

基于IA-SVM模型的热采井套管寿命预测

骆正山, 毕傲睿

(西安建筑科技大学管理学院, 陕西 西安 710055)

摘要:为解决稠油开采过程中热采井套管因高温高压、泥沙磨损等原因导致变形、破损以及泄漏的问题,提出一种基于免疫算法(IA)优化支持向量机(SVM)的套管寿命预测模型。首先,应用免疫算法得到影响支持向量机预测精度的三个参数的精确值;其次,将5个影响套管寿命的因素作为支持向量机的输入变量,套管寿命作为目标输出,建立寿命预测模型;最后以辽河油田运行的套管为例,验证模型预测热采井套管寿命的效果。结果表明,该模型能够很好地预测热采井套管的寿命,误差小、吻合度高,为套管的维护和维修提供了参考依据。

关键词:稠油开采;热采井套管;寿命预测;免疫算法(IA);支持向量机(SVM)

中图分类号:TE973

文献标识码:A

Prediction on casing life of thermal production wells based on IA-SVM model

Luo Zhengshan and Bi Aorui

(School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China)

Abstract: In order to solve the problems of the distortion, damage and leakage of the casing pipes caused by the high temperature, high pressure, and the sand erosion of the thermal production wells during the exploitation of the heavy oil, a life prediction model based on the immune algorithm (IA) and the support vector machine (SVM) was built. Firstly, the IA was used to obtain the exact values of three parameters that affected the prediction accuracy of the SVM. Secondly, the life prediction model was built by taking 5 factors that affected the life of the casing as the input variables of the SVM, and the casing life as the target output. Finally, the casing wells running in Liaohe oilfield were taken as the examples to test the performance of the model. The results showed that the model could predict the casing life well with small errors and high coincidence. It provides a reference for the maintenance and repair of the casing.

Key words: heavy oil recovery, casing pipe of thermal production well, life prediction, immune algorithm, support vector machine

随着经济的快速发展,全球常规石油储量已无法满足生产需求,各国正从常规原油转向稠油、特稠油的开采^[1],然而由于稠油黏度高、流动阻力大,一般需要使用蒸汽吞吐、蒸汽驱等技术^[2-5]。这些技术在提高开采率的同时,对套管造成了严重的损坏,使套管强度变低、变形、错断,进而导致失效,发生泄漏事故,不但影响油田的日常工作,严重时更会造成生命和财产损失。因此,对稠油热采井套管寿命进行科学的预测,制定相应的维修策略,对保证管道的安全

运行和高效生产有着重要的现实意义。

目前国内对管道寿命预测方法的研究已取得一定的成果。王早祥等^[6]应用神经网络模型,确定套管射孔段强度系数与影响因素的关系,建立强度预测模型;李琴等^[7]提出利用遗传算法优化的BP神经网络模型并对腐蚀管道的压力进行了预测;王如君等^[8]结合灰色模型和马尔科夫链,利用少量数据对管道的最大腐蚀深度和腐蚀状况进行了预测等。本文考虑热采井套管运行环境的复杂性以及开采技术的特殊性,从

收稿日期:2017-03-05。

第一作者简介:骆正山(1969—),男,博士,教授,博士生导师,从事管理科学与工程、信息管理与信息系统、油气管道风险评估等学科领域的教学与科研。

基金项目:国家自然科学基金项目“陆上油气管线风险评估技术研究”(61271278);陕西省教育厅项目“盐穴型地下储气库完整性评价技术的研究”(16JK1465);陕西省重点学科建设专项资金“管理科学与工程重点学科”(E08001)。

多维角度出发,将套损因素作为支持向量机输入,并针对支持向量机参数值的不准确性,利用免疫算法进行优化,最终实现套管寿命的精确预测,为稠油开采的安全运行和套管维修提供一定的理论支持。

1 基本原理

1.1 支持向量机

支持向量机(SVM, support vector machine)是基于统计学习理论的一种智能预测方法,在数据挖掘、回归分析、非线性系统建模与识别、函数逼近等领域应用广泛^[9-10]。该方法通过结构风险最小化原理^[11]来提高泛化能力,能比较好地解决小样本、非线性、高维数和局部极小点等问题。其中对于非线性回归问题,SVM需要选择核函数 $K(x_i, x_j)$ 简化非线性逼近,使函数逼近求解绕过特征空间,直接在输入空间上求解,从而将非线性问题转换为可线性拟合的优化问题输入,避免高维空间复杂的点积运算。

