

基于混沌离散 PSO 算法的乙烯 生产过程模型变量的选择

王 凯, 韩雪峰

(南京工业大学 自动化与电气工程学院, 南京 211816)

摘要: 管式炉裂解技术是当今乙烯生产的主要技术, 乙烯生产过程中有多种影响质量指标的变量, 将全部变量参与建模显然会增加建模难度和复杂性, 并且大量的冗余信息会降低模型的精度; 针对这一问题, 提出了基于混沌思想的离散粒子群 (PSO) 算法进行模型变量的选择; 首先, 采用混沌离散 PSO 算法得到建模的最优输入变量集合, 再通过偏最小二乘法 (PLS) 对所选变量进行建模; 结果表明, 该方法可更有效地克服传统粒子群算法容易陷入局部最优的问题并建立较高的模型精度。

关键词: 乙烯生产; 管式炉裂解; 混沌离散粒子群优化算法; 变量选择; 部分最小二乘法

Selection of Modeling Variables Based on Chaotic and Discrete PSO in Process of Ethylene Production

Wang Kai, Han Xuefeng

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

Abstract: Tube cracking is the main technology in ethylene production and the quality index is influenced by many variables. Too many variables will obviously result in the increase of complication and difficulty of model structure. Besides, a lot of redundant information will reduce the accuracy of the model. To solve the problem, an algorithm based on chaotic and discrete PSO was proposed in this paper. Optimal input variable combination of ethylene production process is obtained by chaotic and discrete PSO algorithm. Then the model is set up based on the selected variables by PLS algorithm. The results show that the algorithm can make particles avoid falling into local optimum and get a higher accuracy model.

Keywords: ethylene production; tube cracking; chaotic and discrete PSO; variable selection; PLS

0 引言

在乙烯生产技术日趋成熟的今天, 进一步提高乙烯收率仍是乙烯生产商广泛关注的问题。对相同的裂解原料而言, 在相同工艺设计的装置中, 乙烯收率提高 1%。单位能耗、物耗大约能降低 3%。因此提高乙烯收率是降低乙烯装置能耗的最基本因素, 进而降低生产成本。为了提高乙烯收率, 优化乙烯生产操作条件是先决任务。想要得到系统的最优操作条件, 关键在于建立乙烯生产过程的数学模型, 最终提高乙烯收率, 降低生产成本。但是, 乙烯生产过程是一个非常复杂的化工过程, 有诸多影响乙烯收率的变量。如果将所有影响因素都作为建模变量, 将会引入大量冗余变量或者无关变量, 使得模型复杂化, 建模困难, 从而降低了模型预测精度。所以, 必须对建模变量进行筛选以提高模型预测精度。不少学者对建模变量的选择方法进行了研究^[1-3], 由于这些方法在变量剔除过程中用交叉验证法确定 PLS 主元个数的方法不能保证 PLS 回归模型系数的绝对值能反映相应自变量对因变量的贡献程度, 因此, 群智能优化算法^[4]的建模变量选择方法得到广泛关注。

粒子群优化算法 (Particle Swarm Optimization, PSO), 又称鸟群优化算法。其基本思想源于对鸟类觅食过程中迁徙和

聚集的模拟, 通过鸟之间的集体协作和竞争达到目的。PSO 算法是一种新颖、高效并行的智能优化算法, 寻优时收敛速度快、求解精度相对较高, 不需要目标函数的梯度信息, 算法易于编程实现, 易于与其他方法结合等优点。缺点是对于有多个局部极值点的函数容易陷入局部极值点。因此, 提出基于混沌思想的离散 PSO 建模变量选择方法^[5-6], 应用于乙烯工业生产过程的建模变量的选择, 并进行了仿真实验, 结果表明, 通过该方法得到的最优变量组合建立的模型, 可以有效地避免系统陷入局部最优, 减小了模型的泛化误差, 提高了模型的预测精度。

1 管式炉裂解制乙烯工艺

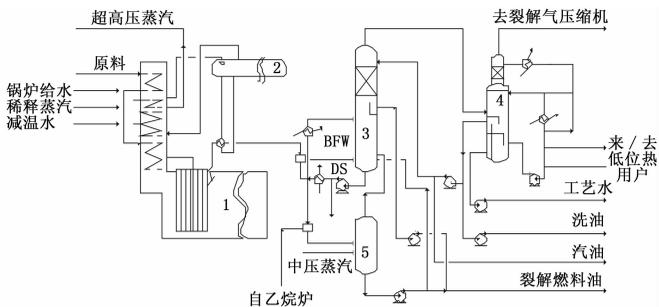
管式炉裂解制乙烯工艺包括裂解与急冷技术、压缩制冷技术和分离技术等。其中裂解与急冷技术是影响乙烯回收率的重要环节。乙烯生产过程的裂解与急冷工艺过程^[7]如图 1 所示。

