

基于模糊聚类 and 模糊支持向量机的湿法炼锌净化除钴过程建模^①

朱红求^② 阳春华 桂卫华

(中南大学信息科学与工程学院 长沙 410083)

摘 要 提出了一种基于模糊 C-均值(FCM)聚类和模糊支持向量机(SVM)方法相结合的湿法炼锌净化除钴过程建模方法。该方法针对样本空间影响支持向量机泛化性能和样本数量影响计算复杂度的问题,首先采用模糊聚类将学习样本分类,然后在各个类的样本空间内采用模糊支持向量机进行训练,并对各支持向量机模型的输出加权作为过程模型的输出。以净化除钴过程生产数据进行实验验证的结果表明,该方法明显减少了模型的训练时间,模型具有精度高、泛化性能好等特点,可以用于净化过程的优化控制。

关键词 净化除钴过程, 模糊聚类, 支持向量机(SVM)

0 引 言

硫酸锌溶液净化过程是湿法炼锌生产中最为重要的一个过程,此过程的效果好坏不仅影响到电解过程电流效率和电能消耗,还影响到后续工序的正常生产和产品质量^[1]。钴是危害电解过程锌析出的最主要的杂质之一,也是最难除的杂质之一,净化除钴成为湿法炼锌过程的关键工段^[2]。目前,操作人员主要凭经验操作,往往为保证出口溶液杂质离子浓度合格而盲目过量添加锌粉,导致锌粉消耗大。在现有工艺条件和设备下,通过过程的操作优化减少锌粉消耗,具有重要的现实意义。

净化过程是一类液固两相长流程化学反应过程,其非线性大,流程长,影响因素多,建立净化过程的机理模型相当困难。基于工业过程大量的现场数据,采用支持向量机(SVM)建模成为当前的研究热点^[3-5]。但支持向量机的泛化能力受到样本数据分布特别是孤立点或者异常点的影响较大,学习算法的复杂度取决于支持向量的个数,这些问题影响了其应用效果。本文提出了一种结合模糊聚类和模糊支持向量机的过程建模方法,首先将学习样本聚类,然后对每一类的样本分别采用支持向量机进行描述,最后对各个子模型的结果综合。现场数据验证结果表明,该方法具有较好的泛化能力,模型预测精度高,能满足现场工艺要求。

1 模糊聚类分析

聚类分析是一种常用的数据分析方法,它根据预先设定的标准,将待分类的样本集根据样本的相似性分割成不同的类。模糊 C-均值(fuzzy C-means, FCM)聚类算法把模糊数学引入到聚类分析中,更具合理性,是使用最广泛的聚类技术之一^[6]。FCM 聚类算法是通过随机选取 C_f (C_f 为聚类数)个点作为初始聚类中心,通过迭代过程极小化目标函数实现样本的聚类。

设 $X = \{x_1, x_2, \dots, x_N\}$ 为样本集, x_j 为 D 维样本,即 $x_j \in R^D, j = 1, 2, \dots, N$ 。将输入样本集 X 划分为 C_f 个类别,那么输入样本分别属于 C_f 个类别的隶属度,可以表示为隶属度函数矩阵 $U = [u_{ij}]$, 其中 $u_{ij} (1 \leq i \leq C, 1 \leq j \leq N)$ 表示第 j 个样本属于第 i 个类别的隶属度,显然,它应该满足以下 3 个条件:

$$\begin{cases} \sum_{i=1}^{C_f} u_{ij} = 1, & 1 \leq j \leq N \\ 0 \leq u_{ij} \leq 1, 1 \leq i \leq C_f, 1 \leq j \leq N \\ 0 < \sum_{j=1}^N u_{ij} < N, & 1 \leq i \leq C_f \end{cases} \quad (1)$$

设 $V = \{v_1, v_2, \dots, v_{C_f}\}$ 是聚类中心,FCM 聚类算法的目标函数为^[7]

^① 863 计划(2009AA04Z124, 2009AA04Z124), 国家自然科学基金(60874069)和湖南省自然科学基金(09JJ3122)资助项目。

^② 男, 1970 年生, 博士生, 讲师; 研究方向: 复杂工业过程的建模与优化; 联系人, E-mail: hqcsu@mail.csu.edu.cn (收稿日期: 2009-08-05)

$$J(U, V) = \sum_{i=1}^{C_f} \sum_{j=1}^N u_{ij}^m \|x_j - \nu_i\|^2 \quad (2)$$

其中 $m(1 < m < \infty)$ 为权重系数,最佳取值范围为 1 到 2.5 之间,本文中取值为 2。 $\|x_j - \nu_i\|^2$ 表示每一个样本与聚类中心的距离。 $J(U, V)$ 表示各类中样本到聚类中心的加权距离平方和, $J(U, V)$ 越小说明聚类效果越好。聚类准则是求得适当的模糊划分矩阵 U 与模糊聚类中心 V , 使得目标函数 $J(U, V)$ 达到极小值。

