

摘要

针对目前赖氨酸生产过程中发酵产品质量难以实时测量, 现有软测量模型精度不高、测量结果受不确定因素影响较大的问题, 提出一种基于ISCA-SVR的赖氨酸发酵过程多模型软测量方法。首先, 利用一种改进的满意聚类算法(Improved Satisfactory Clustering Algorithm, ISCA)将样本数据集根据输入数据间的相似性准则划分为c个子集; 其次, 利用支持向量回归机(Support Vector Machine Regression, SVR)对每个数据集分别构建子模型; 然后对每个子模型进行加权融合, 并建立了赖氨酸发酵过程中关键变量的智能实时监控体系, 并建立网络通信模块, 与互联网相连实现资源共享。实验仿真结果表明, 相较于传统单一SVR预测模型, ISCA-SVR模型预测精度提高了4.5%, 提高了泛化能力, 并给其他复杂发酵过程参数的软测量提供了理论参考。

提出的方法

1. 算法总体结构图

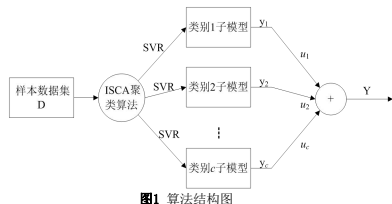


图1 算法结构图

2. 实现过程

- 基于改进满意聚类的数据集划分方法的具体步骤为:
 - Step 1: 令初始聚类个数 $c=2$; 最大聚类数 $c_{max} = \min(\frac{N}{m})$ (N 为数据集样本总数, m 为聚类子集期望最少样本数, 本文取 $m=30$)。
 - Step 2: 利用 FCM 算法将样本集合 D 进行聚类, 得出隶属度矩阵 $U=[u_{ij}]_{c \times N}$, 其中 u_{ij} 是第 j 个样本对第 i 类的隶属度, u_{ij} 范围在 $[0,1]$ 内, 然后根据每组样本所属子集的隶属度选取最大值进行分类, 将 D 分为 c 个子集 $\{D_1, D_2, \dots, D_c\}$;
 - Step 3: 对聚类后生成的每个子集采用最小二乘算法辨识各线性子模型的参数。
 - Step 4: 用隶属度矩阵 $U=[u_{ij}]_{c \times N}$ 作为连接各子模型的系数。
 - Step 5: 计算聚类性能指标 Sc 。若 $c < c_{max}$, 转 Step 6, 否则转 Step 8。
 - Step 6: 在样本集中, 根据隶属度矩阵 U 找出一个与各子集最不相似的样本 X_{ij} ($i=1, 2, \dots, c$) 做新的聚类中心 v_{c+1} 。为避免噪声, 一般应找出几个类似的样本求其平均值, 作为新的聚类中心 v_{c+1} 。
 - Step 7: 令 $c=c+1$ 以 v_1, v_2, \dots, v_{c+1} 为新的聚类中心, 重新计算新隶属度矩阵 U , 重新划分数据集。转 Step 3;
 - Step 8: 选取系统性性能指标最小时所对应的 c 的值作为最优聚类。

2.2 支持向量回归机

对于回归问题, 对于给定训练样本集 $D=(x_1, y_1), \dots, (x_n, y_n)$, $x, y \in R$, 首先考虑用线性回归函数 $f(x)=w \cdot x + b$ 。首先根据结构风险最小化原则将其归纳为寻找最小 w 的凸优化问题; 其次引入 Lagrange 函数, 并根据鞍点定理和 KKT 条件, 引入核函数, 上述凸优化问题转化成了

$$\max_{a, a'} - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left\{ (a_i - a'_i) (a_j - a'_j) K(x_i, x_j) \right\} + \sum_{i=1}^n \left\{ a_i (y_i - \varepsilon) - a'_i (y_i + \varepsilon) \right\}$$

