度的最大值, m 为图像的实际最大灰度级数, 即:

$$w_1 = \frac{1}{m} \sum_{i=1}^m ip(i). \quad (1)$$

2) 光强方差(w_2): 若 M, N 为图像的高度和宽度 (均为像素数), $f(i, j)$ 为第 (i, j) 处的灰度, \bar{f} 为平均光强, 则:

$$w_2 = \sum_{i=1}^M \sum_{j=1}^N [f(i, j) - \bar{f}]^2 / (M \times N). \quad (2)$$

3) 断面红火层面积(w_3): 设 t 为红火层与非红火层分界的灰度阈值, 则:

$$w_3 = \sum_{f(i,j) > t} 1 / M \times N. \quad (3)$$

4) 断面图像灰尘量(w_4): 设 h_1 为红火层下缘的高度, h_2 为红火层上缘的高度, 则:

$$w_4 = \sum_{h_1 < f(i,j) < h_2} 1. \quad (4)$$

5) 火焰占满度(w_5): 若一幅 $M \times N$ 数字火焰图像, 假设其每一个像素所占的面积为 s , 那么整幅图像的面积:

$$s_{total} = M \times N \times s. \quad (5)$$

对该图像进行分割去除背景后, 假设火焰的图像面积 S_{flame} , 并且其总的像素数为 ψ , 则: $s_{flame} = \psi \times s$, 火焰丰度为:

* 收稿日期: 2006-06-11

作者简介: 谢志江(1963-), 男, 湖南湘乡人, 重庆大学教授, 博士生导师, 主要从事机械工程测试及设备故障诊断方面的研究.

$$w_5 = \frac{s_{flame}}{s_{total}}, \quad (6)$$

它反映了火焰占满程度。

2 确定最佳聚类数

设 $U = \{u_1, u_2, \dots, u_n\}$ 为有限样本集, U 中的元素有 m 个特征, 即 $u_i = (u_{i1}, u_{i2}, \dots, u_{im}), i = 1, 2, \dots, n$, 要 U 把化为 $C(2 \leq C \leq n)$ 类: A_1, A_2, \dots, A_c . 对应 U 的一个划分, 给出它的聚类中心^[4-5]: $V = \{v_1, v_2, \dots, v_c\}$, 其中 $v_i = \{v_{i1}, v_{i2}, \dots, v_{im}\}$, 令:

$$F(k) = \frac{\sum_{i=1}^c n_i (v_{ik} - \bar{v}_k)^2 (n - c)}{\sum_{i=1}^c \sum_{j=1}^{n_i} (u_{ijk} - \bar{v}_k)^2 (c - 1)}, k = 1, 2, \dots, p. \quad (7)$$

再令:

$$\text{mixed-}F = \frac{\sum_{k=1}^p \frac{1/F(k)}{\sum_{k=1}^p 1/F(k)} F(k)}{\sum_{k=1}^p 1/F(k)} = \frac{p}{\sum_{k=1}^p 1/F(k)}. \quad (8)$$

式(7)中 n_i 表示第 i 类样本数; v_{ik} 是 i 类样本第 k 个变量的聚类重心; \bar{v}_k 是第 k 个变量聚类重心的平均值; u_{ijk} 为 i 类第 j 个样本的第 k 变量值. 理论上可以证明, 统计量 $F(k)$ 和 Mixed- F 服从自由度为 $(c-1, n-c)$ 的 F 分布. 由于聚类分析中各类的边界是模糊的, 因而应用模糊划分熵 $H_{ij}(U; c)$ 来评价聚类效果的好坏, 它反映了 i, j 两类火焰图像样本的重叠程度^[6-7]. 令:

$$H_{ij}(U; c) = \left| \frac{1}{n_i} \sum_{k_i=1}^{n_i} (-u_i k_i \log 2 u_i k_i - u_j k_i \log 2 u_j k_i) \right| + \left| \frac{1}{n_j} \sum_{k_j=1}^{n_j} (-u_j k_j \log 2 u_j k_j - u_i k_j \log 2 u_i k_j) \right|, \quad (9)$$

