

催化裂化反再系统的故障诊断方法研究

程 明, 栾秋波

(南京工业大学 自动化与电气工程学院, 南京 211816)

摘要: 考虑到石油化工过程系统复杂, 变量繁多, 非线性关系极强, 故障样本数据难于获取, 故利用支持向量机对炼油厂催化裂化装置反应再生子系统的故障状态进行模式识别; 且支持向量机参数 C 、 σ 对分类精度有很大影响, 采用了改进的遗传算法对其进行优化; 并采用了决策树算法进行多类分类, 根据类间分离测度, 从最难分类 (类间分离测度最小) 的两类样本集开始训练, 将其合并为一个类簇后同其他样本集一起, 再从中寻找最难分类的两个样本集合并, 如此逐步合并最终得到训练模型; 实验结果表明, 利用改进的遗传算法优化惩罚系数 C 和核函数参数 σ 后, 缩短了分类时间, 提高了分类准确率, 基于决策树算法的支持向量机有效地解决一对一和一对多分类算法中存在的无法辨识区域的问题, 能很好地识别故障类型, 对催化裂化装置的故障诊断有显著的指导作用。

关键词: 支持向量机; 决策树; 催化裂化; 故障诊断; 遗传算法

Research on Fault Diagnosis of Reaction—Regenerator System of FCCU

Cheng Ming, Luan Qiubo

(College of Automation and Electrical Engineering, Nanjing University of Technology, Nanjing 211816, China)

Abstract: The system of petrochemical process is complex, its variables are various, and their relationship is strongly nonlinear, so use support vector machine to identify the pattern of the fault state on the regenerative subsystem of the catalytic cracking unit in oil refineries. Taking into account the parameters C 、 σ have a great influence on the classification accuracy, so an improved genetic algorithm is used to optimize the parameters; And chose decision tree algorithm to classify, according to the separation measure between different categories, start training from the more difficult and most difficult classification sample sets, that is the separation measure is minimum, then combine them into a class cluster and put together with other sample sets, find two of the most difficult classification sample sets and combine them again, combine clusters like this and get the training model finally. The results show that after using the improved genetic algorithm to optimize penalty factor C and the kernel function parameter σ , the time of classification is shortened and the accuracy of classification is improved. The algorithm based on decision tree support vector machine effectively solve the problem that area cannot be identified that exists in one against one and one against other multiple classification algorithm, can identify the fault types well and has a significant guidance on the fault diagnosis of FCCU.

Key words: SVM; decision tree; FCCU; fault diagnosis; GA

0 引言

催化裂化装置是炼油厂中最为重要的装置之一, 核心为反应再生子系统, 其运行的好坏直接影响到产品质量、收率和消耗等经济性指标^[1-2]。因此, 对反再系统的故障诊断研究有着十分重要的意义。

目前对于催化裂化装置的诊断有基于神经网络、基于专家系统等方法。基于神经网络的诊断方法存在过学习和欠学习的情况, 泛化能力不足。基于专家系统的诊断方法存在局限性, 当遇到一个没有相应规则与之对应的新故障时, 系统则显得无能为力^[3]。支持向量机 (SVM) 是以统计学习理论为基础, 基于结构风险最小化原则的新型机器学习技术。由于其在处理小样本问题和模式识别问题上的出色表现, 正适合处理石油化工过程故障诊断领域中很难获得故障样本这一问题^[4]。遗传算法 (GA) 是一种借鉴生物界的进化规律演化而来的搜索全局最优解的方法^[5], 利用改进的遗传算法 (IGA) 对支持向量机

的参数进行寻优, 可以提高分类结果的准确率。

1 支持向量机的基本原理

支持向量机是有 Vapnik 与其领导的贝尔实验室的研究小组一起开发出来的一种建立在统计学习理论的 VC 维和结构风险最小化原理基础上的新的机器学习技术^[6-7]。

1.1 线性可分问题

SVM 方法是从线性可分情况下的最优分类面发展而来的, 其基本思想可用如图 1 所示的二维平面的情况来说明。

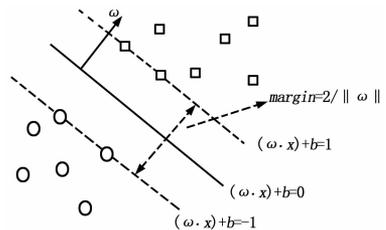


图 1 两类线性可分的最优超平面

在图 1 中, 方框点和圆点代表两类样本, 中间的实线为分类线, 其附近的两条虚线分别为过各类中离分类线最近的样本且平行与分类线的直线, 它们之间的距离就是分类间隔 (mar-