从灌区来的液体原料经急冷水预热后进入裂解炉, 循环乙烷在深冷系统中被裂解气蒸发, 在冷箱中用丙烯制冷剂再加热, 然后与新鲜乙烷原料混合。原料在进入裂解炉前用急冷水加热, 再在对流段最上端盘管中预热, 加入稀释蒸汽, 并注入微量 DMDS。C₅ 和蒸汽在流量控制下, 加到总的裂解炉进料中去。烃与蒸汽的混合物再返回对流段, 在进入辐射段之前进一步预热。四组辐射段炉管出口在炉膛内两组相连后, 进到每台裂解炉的两台急冷锅炉 (TLE) 中。每台裂解炉的 TLE 均连接到一个公共汽包上的热虹吸热系统。产生 12.4 MPa 蒸汽, 进入每个汽包的锅炉给水, 用急冷油和对流段的烟道气预热。

收稿日期: 2014-01-09; 修回日期: 2014-03-17。

基金项目: 江苏省高校自然科学基金项目 (09KJB510003)。

作者简介: 王 凯 (1989-), 男, 江苏省徐州人, 硕士研究生, 主要从事工业过程建模优化方向的研究。



1. 裂解炉; 2. 高压蒸汽包; 3. 汽油分馏塔; 4. 急冷水塔; 5. 黏度控制塔
图 1 裂解和急冷部分示意图

蒸汽在 TLE 中产生，并在裂解炉对流段的盘管中加热到 520 ℃，过热器出口温度由锅炉给水注入量来控制。调节温度后，蒸汽返回对流段。从各 TLE 出来的裂解物料汇入一条总管中，经油急冷后送到汽油分馏塔，裂解气被进一步冷却，裂解燃料油从塔底抽出并送往汽提塔，汽油和较轻组分从塔顶蒸出。从裂解气中回收的热量，经过一个急冷油循环系统用于产生稀释蒸汽。汽油分馏塔采用在急冷塔中冷凝的汽油回流。配备一个黏度控制系统，以维持循环急冷油的黏度在一个可接受的范围内。从乙烷裂解炉出来的物料，在一个独立的油急冷器中与一部分循环的急冷油接触，汇合后物流送往黏度控制塔，塔顶馏分返回到汽油分馏塔，塔底抽出的重质燃料油，用急冷水冷却并送往界区外。从汽油分馏塔中部采出部分轻质燃料油组分进入轻质燃料油汽提塔，经蒸汽汽提控制闪点，塔顶气相返回汽油汽提塔，塔底轻质燃料油经冷却后送出界区与重质燃料油混合后送出界区。来自汽油分馏塔的塔顶气体与急冷塔中循环的水直接逆流接触，被冷却并部分冷凝。来自急冷塔的循环热水向一些工艺单元提供低位热能。急冷水经冷却水和预热裂解炉进料进一步冷却。急冷塔顶部气体温度为 40 ℃、压力约为 0.04 MPa (G)，送往裂解气压缩机。在急冷塔冷凝的汽油，从塔底的循环急冷水和冷凝的稀释蒸汽中分离出来。

2 基于混沌离散 PSO 建模变量的选择方法

乙烯生产过程的建模变量选择是一个离散组合优化问题，其旨在众多的变量中筛选出对因变量影响较大的变量作为输入变量建模，使模型的预测精度变高，而传统的 PSO 是一个连续的寻优过程。本文通过引入隶属度函数，将连续的优化过程离散化，同时采用混沌的思想，充分利用 PSO 的优点找到最优的建模变量组合。偏最小二乘法是种快速、有效的建模方法，能够克服多元回归中的多重共线性问题。结合彼此的优点，用 PLS^[8-9] 回归计算 PSO 的适应值，使变量的选择更加快速、精确。