构造目标函数^[8]:

$$J_2(U, V) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \sum_{k=1}^m (x_{jk} - v_{ik}). \quad (10)$$

在满足 $J_2(U, V)$ 达到最小时, 找到最佳分划矩阵 U^* 和对应的聚类重心 V^* , 即:

$$J_2(U^*, V^*) = \min(J_2(U, V)), U \in M_{fc}. \quad (11)$$

式(11)中 M_{fc} 表示:

$$M_{fc} = \left\{ u_{c \times n} \mid u_{ij} \in [0, 1], \forall i, j; \sum_{i=1}^c u_{ij}, \forall j; 0 < \sum_{j=1}^n u_{ij} < n, \forall i. (i = 1, 2, \dots, c; j = 1, 2, \dots, n) \right\}. \quad (12)$$

3 数据验证与讨论

从某烧结厂 6# 烧结机按照不同烧结工况采集 13 幅烧结机尾断面图像, 其中样本编号 0~9 作为分类器训练样本, 10~12 为测试样本(已知 10 号 FeO 含量适

当, 11 号 FeO 含量偏低, 12 号 FeO 含量偏高), 经图像预处理后, 提取平均光强 (w_1), 光强方差 (w_2), 断面红火层面积 (w_3), 断面图像灰尘量 (w_4), 火焰占满度 (w_5) 等特征向量, 如表 1 所示。

表 1 聚类器训练和测试样本特征向量

样本编号	w_1	w_2	w_3	w_4	w_5
0	53.5	0.130	0.013 0	0.79	0.013
1	51.1	0.130	0.006 0	0.79	0.200
2	41.6	0.110	0.001 8	0.81	0.180
3	55.9	0.130	0.003 3	0.75	0.240
4	50.8	0.118	0.001 2	0.75	0.240
5	78.0	0.122	0.019 0	0.68	0.290
6	54.9	0.115	0.004 0	0.76	0.230
7	54.1	0.126	0.007 0	0.79	0.200
8	52.7	0.116	0.004 2	0.79	0.200
9	51.4	0.131	0.004 5	0.75	0.240
10	47.3	0.122	0.004 6	0.69	0.300
11	68.0	0.132	0.023 0	0.70	0.310
12	41.7	0.130	0.001 7	0.79	0.170

记为 $u_i (i = 1, 2, \dots, 10)$, 作为论域: $U = \{u_1, u_2, \dots, u_{20}\}$, 每一个对象具有 5 个特征:

$$u_i = (u_{i1}, u_{i2}, u_{i3}, u_{i4}, u_{i5}), i = 1, 2, \dots, 10$$

其中 $u_{i1}, u_{i2}, u_{i3}, u_{i4}, u_{i5}$ 分别代表第 i 个图像样本 I_{max}, σ, H, E, A 特征值. 取 $p = 9$, 共有 $p - 2 = 7$ 种有意义的聚类方案, 对每种方案计算 Mixed- F 值, 得到对应不同分类的 Mixed- F , 见表 2 所示。

表 2 不同分类数对应的 Mixed- F 值

分类数	2	3	4	5	6	7	8
Mixed- F	1.78 40.	80.20.	30.13.	60.64	0.2.60	1.70	

从表 2 对应的 Mixed- F 值可知, 分类数为 3 时 Mixed- F 最大, 因此确定最佳聚类数为 3. 应用 ISODATA 模糊聚类算法, 确定最佳分划矩阵 U^* 和对应的聚类重心 V^* , 对辐射图像样本进行聚类, 聚类结果表 3 所示。

表 3 训练样本聚类结果

类 别	训练 样本	聚类重心					FeO 含 量范围/%
		w_1	w_2	w_3	w_4	w_5	
1	0, 1, 3, 4 6, 7, 8, 9	78.00	0.122	0.0190	0.680	0.290	7~9
2	2	53.05	0.125	0.0054	0.771	0.195	9~12
3	5	41.60	0.110	0.0018	0.810	0.180	>13

