

SVM 性能的免疫鱼群多目标优化研究

刘 胜¹, 李高云^{1,2}, 江 娜¹

(1. 哈尔滨工程大学 自动化学院, 黑龙江 哈尔滨 150001; 2. 中国船舶重工集团公司第七〇七研究所九江分部, 江西 九江 332007)

摘 要: SVM 算法的训练精度和训练速度是衡量其性能的 2 个重要指标. 以这 2 个指标为目标变量建立 SVM 性能多目标优化问题的数学模型, 采用直接对多个目标同时进行优化的方法求得问题的 Pareto 近似解集. 在求解 Pareto 近似解集时, 将免疫原理中的浓度机制引入基本鱼群算法中, 形成一种改进的免疫鱼群算法. 以非线性动态系统仿真数据为样本数据, 并采用改进的免疫鱼群算法求解 SVM 性能多目标优化问题的 Pareto 近似解集. 仿真结果表明, 在解决多目标优化问题时, 免疫鱼群算法相对于基本鱼群算法和遗传算法具有更好的优越性.

关键词: 支持向量机; 多目标优化; Pareto 近似解集; 免疫鱼群算法

中图分类号: TP183 **文献标识码:** A **文章编号:** 1673-4785(2010)02-0144-06

Multi-objective optimization of an immune fish swarm algorithm to improve support vector machine performance

LIU Sheng¹, LI Gao-yun^{1,2}, JIANG Na¹

(1. College of Automation, Harbin Engineering University, Harbin 150001, China; 2. Jiujiang Branch of 707 Research Institute, China Shipbuilding Industry Corporation, Jiujiang 332007, China)

Abstract: Accuracy and speed when training a support vector machine (SVM) algorithm provides critical measurements of the algorithm's performance. To optimize performance, a mathematical model of multi-objective optimization with improvements in these two parameters as goals was established. A Pareto approximate solution set was obtained by optimizing multiple targets simultaneously. In the process of finding the Pareto approximate solution set, a concentration mechanism from an immune algorithm was introduced into the basic artificial fish swarm algorithm. This produced significant improvements and resulted in the proposed immune fish swarm algorithm. Taking the non-linear dynamic system simulation data as sample data, a Pareto approximate solution set of multi-objective optimization of SVM performance was obtained using the improved algorithm. Simulation results showed that, for solving multi-objective optimization, the immune fish swarm algorithm was superior to both a basic artificial fish swarm algorithm and to genetic algorithms.

Keywords: support vector machines; multi-objective optimization; Pareto approximate solution set; immune fish swarm algorithm

支持向量机 (support vector machine, SVM)^[1-2] 算法的训练精度是最受关注的性能指标之一, 同时训练速度也是一个关键因素, 尤其在学习速度要求较高的场合. 本文综合考虑算法的训练精度和速度, 从多目标优化 (multi-objective optimization, MOO) 的角度出发, 对 SVM 回归算法的 2 个指标进行研究和探讨. Schaffer 首次采用进化算法研究了 MOO 问题.

目前, MOO 问题的求解方法可以大致分为 2 类, 一类是将 MOO 问题转化为单目标优化问题进行求解, 这种方法最终只能求出一个综合了各个目标分量信息的最优解, 但各个目标的重要程度是在构造相应的单目标问题时确定的, 不具有普遍性. 另一类方法是直接对多个目标同时进行优化, 即多目标问题的直接求解方法. MOO 直接求解法多数为智能类算法, 都是同时求出多目标规划问题的 Pareto 最优解集. Shelokar 等人^[3] 提出了一种基于 Pareto 支配

收稿日期: 2009-03-24.

基金项目: 黑龙江省自然科学基金资助项目 (A2004-19).

通信作者: 刘 胜. E-mail: liu.sch@163.com.

概念的多目标蚁群算法;Coello 等人^[4]提出了多目标粒子群算法,将 Pareto 支配关系与粒子群优化相结合,改善了算法的搜索能力.其方法都挺不错,但算法比较复杂.

如何求解 Pareto 最优解集仍是关键问题,本文将免疫原理中的浓度机制引入基本人工鱼群算法 (artificial fish swarm algorithm, AFSA) 中,改进了基本鱼群算法,即免疫鱼群算法,并采用该算法来求取训练速度和训练精度的 Pareto 近似解集.