采用拉格朗日极值法求式(2)的极小值,可得到如下解:

$$\nu_i = \frac{\sum_{j=1}^N (u_{ij})^m x_j}{\sum_{j=1}^N (u_{ij})^m}, i = 1, 2, \dots, C_f \quad (3)$$

$$u_{ij} = \frac{\left(\frac{1}{\|x_j - \nu_i\|^2} \right)^{1/m-1}}{\sum_{k=1}^{C_f} \left(\frac{1}{\|x_j - \nu_k\|^2} \right)^{1/m-1}} \quad (4)$$

通过迭代算法就可以近似地得到目标函数的最优值。反复迭代直到 $\|U^{k+1} - U^k\| \leq \epsilon$ 或达到最大迭代次数时停止迭代,得到最终的划分矩阵 U 和聚类中心矩阵 V 。

基本的 FCM 聚类算法依赖于样本之间的距离测度,很多情况下,FCM 使用假设各特征属性具有相同权重的欧拉距离算法,这种假设严重影响了 FCM 聚类的性能。但实际上不同的样本特征属性对于不同类别的贡献可能不同,当样本数据特征属性对分类结果影响差异较大的情况下,基本的 FCM 聚类算法将难以产生高质量的聚类结果。因此,通过对不同的属性赋以不同的权重,减少一些次要属性对聚类的干扰,提高样本聚类的效果。

设权重系数为 w ,则在样本空间定义样本 x_i 与 x_j 的加权欧氏距离为

$$\|x_j - \nu_i\|^2 = d_{ij}^2(x_i, x_j) = \sum_{k=1}^D w_k \times (x_{ik} - x_{jk})^2 \quad (5)$$

对于训练集中的每个样本,按照对各聚类中心的最大隶属度原则归入相应的类,如果某一类归入的样本数达不到指定的数目,则删除该类并将类内的样本按照隶属度最大原则并入其他类。同时,对每类内的样本进行加权欧式距离计算,将距离小于某一阈值的两个样本,删除样本时间最早的样本,这样以减少类内样本的数量。

2 模糊支持向量机

在生产过程中,由于矿源变化频繁,导致除钴过程入口溶液杂质离子浓度波动增大,样本波动范围增大,此外,生产过程中根据车间调度而进行的一些生产调整常常会造成一些孤立点的存在,样本数据的这些特点影响了过程支持向量机模型的训练。同时,在净化过程中,锌粉加料、溶液搅拌等设备随着时间的推移其性能会缓慢变化。由于支持向量机获得的最优平面仅依靠小部分样本,因而支持向量机对于训练集中数据噪音或者孤立点过分敏感^[8]。为了解决这一问题,模糊支持向量机方法对数据样本引入模糊隶属度,在分割超平面学习时不同的样本数据具有不同的影响效果^[9-11]。本文提出了基于时间域和空间域的复合模糊隶属度函数,以提高支持向量机模型的泛化性能。

根据 FCM 聚类的结果,对每一类内的样本采用模糊支持向量机进行建模,根据样本空间的分布和时间域上的分布情况,对每一类内不同的样本采用不同的模糊隶属度 μ_i ,模糊化后的建模样本集为 (x_i, y_i, μ_i) (其中: $i = 1, 2, \dots, N; 0 < \mu_i < 1$)。则支持向量机中的目标函数可重写为

$$\min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{C}{2} \sum_{i=1}^N \mu_i e_i^2 \quad (6)$$

$$\mu_i = \delta_i \mu_{ij} \quad (7)$$

$$\delta_i = \frac{N-i}{N} (\delta_{\max} - \delta_{\min}) + \delta_{\min} \quad (8)$$

式中 e_i 为松弛变量, C 为惩罚因子, μ_{ij} 为样本 x_i 属于第 j 个类的模糊隶属度值, δ_i 为基于时间域的线性加权系数,基于远小近大原则,减少样本数据中距离当前时刻较远的数据对训练过程造成的干扰。

通过对各个子类样本进行训练,获得各个子类的支持向量机模型,对各个子模型的输出进行综合,得到过程模型的最终输出:

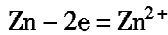
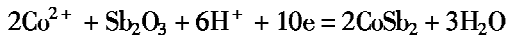
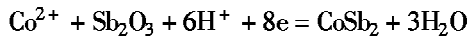
$$y_i = \frac{\sum_{k=1}^C \mu_{ik} y_k}{\sum_{k=1}^C \mu_{ik}} \quad (9)$$