为了进一步验证分类的可靠性, 根据式(3)求解各类之间的模糊划分熵值, $H_{12}(U; c) = 0.14; H_{23}(U;$

$c) = 0.28; H_{13}(U; c) = 0.23.$

当各类别之间 $H_{ij}(U; c) \geq 0.70$ 认为类之间存在较大的相似性,可以归为同一类.从 $H_{ij}(U; c)$ 值可知,辐射图像样本归为 3 类是合理的,符合实际的 3 种烧结工况条件下不同 FeO 含量烧结机尾断面辐射图像的特征.将训练样本送入分类器,聚类结果如表 4 所示.

表 4 测试样本聚类结果

类别	训练样本	聚类重心					FeO 含量范围/%
		w_1	w_2	w_3	w_4	w_5	
1	0,1,3,4 6,7,8,9,11	73.00	0.127	0.02100	0.690	0.300	7~9
2	2,10	52.41	0.124	0.00531	0.762	0.207	9~12
3	5,12	41.65	0.120	0.00175	0.800	0.175	>13

从表 4 可知,11 号样本聚为第 1 类(FeO 含量偏低),10 号样本聚为第 2 类(FeO 含量适当),12 号样本聚为第 3 类(FeO 含量偏高),符合测试样本 FeO 含量聚类分布特征,该分类器能用于预报烧结矿 FeO 含量.

4 结论

对不同 FeO 含量范围烧结断面图像的特征进行了聚类分析,首先用 Mixed-F 统计量来确定最佳聚类

数,然后应用模糊划分熵来确定各类间的重叠程度,确定并验证最终分类数的正确性,最后应用模糊 ISODATA 法进行辐射图像的分类,以确定烧结断面图像与 FeO 含量的聚类相关性,满足现场实时预报的要求.

参考文献:

- [1] 张石,张宏勋,郑春清. 利用数字图像处理技术在线测量烧结矿 FeO 含量的研究[J]. 仪器仪表学报, 1998, 19(5): 489-451.
- [2] 蒋大军. 基于人工神经网络的烧结矿 FeO 预报系统[J]. 烧结球团, 2005, 30(3): 30-34.
- [3] 孙才志,王敬东,潘俊,等. 模糊聚类分析最佳聚类数的确定方法研究[J]. 模糊系统与数学, 2001, 15(1): 91-93.
- [4] 朱永红,杨敏. 模糊 ISODATA 聚类分析方法的 VC 实现[J]. 计算机应用, 2001, 21(8): 141-143.
- [5] BEZDEK JC. Pattern Recognition With Objective Function Algorithms[M]. New York: Plenum Press, 1981.
- [6] 钱夕元,邵志清. 模糊 ISODATA 聚类分析算法的实现及其应用研究[J]. 计算机工程与应用, 2000, 15: 69-71.
- [7] 王立新. 模糊系统与模糊控制[M]. 王迎军译. 北京: 清华大学出版社, 2003.
- [8] 于达仁,胡清华,鲍文,等. 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): 204-207.

Application and Research on Fuzzy Cluster with Sintering Quantity

XIE Zhi-jiang¹, TONG Ying¹, OUYANG Qi^{1,2}, QUAN Guo-zheng¹

(1. College of Mechanical Engineering, Chongqing University, Chongqing 400030, China;

2. College of Materials Science and Engineering, Chongqing University, Chongqing 400030, China)

Abstract: The sintering parameters, such as sinter trolley speed, stuff thickness and ignition temperature etc. have important influence on sinter section image character distribution, and the sinter chemical ingredient have complicated non-linear relation with sintering section image characters. The paper applied Mixed-F statistic analysis to determine the optimal class number of Fuzzy cluster of sintering section radiant image, and fuzzy partition entropy analysis to verify whether the class number is optimal. It is be proved that the algorithm is right and effective, and its application is helpful for the online forecasting sintering FeO range.

Key words: clustering; radiant image; FeO content forecasting; clustering entropy

(编辑 姚 飞)