SVM预测回归函数为:

$$f(x) = \sum_{x_i \in SVs} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (1)$$

式中: α_i, α_i^* 为拉格朗日乘子, b 为偏置, SVs 为支持向量集; $K(x_i, x_j)$ 为核函数,用于代替高维空间的内积运算,减少计算量,常用的有多项式核函数、高斯径向基核函数等。其中高斯径向基核函数适用性较强,通过选择参数可以适用于任意分布的样本:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

式中: x_i 为输入样本值, x_j 为核函数的中心, σ 为核参数。

拉格朗日乘子和偏置由下式确定:

$$\min_{\alpha_i, \alpha_i^*} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) + \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \quad (3)$$

约束条件为: $\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i^* \leq C, i = 1, 2, \dots, n; C$ 为惩罚参数, ε 为不敏感损失参数。

从式(1)至(3)可以看出参数 C, σ 和 ε 影响着支持向量机的预测精度和泛化能力。惩罚因子 C 值越高说明越不能容忍误差,但是容易过拟合,值越小则

容易欠拟合,过大或过小都会导致模型泛化能力变差; σ 反映训练样本数据的分布特性;不敏感损失参数 ε 是一个大于零的值,是忽略真实值在某个范围内的误差,决定着支持向量的个数。目前对于如何得到这三个参数的具体值还没有特别有效的方法,一般是通过先验知识、网格交叉验证试算等方式来确定,但这些方法主观、不精确、易陷入局部最优,并且在样本较大时运算量大、耗时长,严重影响支持向量机的实际使用。

1.2 免疫算法

免疫算法(IA, immune algorithm)是一种受生物学免疫系统启发,将免疫机理结合实际应用而形成的一种人工智能模拟算法,有快速的全局搜索、动态维护自我记忆的能力,以及自我、非我的抗原区别机制^[12],对于外界干扰能够很强烈地维持整体平衡,具有高稳定性和可靠性。目前已应用于参数优化、机器学习、数据处理、异常和故障诊断、工业设计等领域^[13],并取得了良好的效果。因此,本文将用其解决支持向量机的参数取值问题。

2 寿命预测模型的建立

本文通过免疫算法对支持向量机参数进行选值,以期提高支持向量机的泛化、学习能力和预测精度,然后应用具有最优参数的支持向量机进行热采井套管的寿命预测,模型建立如下(图1):

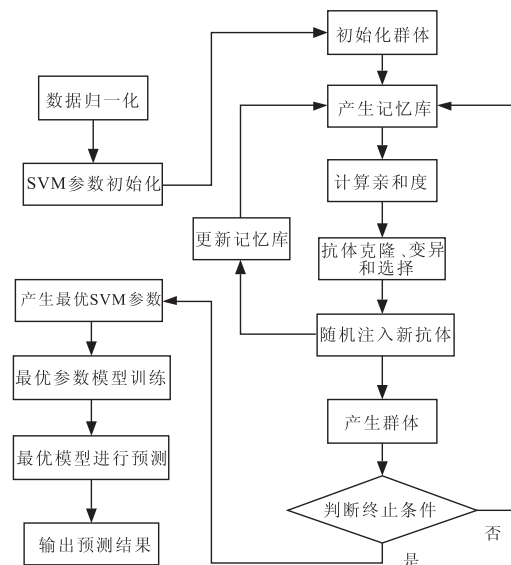


图1 IA-SVM模型预测过程

Fig. 1 Prediction process of model IA-SVM

1) 数据准备和预处理。将套管破损因素作为支持向量机的输入变量集 $X(x_1, x_2, x_3, \dots, x_n)$, 其中 $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ij})$ 为破损因素, x_{ij} 为第 i 个破损因素下的第 j 个样本值, $i=1, 2, 3, \dots, n, j=1, 2, 3, \dots, m$; 对应的套管寿命作为输出集 $Y(y_1, y_2, y_3, \dots, y_n)$; 将样本数据分为训练样本和测试样本, 为便于计算分析且消除量纲的影响, 将数据标准化处理, 如下:

$$x_{ij}^* = (x_{ij} - \bar{x}_i) / \sqrt{\frac{1}{n-1} \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2} \quad (4)$$