2.1 标准 PSO 算法

粒子群优化算法是由 Eberhart 和 Kennedy^[10] 于 1995 年提出的一种基于迭代的优化方法，也是一类基于群智能 (Swarm Intelligence) 的随机优化算法。设粒子种群规模为 n ，每个粒子可以看做 $n \times D$ 维空间中的一个微粒。该微粒在搜索空间中以一定的速度飞行，飞行速度由个体和群体的飞行经验进行动态调整。设第 i 个粒子的当前位置和速度分别表示为 $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ 和 $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$ 。每个粒子位置对应一个由适应度函数求得的适应值，用以评价粒子位置的

优劣。粒子 i 所经历过的最好位置和种群经历过的最好位置分别表示为 $p_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ 和 $p_g = (p_{g1}, p_{g2}, \dots, p_{gD})$ 。则其速度和位置更新方程为：

$$v_{id}(t+1) = \omega \times v_{id}(t) + c_1 r_1 \times (p_{id}(t) - x_{id}(t)) + c_2 r_2 \times (p_{gd}(t) - x_{id}(t)) \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

其中： ω 是惯性权重， r_1 、 r_2 是 (0, 1) 上的随机数； c_1 、 c_2 是加速系数 (或称学习因子)。 ω 可以控制历史速度对当前速度的影响，较大的 ω 可以加强 PSO 算法的全局搜索能力，而较小的 ω 能加强局部搜索能力，但系统容易陷入局部最优。 c_1 、 c_2 分别用于调整粒子的自身经验和社会经验在其运动中所起的作用，表示将每个粒子推向 p_i 和 p_g 位置的统计加速项的权重。低的值允许粒子在被拉回前可以在目标区域外徘徊，而高的值则导致粒子突然冲向或越过目标区域，一般取 $c_1 = c_2 = 2$ 。

2.2 混沌离散 PSO 算法

基本 PSO 算法能够有效地处理连续优化的问题，而建模变量的选择是离散的优化组合问题。因此，一种离散形式的二进制优化^[11] (binary PSO, BPSO) 算法提出来，在 BPSO 中，一个二进制的空间就表示一个超立方空间，每个粒子用一个二进制变量来表示，可以通过该二进制变量的某些位在 {0, 1} 之间的翻转来实现粒子在这个超立方空间的移动。

设候选建模变量集合为 $m \times D$ 维空间。则每个粒子的位置 x_i 表示为一组长为 D 的 0、1 编码组合，其代表变量的一种选取组合。由于基本 PSO 算法中粒子初始化有一定的盲目性，影响过程的收敛。这里采用混沌序列初始化粒子的位置和速度，由于混沌带有精确内在规律的随机性以及遍历性。混沌搜索会比盲目无序地随机搜索更具有优越性，可以避免算法陷入局部最优。随机产生一个 m 维每个分量数值在 0~1 之间的向量，引入一个典型的混沌系统 Logistic 方程，得到 M 个向量，然后将各个向量的各个分量载波到对应变量的取值区间，将混沌引入优化变量使其呈现混沌状态。可以有效地改变种群的多样性并帮助惰性粒子逃离局部最小点。Logistic 方程如下：

$$s_{k+1} = \mu s_k (1 - s_k) \quad (k = 0, 1, 2, \dots) \quad (3)$$

式中， μ 为控制参数，取 $\mu = 4$ ， $s_0 \in [0, 1]$ 。

采用式 (1) 与式 (2) 对粒子的位置和速度进行更新。但粒子每次更新后的位置不再是 0、1 二进制数。因此，为了维持粒子的二进制形式，这里引入 Sigmoid 函数，使粒子位置更新后的 x_i 转化成 0 或 1。Sigmoid 函数如下：

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

从函数式 (4) 可以看出 x 值越大，粒子的位置 x_i 选 1 的概率就越大； x 值越小， x_i 选 0 的概率就越大。其函数曲线如图 2 所示。经分析，可以将粒子位置的 Sigmoid 函数作为判断该粒子元素为 0 或 1 的隶属度函数。在传统离散 PSO 中，将粒子的飞行速度 $v_{id}(t+1)$ 作为 Sigmoid 函数的自变量，来判断粒子位置 $x_{id}(t+1)$ 的 0、1 取值。因这种间接方式对 $x_{id}(t+1)$ 的隶属度计算不利于提高收敛速度，所以本文采取对公式 (2) 更新后的 $x_{id}(t+1)$ 进行隶属度计算。即当 $\text{Sigmoid}(x_{id}) \leq S$ 时， $x_{id} = 0$ ，否则 $x_{id} = 1$ 。考虑到粒子初始