1 SVM 性能多目标优化方法

1.1 Pareto 最优解集

本文采用直接对多个目标同时进行优化的方法来求解 SVM 性能的 MOO 问题,其关键是求解 Pareto 最优解集.求解 SVM 性能多目标优化问题的 Pareto 最优解集,如图 1 所示.其中, r 反映训练精度指标, t 反映学习速度指标.

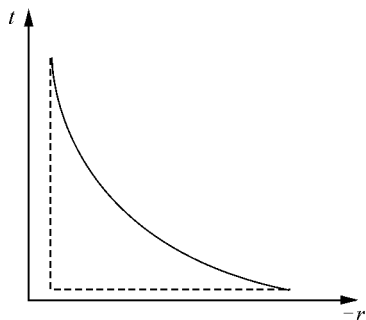


图 1 SVM 性能的多目标 Pareto 解集
Fig.1 Sketch map of multi-objective Pareto solution set of SVM performance

根据算法的具体应用场合,即针对训练速度和训练精度的不同要求,从 Pareto 解集中选择耗时与精度适宜的解.同时,可以通过 Pareto 解集中解的分布情况来判断单个目标的变化趋势,在选择最优解时以某个目标较小的牺牲来换取另一个目标较大的性能提升,使得总体效益最大化.

1.2 免疫鱼群算法

2002 年李晓磊等提出了人工鱼群随机搜索优化算法,简称鱼群算法^[5-6].该方法根据实际鱼的运动、聚集等行为构造人工鱼,通过觅食、聚群及追尾行为等改变自身的位置,经过一段时间后,各人工鱼在各极值处聚集,再由各人工鱼局部寻优找到全局最优值.

基本鱼群算法容易丧失群体的多样性,优化结

果常收敛至单个解,有时甚至出现整个群体都趋同的现象.显然,无论该单个解是否为全局最优解或近似最优解,均未实现多目标优化的目的.因此必须采取措施保持进化群体的多样性,避免未成熟收敛产生局部劣解,使优化结果呈现为一组不同的 Pareto 最优解.

本文深入分析免疫算法^[7]的机理特征,以下 2 条免疫机理特征尤其值得借鉴.

1)产生多样抗体的能力:通过细胞的分裂和分化作用,免疫系统可产生大量的抗体来抵御抗原,对应于鱼群算法中个体的多样性,这种机制可提高鱼群算法的全局搜索能力而不陷于局部解.

2)自我调节机制:免疫系统具有维持免疫平衡的机制,通过对抗体的抑制和促进作用能自我调节产生适当数量的必要抗体,对应于鱼群算法中个体浓度的抑制和促进.模拟这一功能可提高鱼群算法的局部搜索能力.

基于以上免疫机理的特点,将免疫原理中的浓度机制引入到基于多目标优化的鱼群算法的适应度函数设计中,旨在改进算法的群体多样性.免疫鱼群算法流程图如图 2 所示.

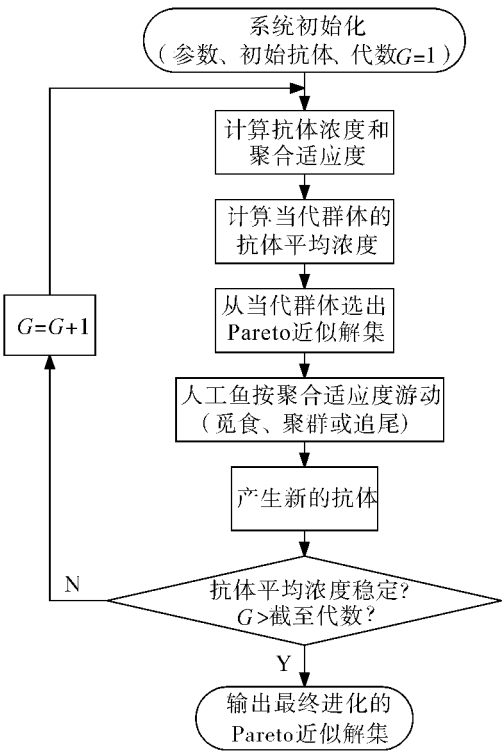


图 2 改进的免疫鱼群算法流程图
Fig.2 Flow chart of improved immune fish swarm algorithm

图 2 中,“抗体平均浓度稳定”是指在进化过程中

的后若干代与前若干代的抗体平均浓度的差值不超过一定的范围,则可以认为已趋于稳定.也可以使用传统的方法作为进化停止准则,即进化到规定代数.