式中: \bar{x}_i 为破损因素 i 的样本平均值, $i=1, 2, 3, \dots, n, j=1, 2, 3, \dots, m$ 。

2) 核函数选择。选择高斯径向基核函数, 见式(2)。

3) 初始化。将寿命实际值 y_i 设置为抗原, 令参数向量 $v_i = (C_i, \varepsilon_i, \sigma_i)$, 设置参数取值范围分别为 $[0.1, 5]$ 、 $[0, 1]$ 和 $[1, 100]$; 设参数向量规模为 N 、记忆库规模为 M , 在取值范围内随机生成初始群体 $A(0)$ 和初始记忆库 $R(0)$ 并进行二进制编码; 令迭代次数 t , 最大迭代次数 T , 初始值 $t=0$, 迭代到第 t 次时, 抗体群体为 $A(t)$ 。

4) 亲和力计算。亲和力分两种: 抗原和抗体的亲和力以及抗体之间的亲和力。因为模型利用破损因素作为支持向量机的输入变量, 结合初始群体 $A(0)$ 得到初始寿命预测值; 将初始寿命预测值作为抗体, 结合抗原计算亲和力 $F(0)$, 计算公式为:

$$F(0) = 1 / \sqrt{\frac{1}{s} \sum_{i=1}^s (y_i - \hat{y}_i)^2} \quad (5)$$

式中: y_i 为套管实际寿命, a ; \hat{y}_i 为在抗体 v_i 下的预测值, a ; s 为训练样本数。

单个抗体 \hat{y}_i 和抗原 y_i 的亲和力描述为:

$$F_i(0) = 1 / \|y_i - \hat{y}_i\| \quad (6)$$

式中: $\|y_i - \hat{y}_i\|$ 为抗体和抗原的欧拉距离。

5) 克隆、变异和更新。计算 $A(0)$ 的浓度, 抗体的浓度体现了相似抗体的规模, 利用浓度控制抗体的促进和抑制, 防止参数过早收敛。浓度公式为:

$$D(0) = \frac{1}{N} \sum_{i=1}^N H_{i,j}(0) \quad (7)$$

$$H_{i,j}(0) = \begin{cases} 1, & \|v_i - v_j\| > \tau \\ 0, & \text{others} \end{cases} \quad (8)$$

式中: τ 是阈值, $\|v_i - v_j\|$ 为抗体 i 和抗体 j 的欧拉距离。

抗体选择的原则是高亲和度的抗体得到促进, 而高浓度的受到抑制。由浓度 $D(0)$ 和亲和力 $F(0)$ 得到抗体的克隆数量公式为:

$$k = \text{round}(\lambda \frac{F(0)}{D(0)}) \quad (9)$$

式中: $\text{round}(\cdot)$ 为取整函数; λ 为伸缩因子, 用于防止克隆数量过大或过小, 因为亲和力一般大于 1, 浓度的期望值取中间量 0.5, 设定克隆数目的期望数目大约为 6, 故取 $\lambda=3$ 。由式(6)计算单个抗体的亲和力, 由高至低克隆 k 个得到新群体 $C(0)$ 。

对 $C(0)$ 进行变异, 一般抗体变异的概率与亲和力 $F(0)$ 成单调递减关系, 根据参数寻优的特点, 本文提出变异概率:

$$e(0) \in [0, f(0)] \quad (10)$$

$$f(0) = 1 - \exp(-F(0)) \quad (11)$$

式中: $e(0)$ 为 $[0, f(0)]$ 上的随机数。变异算子为: 针对 v_i 二进制编码的每一位, 在 $[0, 1]$ 中生成随机数并与 $e(0)$ 比较, 随机数小于 $e(0)$ 的则变异。经过变异后得群体 $V(0)$ 。

用 $V(0)$ 中亲和力高的抗体替换 $R(0)$ 中亲和力低的抗体, 之后将 $V(0)$ 并入 $A(0)$, 去掉重复和亲和力低的抗体, 然后随机产生部分新值替换 $A(0)$ 中亲和力低的抗体, 最后去掉亲和力大于阈值的抗体, 保持种群规模 N ; 转步骤 2), 令 $t=1$ 。