速度 $v_{iD}(0)$ 取 $(0, 1)$ 之间的随机数, 且 $x_{iD}(t)$ 为 0、1 二数值, 则迭代后 $x_{iD}(t+1) > 0$ (即阈值 $S > 0.5$) 的概率大, 所以阈值 S 的选取也需要大些。本文取方程 $x = 1 / (1 + \exp(-x))$ 的最优解作为阈值 S ($S = 0.659 0$)。

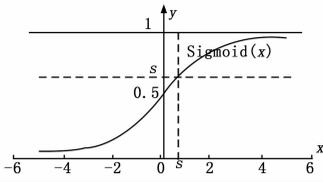


图 2 Sigmoid 函数曲线

2.3 适应度函数

在 PSO 算法中, 适应度函数的适应值是粒子最优位置更新的依据, 根据所选取的不同适应度函数, 所对应的 PSO 算法的准确性和收敛性也有所不同。因此, 适应度函数的选取至关重要。部分最小二乘法 (Partial Least Square, PLS) 能够消除原始数据间的共线性和噪声的影响, 并兼具有快速性和有效性等优点。将其应用于适应值的计算, 能够有效提高收敛的速度和方向性。

依据 PLS 回归算法, 本文采用文献 [12] 提出的适应度计算公式 (5) 来评价各个粒子的适应值。

$$R^2_{pred} = \max_l \left[1 - \frac{\sum_i (y_{i,yu} - y_{-i,pred})^2}{\sum_i (y_{i,yu} - y_{jun})^2} \right] \quad (5)$$

式中, $y_{i,yu}$ 为样本 i 的实际测量值, y_{jun} 为 $y_{i,yu}$ 的平均值, $y_{-i,pred}$ 为用除去第 i 个样本的数据建立的模型对 $y_{i,yu}$ 的预测值, l 为式 (4) 取得最大值时的主元个数。 R^2_{pred} 的值越接近 1 则模型的预测能力越好。

3 乙烯生产过程仿真实验

将上文提出的算法结合偏最小二乘法用于乙烯生产过程的建模中, 乙烯生产过程具有设备种类繁多和危险度高等特点, 是一个极其复杂的工艺过程, 影响乙烯生产过程的变量很多。如果将所采集到的全部变量用于建模, 必然产生信息冗余, 而将少数变量参与建模, 则又会漏掉部分重要变量, 降低模型的预测精度。因此, 必须对变量进行选择以提高模型的预测精度。根据对乙烯生产工艺的分析, 从现场采集到了 17 个候选变量, 包括裂解温度、炉管温度、急冷锅炉出口温度、炉膛温度、炉壁温度、横跨温度、排烟温度、裂解炉出口压力 (COP)、真实停留时间、表观停留时间、积分停留时间、投料量、稀释比、空气过剩系数、炉管内流体压降、燃烧器底部供热与侧壁供热比及质量流速。样本为 2013 年 8 月~11 月现场数据剔除空值后的 121 组数据。

3.1 算法步骤及仿真实验

3.1.1 算法步骤

将采集到的 17 个候选变量用本文提出的混沌离散 PSO 算法进行变量选择, 具体步骤如下:

Step1: 初始化粒子种群和相关参数。惯性权重 ω 、学习因子 c_1 、 c_2 。

Step2: 随机产生一个 m 维每个分量数值在 $0 \sim 1$ 之间的向

量, $Z_1 = (z_{11}, z_{12}, \dots, z_{1m})$, 利用典型的混沌系统 Logistic 映射, 得到 N 个向量 Z_1, Z_2, \dots, Z_M 。将 Z_j 的各个分量截波到对应变量的取值区间。计算目标函数值, 选择较好的 n 个粒子作为初始种群, 并随机初始 n 个粒子的速度。

Step3: 根据 PLS 回归算法和式 (5) 计算各个粒子的适应度函数, 并求其适应值。

Step4: 如果当前粒子位置的适应值优于其经历过的最好位置的适应值, 则将该粒子当前位置替换为其最好位置。

Step5: 如果当前种群位置的适应值优于种群所经历过的最好位置的适应值, 则将种群的当前位置替换为种群最好位置。

Step6: 根据式 (1) 式 (2) 对当前粒子的位置和速度进行调整, 并根据式 (4) 对更新后的粒子进行离散化。

Step7: 若满足寻优终止条件, 则终止寻优, 否则返回 Step3。

3.1.2 仿真实验

用本文提出的方法将采集到的 17 个候选变量进行变量选择。实验步骤如下:

1) 将 17 个候选变量做归一化处理, 将离散 PSO 的参数初始化, 种群数 $n = 22$ 、惯性权重 $\omega = 0.8$ 、学习因子 $c_1 = c_2 = 0.7$ 。最大限速 $V_{max} = 4$ 。

2) 根据文中的步骤, 进行迭代优化, 迭代过程中适应值的变化如图 3 所示。经优化后, 得到裂解温度、横跨温度、稀释比、积分停留时间、裂解炉出口压力及投料量 6 个变量组合, 其序列如表 1 所示。

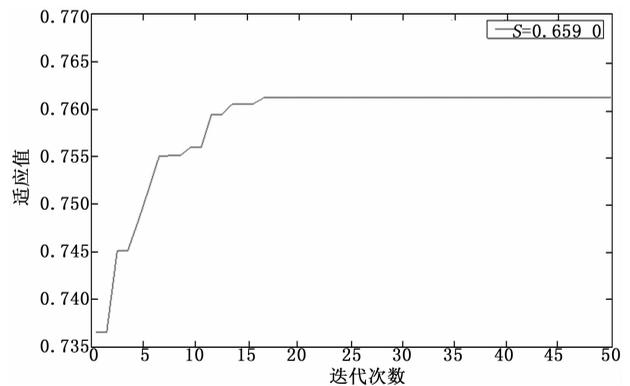


图 3 迭代过程中 PSO 的适应值变化曲线

表 1 预测泛化误差比较

变量集	所选变量序列	RMSE	MAXE
原变量集	1~17	0.189 2	0.987 1
少量关键变量	1;7;12;15	0.173 4	0.921 3
离散 PSO	1;3;7;8;12;15	0.132 2	0.817 5

3) 用得到的 6 个变量用 PLS 进行建模预测, 其中前 90 组训练模型, 后 31 组模型预测。预测曲线如图 4 所示。

4) 将运用混沌离散 PSO 算法选取的 6 个变量组合与原来全部变量组合和部分关键变量组合进行的 PLS 回归预测作比较, 结果在表 1 中列出。

对于模型预测精度的评价, 本文采用泛化均方根误差^[13]和最大泛化误差。具体如下:

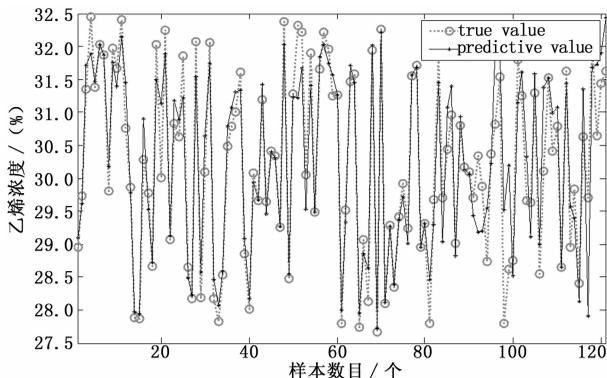


图 4 基于混沌离散 PSO 选取变量 PLS 回归预测

$$\begin{cases} RMSE = \frac{\|Y - Y_t\|_2}{\sqrt{M}} \\ MAXE = \max_M |Y - Y_t| \end{cases}$$

式中, $RMSE$ 为泛化均方根误差, 而 $MAXE$ 为最大泛化误差, M 为测试样本数, Y 为测试样本的实际输出, Y_t 为测试样本的预测输出。

3.2 实验结果分析

采用本文提出的混沌离散 PSO 算法, 经过多次迭代试验, 均可以得到全局最优解, 表明该方法有效, 由表 1 可以看出, 将所有原变量组合, 少量关键变量组合和本文所选的变量组合进行用于 PLS 建模, 本文所选的变量组合的泛化均方根误差和最大泛化误差比其它两种变量组合都小, 说明在乙烯生产过程中变量的选择或多或少都会影响模型的预测精度。而采用本文提出的方法进行变量选择建模减小了模型的泛化误差, 提高了模型预测精度。