式中 y_i 为样本 x_i 对应的输出, μ_{ik} 为样本 x_i 对第 k 个聚类的隶属度, y_k 为样本 x_i 对应第 k 个子模型的输出。

3 过程建模与验证分析

净化除钴过程流程长,溶液从流入反应器到流

出,前后需要 2h 以上,滞后时间大。净化除钴主要是通过添加锌粉和铈盐置换溶液中的钴离子^[12],铈盐作为活化剂,化学反应方程为



从上述反应方程式可以看出,硫酸锌溶液中加锌粉铈盐除钴过程复杂,并非简单金属置换反应,存在复杂的非线性关系。此外,影响的因素众多,包括反应温度、溶液 pH 值、搅拌速度、锌粉添加频率、铈盐流量等因素,且因素间互相影响,难于建立基于机理分析的过程模型。

通过机理研究与数据分析,影响净化除钴过程效果的主要因素包括入口镉离子浓度、入口溶液温度、入口溶液流量、铈盐流量、锌粉添加频率和入口钴离子浓度等因素。以上述因素为过程模型的输入变量,以净化过程出口溶液钴离子浓度作为过程的输出变量,收集生产过程数据记录作为训练样本,采用模糊聚类方法将样本集分类,每个分类中采用模糊支持向量机训练,再将各模型数据集成,就可以得到过程模型的输出。

为检验该过程建模方法的有效性,本文采用净化除钴生产过程中的数据,在经过数据预处理和样本相似性计算处理后,获得了 3200 组数据,随机选取其中的 100 组作为测试集,用来比较分析算法的性能。随机选取其中的 100 组数据作为验证样本集,采用粒子群算法对过程属性选择、权重系数等参数优化。剩下的 3000 组生产数据构建模型训练样本集。根据净化过程数据的分布情况,聚类个数初始值设定为 8,每个子集的样本不少于 50 个。通过聚类分析后,对每个类内的数据进行相似性处理,去除相似度大于 0.9 的样本。每一个样本集采用模糊支持向量机进行训练,惩罚系数 C 以及径向基核函数的宽度参数 σ 初始值相应选取为 40、0.5。

为了验证所提方法的性能,分别在训练样本集中随机选取了 500 组、1000 组、2000 组和 3000 组数据作为训练样本集,以此分析所提方法在计算复杂度方面的性能。在 P4 2.8G/512M RAM 计算机上,在 Windows XP 操作系统和 MATLAB 7.01 仿真软件平台下,从训练时间和测试集估计精度上进行了比较,比较结果如表 1 所示。其中普通最小二乘支持向量机核函数为径向基函数(RBF)(σ 为 0.5), C 为 100,采用相对均方根误差($RMSE$)作为评价指标。

表 1 两种算法的测试结果比较

训练样本数	方法	训练时间(s)	$RMSE$
500	标准 SVM	4.19	9.931%
	FCM-SVM	4.0523	9.877%
1000	标准 SVM	22.23	8.98%
	FCM-SVM	4.2558	8.71%
2000	标准 SVM	112.68	8.77%
	FCM-SVM	7.1031	8.472%
3000	标准 SVM	308.79	8.198%
	FCM-SVM	17.1504	7.954%

从表 1 中可以看出,标准 SVM 模型和 FCM-SVM 模型的相对均方根误差均不超过 10%,说明基于支持向量机的过程模型具有较好地跟踪样本数据变化趋势的能力,FCM-SVM 模型的泛化性能好于标准 SVM 模型。此外,可以明显看出,经过模糊聚类数据预处理的 SVM,训练时间大为缩短,有效提高了在大样本情况下的计算性能。当样本数较少时,时间减少并不显著,当学习样本集中样本数超过 1000 时,时间减少约 80.86%,样本数继续增加则训练的时间减少得更多,在样本数超过 2000 时 FCM-SVM 方法减少训练时间达 93.7%;在样本数超过 3000 时 FCM-SVM 方法减少训练时间达 94.45%。因此,在样本数较大时,FCM-SVM 方法可以大幅度减少训练时间,显著改善支持向量机的计算性能。

采用净化过程生产现场的 100 组生产数据,对所建立的过程模型进行了验证,结果如图 1 所示。

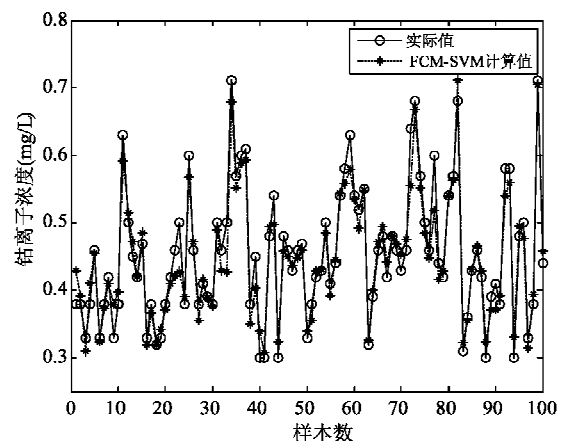


图 1 FCM-SVM 模型验证结果

通过分析,FCM-SVM 模型仿真结果中相对误差大于 10% 的样本仅占 9%,小于 5% 的样本达 71%,最大相对误差为 15.15%,平均相对误差为 4.28%,说明模型具有较高的精度。而相对均方根误差为