6) 终止条件。若 $t=T$, 则迭代终止, 比较 $A(t)$ 和 $R(t)$ 中的最优个体作为输出结果, 得到最优参数解 v^* 并赋给支持向量机。

7) 寿命预测。将标准化的数据 x_{ij}^* 和最优参数 v^* 代入式(1)进行训练, 求解拉格朗日参数 α_i 、 α_i^* 和偏置 b , 用测试样本完成对热采井套管的寿命预测。

3 实例分析

目前辽河油田稠油投产井使用套管多为 N80 型, 套损数量以每年 40 口左右递增, 套损率约为 68%, 平均使用寿命 8~10 年, 套损类型主要有变形、错断、漏失、缩径、破裂等。试验选取 21 组套管数据作样本集, 其中前 16 组数据作为模型训练数据, 后 5 组数据作为模型测试数据, 基本情况见表 1。

3.1 套损影响因素选择

由国内外油田几十年的开采经验可知, 目前稠

表1 样本基本信息
Table 1 Basic information of samples

序号	损坏类型	实际寿命/a	序号	损坏类型	实际寿命/a	序号	损坏类型	实际寿命/a
1	破裂	9.98	8	错断	8.34	15	错断	6.77
2	变形	8.00	9	缩径	7.84	16	缩径	7.66
3	错断	9.18	10	变形	6.81	17	变形	5.39
4	变形	10.71	11	变形	7.90	18	错断	4.57
5	漏失	7.74	12	变形	11.45	19	错断	8.33
6	缩径	4.56	13	破裂	8.24	20	缩径	6.62
7	破裂	9.30	14	错断	6.03	21	变形	6.96

油套损因素主要分为地质因素、工程因素和腐蚀因素三大类,每大类又包括一些具体原因(表2)。

虽然稠油套损因素众多,但是导致决定性破坏作用的因素却是有限的,通过对相关统计资料分析得到辽河油田套损主要原因如下:套损比率随出砂次数的增大而增大,见图2,而油井每次出砂量大致相同,故可将出砂总量作为套损的间接因素;套损部位有90%以上出现在射孔段和射孔段顶部附近区域,见图3。

注汽温度和注汽压力影响套管应力,导致变形等,见图4、图5,其中图4中的注汽压力为16 MPa,图5中的注汽温度为350 ℃;同时,套管和水泥环由于膨胀系数不一致产生的剪切应力大于井底胶结剪切力时会导致套管和水泥环脱开,脱开部位由于出现受力空间导致套损也非常严重。

综上分析,将出砂量、射孔密度、注汽温度、注汽压力、井底胶结剪切力五个因素作为SVM的输入变量。

表2 套损因素
Table 2 Casing failure factors

套损因素	具体因素
地质因素	泥岩膨胀、盐层蠕动、断层运移、油井出砂
工程技术因素	套管强度及井身结构不合理、固井质量差、射孔损坏、注水不合理、酸化压裂、热采温度过高
腐蚀因素	硫化物、CO ₂ 、H ₂ S、各种腐蚀性物质等