在采用免疫鱼群算法求解多目标优化问题时,把目标函数和决策变量作为抗原,保证了所生成的抗体直接与问题相关联;生成的抗体能有效地排除抗原,也就相当于求得了问题的最优解;对与抗原亲和力和高的抗体进行记忆,能促进快速求解,即当遇到同类抗原时,可以快速生成与之对应的抗体.

2 SVM 性能多目标优化数学建模

2.1 目标变量

SVM 性能多目标优化问题的第 1 个目标是训练精度 r ,如果拟合误差过大,则会对下一步的预报造成影响,形成更大的预报误差;第 2 个目标是反映学习速度的时间参数 t ,如果学习时间过长,即使达到了较高的训练精度,也是得不偿失.

2.2 决策变量

影响 SVM 性能的参数主要包括惩罚因子 C ,损失函数参数 ε 和核函数参数.参数 C 、 ε 和核函数参数的选择,对于 SVM 回归估计的拟合精度和学习速度影响很大.

参数 C 可以根据样本数据的特性,决定模型的复杂度和对大于 ε 的拟合偏差的惩罚程度. C 值过大或过小,都会使系统的泛化能力变差.参数 ε 表明了系统对估计函数在样本数据上误差的期望,其值影响了构造回归函数的支持向量数目.核函数的参数反映了训练数据的特性,对于系统的泛化能力影响较大.2 种常见的核函数为

多项式核函数:

$$K_{\text{poly}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^d, \quad (1)$$

RBF 核函数:

$$K_{\text{RBF}} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma^2). \quad (2)$$

式中: $\mathbf{x}_i, \mathbf{x}_j$ 为输入空间的 2 个向量, d 和 σ 分别为多项式和 RBF 核函数的参数.由于每一种核函数对提高 SVM 性能的能力有限,所以引用一个参数 λ 来构造混合模型:

$$\begin{aligned} K_{\text{mix}}(\mathbf{x}_i, \mathbf{x}_j) &= \lambda K_{\text{poly}}(\mathbf{x}_i, \mathbf{x}_j) + \\ &\quad (1 - \lambda) K_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j), \\ 0 &\leq \lambda \leq 1. \end{aligned} \quad (3)$$

但是如何确定 λ , 目前尚未找到一种有效的方法.

只有正确选择上述参数,才可以使 SVM 回归估

计得到较高的拟合精度和较快的收敛速度.综合 C 、 ε 以及核函数参数, SVM 多目标优化的决策变量为

$$\mathbf{X} = \{\lambda, \sigma, d, C, \varepsilon\},$$

$$0 \leq \lambda \leq 1, \sigma, d, C, \varepsilon \geq 0. \quad (4)$$

2.3 目标函数

SVM 性能多目标优化问题中应使拟合精度 r 较高,而学习时间 t 较短;但在目标优化问题中是以函数值最小来建立目标函数的,所以以 $-r$ 和 t 为目标变量来建立目标函数,如式(5)所示:

$$\text{Min } Y = \text{Min}\{t, -r\}. \quad (5)$$

3 SVM 性能多目标的免疫鱼群优化

3.1 算法步骤

1) 抗原识别.

输入抗原,抗原即为待优化的目标函数及其约束条件.

2) 初始抗体生成.

假设鱼群规模定义为 100,在决策空间 \mathbf{X} 中随机产生 100 个不同的人工鱼抗体组成人工鱼初始群体.

3) 抗体适应度计算.

采用非劣分层算法进行个体适应度计算.

① 使用种群非劣分层算法将种群 \mathbf{P} 进行非劣分层.非劣分层的结果是将种群 \mathbf{P} 分成 ρ 组非劣组 \mathbf{P}_j .

$$\mathbf{P} = \cup_{j=1}^{\rho} \mathbf{P}_j. \quad (6)$$

第 1 组 \mathbf{P}_1 的所有个体属于种群的最优非劣组,第 2 组 \mathbf{P}_2 的所有个体属于种群的次优非劣组, ..., 最后一组 \mathbf{P}_ρ 的所有个体属于最劣组.