5.54%,说明过程模型跟踪样本数据变化趋势的能力强,具有较强的泛化能力。此外,根据工厂现场的工艺要求,在溶液杂质离子浓度小于 1mg/L 的范围内,测量绝对误差要求小于 0.1mg/L,而模型的绝对误差最大为 0.085mg/L。可见,模型精度满足现场控制要求,可以用来指导过程的操作优化。

4 结 论

针对湿法炼锌净化除钴过程机理复杂、流程长以及影响因素多等特点,提出了基于模糊聚类和支撑向量机相结合的净化除钴过程模型。通过对输入样本空间的模糊聚类以及样本模糊隶属度的定义,提高了支撑向量机泛化能力并减少了模型的训练时间。采用现场生产数据的验证分析结果表明,所提出的方法在输入样本空间大的情况下显著减少了训练时间,所建立的净化除钴过程模型具有精度高、泛化能力强等特点,该模型可以用于净化过程的过程操作优化。

参考文献

- [1] Raghavan R, Mohanan P K, Verma S K. Modified zinc sulphate solution purification technique to obtain low levels of cobalt for the zinc electrowinning process. *Hydrometallurgy*, 1999, 51(2): 187-206
- [2] Boyanov B S, Konareva V V, Kolev N K. Purification of zinc sulfate solutions from cobalt and nickel through activated cementation. *Hydrometallurgy*, 2004, 73(1): 163-168
- [3] Li D C, Fang Y H. An algorithm to cluster data for efficient classification of support vector machines. *Expert Systems with Applications*, 2008, 34(3): 2013-2018
- [4] Çomak E, Arslan A. A new training method for support vector machines: clustering k-NN support vector machines. *Expert Systems with Applications*, 2008, 35(3): 564-568
- [5] Kulkarni A, Jayaraman V K, Kulkarni B D. Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. *Computers & Chemical Engineering*, 2005, 29(10): 2128-2133
- [6] Chen W C, Wang M S. A fuzzy c-means clustering-based fragile watermarking scheme for image authentication. *Expert Systems with Applications*, 2009, 36(2): 1300-1307
- [7] Chen T. Incorporating fuzzy C-means and a back-propagation network ensemble to job completion time prediction in a semiconductor fabrication factory. *Fuzzy Sets and Systems*, 2007, 158(19): 2153-2168
- [8] Xu Q, Zhou H, Wang Y J, et al. Fuzzy support vector machine for classification of EEG signals using wavelet-based features. *Medical Engineering & Physics*, 2009, 31(7): 858-865
- [9] Lin C F, Wang S D. Training algorithms for fuzzy support vector machines with noisy data. *Pattern Recognition Letters*, 2004, 25(14): 1647-1656
- [10] 杨海军, 太雷. 基于模糊支持向量机的上市公司财务困境预测. *管理科学学报*, 2009, 12(3): 102-110
- [11] Shieh M D, Yang C C. Classification model for product form design using fuzzy support vector machines. *Computers & Industrial Engineering*, 2008, 55(1): 150-164
- [12] 郭忠诚, 阎江峰, 杨显万. Sb_2O_3 在硫酸锌溶液净化除钴过程中的机理. *中国有色金属学报*, 2000, 10(5): 697-700

Modeling of cobalt removal purification process in zinc hydrometallurgy based on FCM-SVM

Zhu Hongqiu, Yang Chunhua, Gui Weihua

(School of Information Science and Engineering, Central South University, Changsha 410083)

Abstract

A modeling method for cobalt removal purification process in zinc hydrometallurgy based on the combination of fuzzy C-means (FCM) clustering and fuzzy support vector machine (SVM) is proposed. Considering that the sample space influences the performance of SVM and the number of samples influences the computation complexity, the method divides the original sample space into subspaces by the FCM clustering firstly, then, trains the samples in each subspace by the fuzzy SVM, and finally, establishes the process model and takes the weighted output of each fuzzy SVM model as the output of the process model. The experimental results show that the proposed method greatly reduces the training time. Moreover, the proposed model has the high accuracy and good generalization performance. It can be used for operation optimization of the purification process of cobalt removal.

Key words: purification process of cobalt removal, fuzzy C-means clustering, support vector machine (SVM)