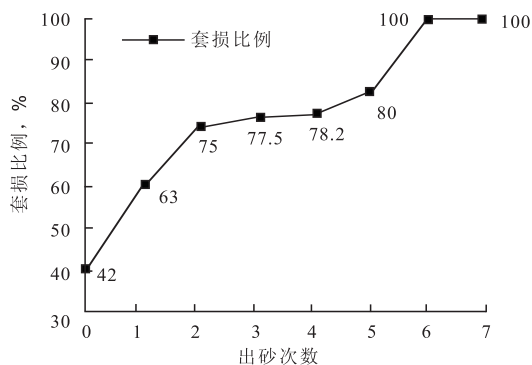


图2 套损比率与出砂次数关系

Fig. 2 Relation between casing failure ratio and sand amount

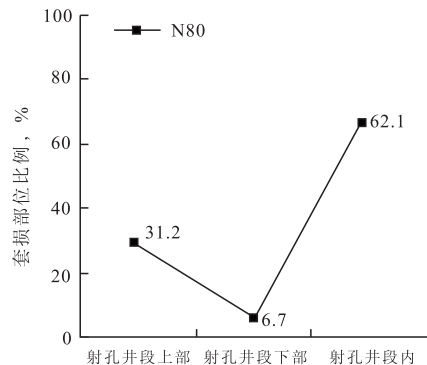


图3 套管损坏部位比例

Fig. 3 Proportion of casing failure parts

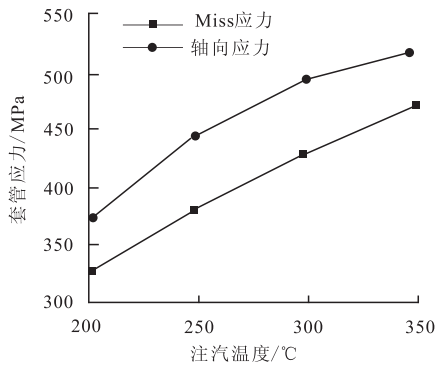


图4 应力对注汽温度变化曲线

Fig. 4 Curves of injected steam temperature and stress

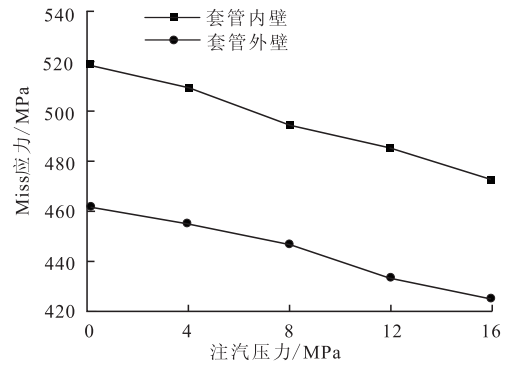


图5 应力对注汽压力变化曲线

Fig. 5 Curves of injected steam pressure and stress

3.2 参数优化和预测

初始化模型:抗体个数 $N=45$,记忆库个数 $M=23$,抑制阈值 $\tau=0.1$,迭代次数最大 $T=300$ 。抗体与抗原亲和度随迭代变化曲线见图6,亲和度随迭代次数增加而趋于稳定;当模型陷入局部最优时,通过克隆、变异和选择,会在局部极值附近进行小范围的扰动变化,然后跳出该局部极值点重新进行搜索。

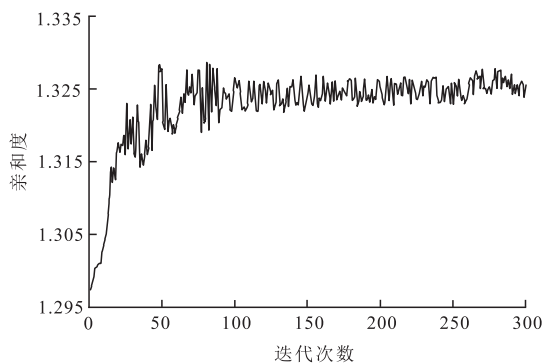


图6 亲和度变化过程

Fig. 6 Process of affinity changing

表3 不同模型预测值和实际值

Table 3 Predicted and actual values of different models

序号	Bp神经网络/a	GA-SVM/a	SVM/a	IA-SVM/a	实际值/a
1	6.15	5.82	6.10	5.53	5.39
2	5.14	4.90	5.10	4.68	4.57
3	9.64	9.13	9.57	8.90	8.33
4	7.98	7.22	7.86	6.98	6.62
5	7.90	7.54	7.90	7.32	6.96

表4 预测模型误差对比

Table 4 Errors comparison of different prediction models

序号	模型	标准差比值	平均相对误差,%	平均绝对误差/a
1	Bp神经网络	1.205	15.27	1.571
2	GA-SVM	1.143	5.27	1.359
3	SVM	1.197	14.38	1.551
4	IA-SVM	1.127	4.49	1.323

通过免疫算法的训练总次数为13 500次,获得的 σ 、 ε 、 C 值分别为:0.248 4、0.009 6、5.433 3。将这3个值代入的SVM模型并用Matlab的libsvm工具包编程进行训练,训练完成后对检验集数据进行预测,预测结果和一些常用模型预测结果见表3。

3.3 试验结果分析:

表4给出了各个模型的数据误差指标,分析对比如下:

1) 对于未经优化的SVM参数 σ 、 ε 、 C 采用网格交叉法筛选,将各参数取值区间分别划分为50、20和100,则共需 10^5 次训练,参数值分别为:0.2、0.05、5,其预测结果平均绝对误差为14.38%;Bp神经网络采用3层网络,其输入层结点数为6,隐含层结点数为4,迭代次数上限2 000次,其预测结果平均绝对误差为15.27%。可以看出采用网格交叉法筛选参数值训练次数比IA-SVM高了一个数量级,效率太低;而Bp神经网络和未经优化SVM的预测结果比IA-SVM平均误差高了近10%,对于提供维修策略的借鉴可

用性较低。

2) 对于GA-SVM模型,训练次数与IA-SVM相同,参数 σ 、 ε 、 C 经过遗传算法优化为:0.263 6,0.012 4,5.855 6,其预测结果平均相对误差比IA-SVM模型略高0.78%;对比IA-SVM和GA-SVM可以发现二者的预测效果较近似,原因是IA和GA均属于随机启发式搜索的进化算法,在形式上类似,但是IA在GA的基础上考虑了外部环境和个体之间的相互作用,因此,具有更加完善的搜索能力,结果也更加精确。

通过分析可知,对于标准差比值、平均相对误差、平均绝对误差这几项数据误差指标,IA-SVM模型的性能均优于其他三种模型;同时,虽然几种预测模型的输出值与实际值相比都非常接近,但是IA-SVM的训练时间更少,预测结果更加符合实际情况,对维修决策也更具有参考价值。

4 结论

1) 利用免疫算法得到影响支持向量机预测能力的3个参数,其效率和精度高于常规使用的网格交叉法和遗传算法,同时避免陷入局部极值的问题,提高了支持向量机的预测精度。

2) 通过对比分析IA-SVM模型、Bp神经网络模型、未优化的支持向量机和GA-SVM模型的预测结果可知,IA-SVM模型的预测结果与实际情况更加吻合,误差指标更小。

3) IA-SVM模型能够对稠油热采井套管的最大寿命进行预测,为及时准确掌握套管使用情况提供依据,也为热采井套管的进一步维修、更换提供了参考。

参考文献

- [1] 张万才,马振生,郭立君,等.热采井套管损坏机理及防治技术——以单家寺油田为例[J].油气地质与采收率,2005,12(2):74-76.
- [2] 万仁溥.现代完井工程[M].北京:石油工业出版社,2000:49-50.
- [3] 李敬元,李子丰,马兴瑞,等.热采井注汽管柱力学分析[J].工程力学,1998,15(3):51-60.
- [4] 张桂林.可复位式预应力固井地锚的研究与应用[J].石油钻探技术,2000,28(3):36-38.
- [5] 刘坤芳,张兆银,孙晓明,等.注蒸汽井套管热应力分析及管柱强度设计[J].石油钻探技术,1994,22(4):36-40.
- [6] 王旱祥,隋允康,李增亮.神经网络在射孔对套管强度影响预测中的应用[J].机械强度,2004,26(5):534-537.
- [7] 李琴,孙春梅,黄志强,等.兰成渝腐蚀管道失效压力的GA-BP神经网络组合预测方法[J].中国安全生产科学技术,2015,11(11):83-89.
- [8] 王如君,王天瑜.灰色—马尔科夫链模型在埋地油气管道腐蚀预测中的应用[J].中国安全生产科学技术,2015,11(4):102-106.
- [9] 哈莉娅·达力列汉,袁勇博,张明媛.基于灰色关联服役构件的可靠度评估研究[J].中国安全科学学报,2011,21(11):115-121.
- [10] 张学工.关于统计学习理论与支持向量机[J].自动化学报,2000,26(1):34-42.
- [11] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Network Letters, 1999, 9(3): 293-300.
- [12] Dasgupta D. Artificial immune systems and their applications [J]. Lecture Notes in Computer Science, 1999, 1(4): 121-124.
- [13] Forrest S, Hofmeyr S A, Somayaji A. Computer immunology[J]. Immunological Reviews, 2007, 216(1): 176-197.

(编辑 尹淑容)