4 结论

在乙烯生产过程中有诸多影响乙烯收率的变量, 而将选取少量变量或者所有变量进行建模都会影响模型的预测精度。本文在传统 PSO 算法基础上, 采用混沌的思想, 用计算粒子位置的隶属度将其离散化的方法, 来实现离散 PSO 算法在建模变量选择中的应用, 最后将该方法应用于乙烯生产过程模型变量选择的仿真实验中, 实验结果表明, 该方法可有效避免传统

PSO 算法容易陷入局部最优的问题, 并得到较好的模型预测精度。

参考文献:

[1] Xu L, Zhang W J. Comparison of different methods for variable selection [J]. *Analytica Chimica Acta*, 2001, 446: 477-483.

[2] Szekeley N. Simplifying the model of a complex industrial process using input variable selection [J]. *Periodica Polytechnica Electrical Engineering*, 2003, 47 (1/2): 141-147.

[3] Gauchi J P, Chagnon P. Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, 58: 171-193.

[4] 雷秀娟. 群智能优化算法及其应用 [M]. 北京: 科学出版社, 2012.

[5] Reynolds C W. Flocks, herds, schools: a distributed behavioral model [J]. *Computer Graphics*, 1987; 21 (4): 25-34.

[6] Liu B, Wang L, Jin Y H. Improved particle swarm optimization combined with chaos [J]. *Chaos, Solitons and Fractals*, 2005, 25 (5): 1261-1271.

[7] 王松汉. 乙烯工艺与技术 [M]. 北京: 中国石化出版社, 2012.

[8] Geladi P, Kowalski B R. Partial least-squares regression: a tutorial [J]. *Analytica Chimica Acta*, 1986, 185 (1): 1-17.

[9] 李丽娟. 最小二乘支持向量机建模及预测控制算法研究 [D]. 浙江大学, 2008.

[10] Kennedy J, Eberhart R. Particle swarm optimization [A]. *Proceedings of IEEE International Conference on Neural Networks* [C]. Perth Australia, 1995: 1942-1948.

[11] Kennedy J, Eberhart R. C. A discrete binary version of the particle swarm algorithm [A]. *Proceedings of the World Multi-conference on Systemic, Cybernetics and Informatics*. Piscataway [C]. NJ: IEEE Service Center, 1997: 4104-4109.

[12] Hasegawa K, Miyashita Y, Funatsu K. GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists [J]. *Journal of Chemical Information and Computer Sciences* (S0095-2338), 1997, 37 (2): 306-310.

[13] Li L J, Su H Y, Chu J. Modeling of isomerization of C8 aromatics by online least squares support vector machine [J]. *Chinese Journal of Chemical Engineering* (S1004-9541), 2009, 17 (3): 437-444.

(上接第 2624 页)

[4] Dasgupta D, Forrest S. Artificial immune systems in industrial applications [C]. *Intelligent Processing and Manufacturing of Materials*, 1999.

[5] Gonzalez, Dasgupta D. Anomaly detection using real-valued negative selection [C]. *Genetic Programming and Evolvable Machine*, 2003.

[6] Zhou J, Dasgupta D. V-detector: An efficient negative selection algorithm with "probably adequate" detector coverage [J]. *Information Sciences*, 2009, 179: 1390-1406.

[7] Hu Z B, Zhou J, Ma P. A novel anomaly detection algorithm based on real-valued negative selection system [J]. *Knowledge Discovery and Data Mining*, 2008.

[8] 刘树林, 王日新, 黄文虎, 等. 基于反面选择算法的压缩机振动在线检测方法研究 [J]. *压缩机技术*, 2001, 5: 9-10.

[9] 窦唯, 刘树林, 孙明, 等. 生物免疫机理在往复压缩机在线状态检测中的应用 [J]. *流体机械*, 2004, 32 (5): 16-19.

[10] 刘树林, 丛蕊, 冷建成, 等. 孟庆武小波包特征免疫检测器在设备异常状态检测中的应用 [J]. *大庆石油学报*, 2005, 29 (6): 101-103.

[11] Zhou J, Dasgupta D. Augmented negative selection algorithm with variable-coverage detectors [A]. *Congress on Evolutionary Computation, CEC2004* [C]. Portland, Oregon, 2004: 123-136.

[12] 程军圣, 于德介等人. 时间-小波能量谱在滚动轴承故障诊断中的应用 [J]. *振动与冲击*, 2004, 23 (2): 34-36.