通过选择拉格朗日乘子 a_i 和 a'_i 大于零的支持向量, 回归函数可重新表示为:

$$f(x) = \sum_{i=1}^p (a_i - a'_i) K(x, x_i) + b$$

其中 p 是支持向量的数目。选择高斯径向基函数核函数。

2.3 子模型加权融合

多模型的输出是所有子模型的权重连接, 可表示为

$$Y = U^T Y_i = u_1 y_1 + u_2 y_2 + \dots + u_c y_c$$

式中 U 为隶属度矩阵。

3. 算法流程图

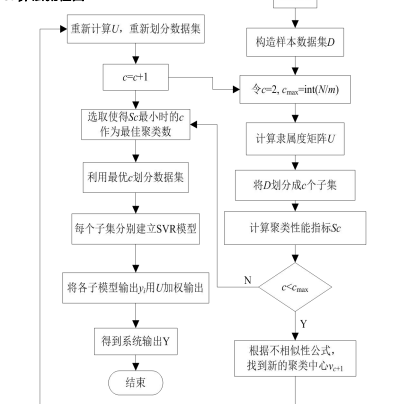


图2 ISCA-SVR算法流程图

4. 硬件系统

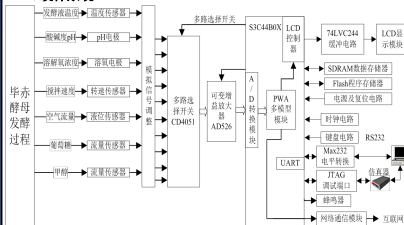


图3 监控系统总体硬件结构图

实验

1. 仿真结果

因篇幅有限, 在此仅给出产物浓度 P 和菌体浓度 X 的预测曲线对比图

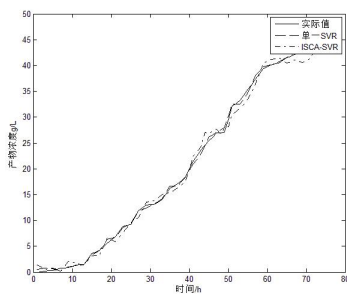


图4 产物浓度预测比较图

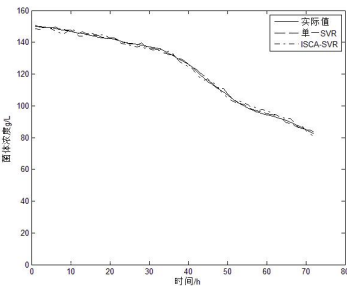


图5 菌体浓度预测比较图

预测输出	ISCA-SVR			单一-SVR		
	X/g·L ⁻¹	S/g·L ⁻¹	P/g·L ⁻¹	X/g·L ⁻¹	S/g·L ⁻¹	P/g·L ⁻¹
一批数据	1.38%	1.63%	1.97%	4.62%	3.36%	5.76%
二批数据	1.43%	1.72%	2.06%	5.11%	2.98%	6.32%
三批数据	1.51%	1.81%	2.01%	6.01%	2.66%	5.69%

表1 误差对比表

结论

ISCA-SVR 预测模型相较于传统 SVR 预测模型, 预测精度提高了 4.5%, 提高了泛化能力和稳定性。

主要参考文献

- [1] Wang Bo, Shahzad Muhammad, Zhu Xiang lin, et al. Soft-sensor modeling for L-lysine fermentation process based on hybrid ICS-MLSSVM[J]. Scientific reports, 2020, 10(1).
- [2] 陈康, 张相芬, 马燕, 等. 基于马尔可夫链和模糊 C-均值聚类的 DTI 图像分割算法[J]. 上海师范大学学报 (自然科学版) 2020, 49(01): 48-55.
- [3] 李行, 李少远, 席裕庚. 基于满意聚类的多模型建模方法[J]. 控制理论与应用, 2003(05): 783-787.
- [4] 高云龙, 王志豪, 潘金艳, 等. 基于自适应松弛的鲁棒模糊 C 均值聚类算法[J]. 电子与信息学报, 2020, 42(07): 1774-1781.

联系方式

联系人: 王星星
手机: 17851567860
邮箱: 1358735540@qq.com