收稿日期: 2014-01-20; 修回日期: 2014-03-18。

作者简介: 程 明 (1964-), 男, 江苏南京人, 副教授, 主要从事过程系统控制理论与优化策略、系统建模理论与仿真技术方向的研究。

gin)。所谓最优分类线就是要求分类线不但能将两类正确分开, 即训练错误率为 0, 而且使分类间隔最大。使分类间隔最大实际上就是对推广能力的控制, 这是 SVM 的核心思想之一^[8]。

在线性可分条件下构建最优超平面, 可以转化为下面的二次规划问题:

$$\begin{cases} \min \frac{\|\omega\|^2}{2} \\ \text{s. t. } y_i((\omega \cdot x_i) + b) \geq 1 \end{cases} \quad (1)$$

式 (1) 中, 最优解 ω^*, b^* 为下面的拉格朗日函数的鞍点:

$$L(\omega, b, a) = \frac{1}{2}(\omega \cdot \omega) - \sum_{i=1}^l a_i [y_i((\omega \cdot x_i) + b)] \quad (2)$$

其中: $a \geq 0$ 为拉格朗日乘数。因此, 决策函数集为:

$$g(x) = \text{sgn}[(\omega^* \cdot x) + b^*] \quad (3)$$

1.2 线性不可分问题

非线性 SVM 问题的基本思想是通过非线性变换, 将输入变量 x 转化到某个高维空间中, 然后再变换空间求最优分类面。这种变换可能比较复杂, 因此这种思路在一般情况下不易实现^[9]。但是要注意到, 上面的对偶问题都只涉及训练样本之间的内积运算:

$$(x_i, x_j) \quad i, j = 1, \dots, l \quad (4)$$

即在高维空间只需进行内积运算, 而这种内积运算是可以用原空间中的函数实现的, 甚至没有必要知道变换的形式。根据泛函的有关理论, 只要一种核函数:

$$K(x_i, x_j) \quad i, j = 1, \dots, l \quad (5)$$

满足 Mercer 条件, 它就对应某一变换空间中的内积。

如果定义:

$$K(x, y) = \Phi(x) \cdot \Phi(y) \quad (6)$$

那么“最大间隔”非线性支持向量机的目标函数就变为:

$$W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{j=1}^l y_j a_j a_i K(x_i, x_j) \quad (7)$$

相应的决策函数为:

$$f(x) = \text{sgn}[\omega \cdot \Phi(x) + b] = \text{sgn}\left[\sum_{i=1}^l y_i a_i K(x_i, x_j) + b\right] \quad (8)$$

选择不同的核函数就可以产生不同的支持向量机, 常用的方法有以下几种:

(1) 线性核: $K(x, y) = x \cdot y \quad (9)$

(2) 多项式核:

$$K(x, y) = [(x \cdot y) + c]^d, c \geq 0 \quad (10)$$

(3) 高斯 (RBF) 核:

$$K(x, y) = \exp\left\{-\frac{\|x - y\|^2}{2\sigma^2}\right\} \quad (11)$$

(4) Sigmoid 核:

$$K(x, y) = \tanh[k(x \cdot y) - \delta], k > 0, \delta > 0 \quad (12)$$

2 支持向量机决策树 (SVMMDT) 算法

在采用一对一算法和一对多算法进行多类分类时, 会存在“拒绝分类”和“重叠区域”, 即无法辨识区。支持向量机决策树的基本思想是与基于二叉树的支持向量机采用反向思维,

SVMMDT 是基于类族展开的, 所谓“类簇”是指多类样本空间分布局部集中可以将其看作一个新集合类, 即形成所谓的“簇”^[10-11]。

2.1 类间分离性测度

通常采用欧氏距离 (类间中心点的距离) 来衡量类间分离测度。但欧氏距离不能很科学评定类间分离的难易程度, 不能代表其类间的分离度。因此, 定义了一种类分布的类间分离性测度来评定两类的分离性强弱。在进行 k 类分类, 训练样本集由类 $x_i, i=1, 2, \dots, k$ 组成, 则 sm_{ij} 表示类 i 和类 j 间的类间分离性测度^[12]:

$$sm_{ij} = \frac{d_{ij}}{(\sigma_i + \sigma_j)} (i = 1, 2, \dots, k) \quad (13)$$

其中: d_{ij} 表示类 i 和类 j 间的中心距离;

$$d_{ij} = \|c_i - c_j\| \quad (14)$$

c_i 是根据训练样本计算出的类中心:

$$c_i = \frac{1}{n_i} \sum_{x \in X_i} x, i = 1, 2, \dots, k \quad (15)$$

$$\sigma_i = \frac{1}{n_i - 1} \sum_{x \in X_i} \|x_i - c_i\|, i = 1, 2, \dots, k \quad (16)$$

其中: n_i 为类 x_i 中的样本个数; σ_i 表示类 i 的标准差, σ_j 表示类 j 的标准差, 其表示类的分布情况。

对于非线性的训练样本集, 经非线性映射 Φ 作用后, 在特定空间 H 中类 i 和类 j 间的分离性测度为:

$$sm_{ij} = \frac{d^H(m_i^i, m_i^j)}{(\sigma_i^H + \sigma_j^H)} \quad (17)$$

$$d^H(z_1, z_2) = \sqrt{K(Z_1, Z_1) - 2K(Z_1, Z_2) + K(Z_2, Z_2)} \quad (18)$$

$$m_\phi = \frac{1}{n} \sum_{i=1}^n \Phi(x_i) \quad (19)$$

$$\sigma^H = \frac{1}{n-1} \sum_{i=1}^n d^H(x_i, m_\phi) \quad (20)$$

其中: $K(\cdot, \cdot)$ 是核函数, $d^H(z_1, z_2)$ 是经非线性映射作用后, 在特征空间 H 中, 训练样本间的欧式距离, 同时输入空间样本的中心经映射后得到不在是特种空间中的训练中心, 特征空间样本的中心向量为 m_ϕ , n 为样本个数, σ^H 特征空间中的类方差。

2.2 SVMMDT 的训练过程及分类过程

支持向量机决策树是在训练过程中从类间最小距离入手, 分类过程中则是类间最大距离。训练时根据类间分离测度, 找到最难分类的两个样本集 (单一类别样本集), 对它们进行训练, 再将它们合并成一个类簇, 并与其余的样本集一起, 再找出最难分类的两个样本集 (可能是单一样本集, 也可能是类簇), 这样逐层合并形成类簇来建立训练模型; 而分类过程则是训练过程的逆过程, 从决策树的根节点到叶子节点逐层判断, 最后得到分类结果。SVMMDT 的训练过程和分类过程分别如图 2 和图 3 所示。

3 改进的遗传优化算法

根据生物进化的模型, John H. Holland 提出了一种优化算法——遗传算法。一般的遗传算法由选择 (Selection)、交叉 (Crossover) 和变异 (Mutation) 3 个部分构成。选择、交

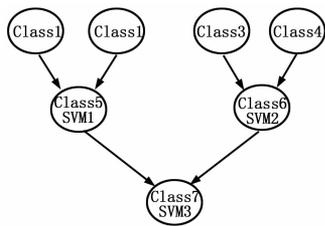


图 2 SVMDT 训练过程

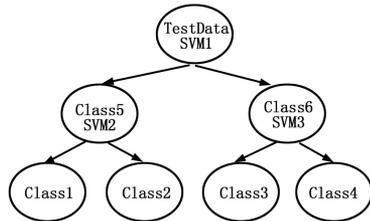


图 3 SVMDT 分类过程

又和变异操作由种群的适配值或适应度函数的概率确定^[13-14]。改进遗传算法与标准遗传算法的主要区别是在选择和交叉两步遗传操作：选择操作中采用了代沟选择和基于适应度值的重插入，确保当前种群中最适应的个体总是被连续地传播到下一代；交叉操作中采用可变交叉概率，使进化后期优化的对象比较容易稳定，以减少无用计算^[15]。

基于改进的遗传算法具体步骤如下：

(1) 给定参数 C 、 σ 的范围，并对其进行二进制编码。

(2) 适应度评估。采用 K -折交叉验证（即将数据集分成 K 份， $K-1$ 份用于训练，1 份用于测试）方法确认评估个体的适应度。本文选取 10 折交叉验证方法。

(3) 设置初始种群数量和最大进化代数。

(4) 计算群体中各个体的适应度，检查是否满足终止条件，若满足，则结束寻优；否则转至步骤 (5)。终止条件为寻优达到最大进化代数或最优个体的适应度相等。

(5) 依次执行遗传算子操作。对于选择操作，使用随机遍历抽样，设置代沟 $ggap$ ，根据适应度值从当前种群中选择 $m \times ggap$ 个个体复制到子代种群中。对于交叉操作，对选择操作中复制的子代种群进行单点交叉。对于交叉概率，本文采用动态参数法，这样做可以使进化后期优化的对象比较容易稳定，以减少无用计算。

(6) 基于适应度值的重插入。由于使用了代沟，子代种群的数量比当前种群数量要小，用子代群体中的个体代替当前种群中最不适应的个体。基于适应度的重插入确保当前种群中 $m \times (1 - ggap)$ 个最适应的个体总是被连续传播到下一代，并且每一代中不创建比现存种群多的个体，计算次数减少，内存要求也小。

(7) 进化代数加 1，检查是否满足终止条件，若满足，则结束寻优；否则转至步骤 (5)。

4 FCCU 反再子系统故障诊断实例

催化裂化反应再生子系统的故障类别多样，引发故障的原因复杂，故障的征兆参数不但数量多，而且它们之间的定量关系也难以用数学关系精确表达^[16-17]。通过研究催化裂化反应再生子系统整个运行过程的特点，结合相关的理论及经验知

识，确定了 4 个主要工艺故障，如表 1 所示，并选取了 12 个操作参数作为故障征兆参数，如表 2 所示。

表 1 催化裂化反应再生子系统主要故障

序号	故障类别
1	催化剂活性降低
2	二次燃烧
3	炭堆积
4	待生剂带油

表 2 催化裂化反应再生子系统主要故障征兆参数

序号	故障征兆参数
1	新鲜原料量
2	反应料位
3	反应温度
4	再生床层温度
5	烧焦罐出口温度
6	稀相温度
7	密相温度
8	反应压力
9	再生压力
10	两器差压
11	再生烟气氧含量
12	催化剂循环量

对于多故障的情况，本文采用基于决策树算法的多分类支持向量机，核函数选择 RBF 核函数。每类故障选取 100 个数据做为训练集，60 个数据做为测试集，对于参数 C 、 σ 的优化，采用改进的遗传算法进行比较，并与标准遗传算法比较，具体参数设置： C 、 σ 的范围均为 $[0, 100]$ ，二进制编码长度为 20，初始种群数量为 20，最大进化代数为 200，变异概率为 0.001。GA 的交叉概率为 0.4，IGA 的代沟为 0.9，交叉概率初始值为 0.4。利用 GA 和 IGA 进行参数优化后的 C 和 σ 的最优值以及分类准确率（以故障 1 为例）如图 4、图 5 所示。

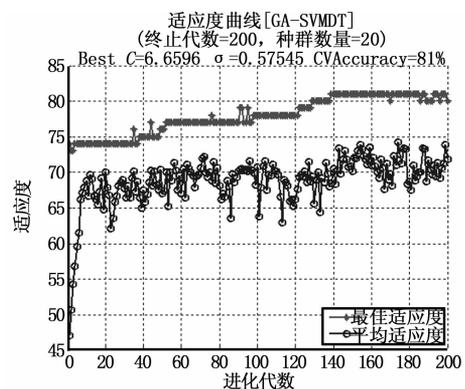


图 4 GA-SVMDT 适应度曲线图

由图 4、图 5 可以看出，对故障 1 进行故障识别时，采用 GA 算法优化 SVM 参数， C 的最优值为 6.659 6， σ 的最优值为 0.575 45，此时的分类准确率为 81%；采用 IAG 算法优化

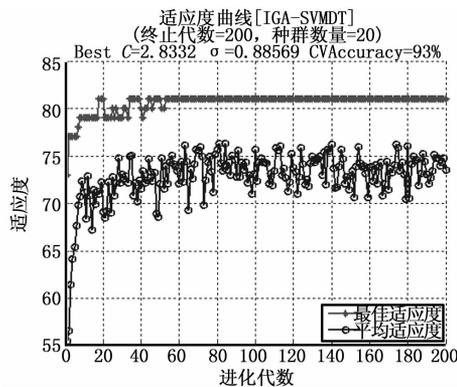


图5 IGA-SVMMDT 适应度曲线图

SVM 参数, C 的最优值为 2.833 2, σ 的最优值为 0.885 69, 此时的分类准确率为 93%。因此, IGA 相对于标准的遗传算法能够获得更优的参数, 分类准确率更高。由图 6 可知, 采用改进的遗传算法分类时所消耗的时间较标准遗传算法也缩短很多, 而对于在线故障检测, 能否快速地诊断出故障极其重要。分别采用 SVMMDT 算法、二叉树算法、一对一算法进行故障分类, 对比其分类准确率, 如图 7 所示。

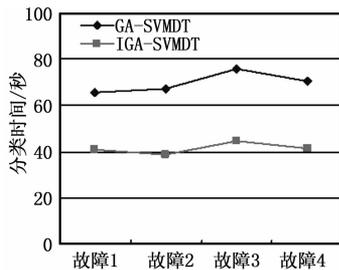


图6 GA-SVMMDT 和 IGA-SVMMDT 分类所用时间

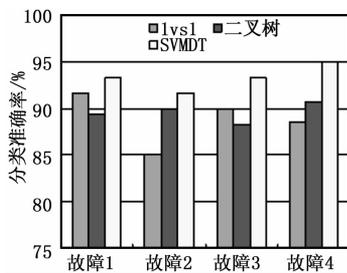


图7 各分类算法分类准确率比较

由图 7 可以看出, 由于 SVMMDT 训练是从最难分类的故障入手, 这在一定程度上避免了二叉树分类算法在根节点的错误分类而产生的误差积累, SVMMDT 的分类准确率高于其他的两种分类算法。

5 结语

本文采用了支持向量机决策树分类算法对催化裂化装置反应再生子系统进行故障诊断, 并利用改进的遗传算法对 SVM 的参数进行优化。实验结果证明, 建立的 IGA-SVMMDT 故障诊断模型效果良好, 缩短了样本数据的训练时间, 提高了故障数据的分类准确率。此外, 将支持向量机与主元分析 (PCA)

相结合, 可以构建一个基于数据驱动的在线故障监测与诊断系统。利用主元分析建立系统低维模型, 在线监测系统的实时工况数据并在故障发生时提取故障特征信息, 然后采用支持向量机进行故障类型的识别。

参考文献:

- [1] 中国石油化工集团公司人事部, 中国石油天然气集团公司人事服务中心. 催化裂化装置操作工 [M]. 北京: 中国石化出版社, 2009.
- [2] 鲁英. 催化裂化故障诊断系统研究 [D]. 西安: 西安石油大学, 2010.
- [3] 赵旭. 基于统计学方法的过程监控与质量控制研究 [D]. 上海: 上海交通大学, 2006.
- [4] 邓乃扬, 田英杰. 支持向量机—理论、算法与拓展 [M]. 北京: 科学出版社, 2009.
- [5] ilhan I. A genetic algorithm - support vector machine method with parameter optimization for selecting the tag SNPs [J]. Journal of Biomedical Informatics, 2013, 46: 328-340.
- [6] V. David Sánchez. Advanced support vector machines and kernel methods [J]. Neurocomputing, 2003, 55: 5-20.
- [7] Yuan S F, Chu F L. Support vector machines based fault diagnosis for turbo-pump rotor [J]. Mechanical Systems and Signal Processing, 2006, 20 (4): 939-952.
- [8] Lv G Y, Cheng H Z. Fault diagnosis of power transformer based on multi-layer SVM classifier [J]. Electric Power Systems Research, 2005, 75 (1): 9-15.
- [9] Azadeha A, Saberib M, Kazemc A. A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization [J]. Applied Soft Computing, 2013, 13: 1478-1485.
- [10] Takahashi F, Abe S. Decision-tree-based multiclass support vector machine [A]. Proceedings of the 9th International Conference on Neural Information Proceeding [C]. USA: Institute of Electrical and Electronics Engineers, 2002.
- [11] Arun Kumar M, Gopal M. A hybrid SVM based decision tree [J]. Pattern Recognition, 2010, 43: 3977-3987.
- [12] Shao Y H, Chen W J. The Best Separating Decision Tree Twin Support Vector Machine for Multi-Class Classification [J]. Procedia Computer Science, 2013, 17: 1032-1038.
- [13] Pournasheer E. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity [J]. European Journal of Medicinal Chemistry, 2009, 44: 5023-5028.
- [14] Huang C L. A GA-based feature selection and parameters optimization for support vector machines [J]. Expert Systems with Applications, 2006, 31: 231-240.
- [15] 孙凯, 段文超. 基于 IGA-SVM 的汽轮机故障诊断研究 [J]. 东北电力大学学报, 2012, 32 (3): 26-30.
- [16] 杜殿林. FCCU 反一再系统基于神经网络和 SDG 模型的混合故障诊断系统研究与开发 [D]. 北京: 北京化工大学, 2006.
- [17] 肖炎良, 侯立刚. 基于 PSO-SVM 的催化裂化装置故障诊断研究 [J]. 化工自动化及仪表, 2010, 37 (10): 55-57.