② 因为种群中最优非劣组的个体与种群中其他的个体相比最接近真正的 Pareto 最优域,所以,分配最高的适应值给最优非劣组的个体,然后逐步地分配较差的适应值给更劣层非劣组的个体.分配给第 1 层最优非劣组的任一个抗体 i 的适应值 $f_i = N = 100$ (N 为抗体种群大小).

③ 在每组内部使用共享函数法. \mathbf{P}_1 组的每个抗体 i 与同一组的另一个抗体 j 的标准化欧拉距离 d_{ij} 计算如下:

$$d_{ij} = \sqrt{\sum_{s=1}^N \left(\frac{x_s^{(i)} - x_s^{(j)}}{x_s^{\max} - x_s^{\min}} \right)^2}. \quad (7)$$

式中: N 为决策变量的个数.

④ 利用式(8)计算共享函数值.

$$\text{Sh}(d_{ij}) = \begin{cases} 1 - \left(\frac{d_{ij}}{\sigma_{\text{share}}}\right)^\alpha, & d_{ij} \leq \sigma_{\text{share}}; \\ 0, & \text{其他.} \end{cases} \quad (8)$$

式中: $\alpha = 2$, 共享函数依据 d_{ij} 在 $(0, 1)$ 之间取值, σ_{share} 是人为设定的小生境半径, 任何与抗体 i 的距离大于 σ_{share} 的抗体 j 对共享函数值不作贡献。

⑤ 在计算完 $|P_1|$ 个共享函数值以后, 利用式(9)计算抗体 i 的小生境数 E_i .

$$\text{nc}_i = \sum_{j=1}^{\mu(r_1)} \text{sh}(d_{ij}). \quad (9)$$

⑥ 依次类推, 直到计算完第 1 层所有的抗体的小生境数. 将抗体的适应度值调整为

$$f_i = f_i / \text{nc}_i. \quad (10)$$

⑦ 再令第 2 层抗体的适应度值为 $f_i = N - \rho / N$, 计算第 2 层所有抗体调整以后的适应度值. 然后, 令第 3 层抗体的适应度值为 $f_i = N - 2\rho / N$, 计算第 3 层所有抗体调整以后的适应度值. 依次类推, 计算出所有抗体的适应度值, 作为抗体的亲和度 a_i .

⑧ 根据抗体浓度的定义, 计算所有抗体的浓度 C_i (第 i 个抗体的浓度).

⑨ 计算抗体的聚合适度:

$$f_i = f_i / (1 + k \ln C_i) \quad (11)$$

作为抗体最后的适应度. 式(11)中 $k > 0$, 为调节系数, 可根据群体规模和经验而定, 本文取值为 5.

4) 产生新抗体.

抗体即为鱼群中的人工鱼, 第 i 条人工鱼 (抗体 i) 的位置用向量 \mathbf{x}_i 表示, 将抗体 i 的适应度值 f_i 作为该位置的食物浓度值 Y_i , 人工鱼 (抗体) 通过觅食行为、聚群行为或追尾行为游动, 到达新的位置, 即产生了新抗体.

① 觅食行为:

$$\mathbf{x}_{\text{inext}} = \begin{cases} \mathbf{x}_{i0} + \text{rand} \cdot \text{step} \frac{\mathbf{x}_{i1} - \mathbf{x}_{i0}}{d_{01}}, & Y_{i1} > Y_{i0}; \\ \mathbf{x}_{i0} + \text{rand} \cdot \text{step}, & Y_{i1} < Y_{i0}. \end{cases} \quad (12)$$

② 聚群行为:

$$\mathbf{x}_{\text{inext}} = \mathbf{x}_i + \text{rand} \cdot \text{step} \frac{(\mathbf{x}_{\text{center}} - \mathbf{x}_i)}{d_{\text{icenter}}}. \quad (13)$$

③ 追尾行为:

$$\mathbf{x}_{\text{inext}} = \mathbf{x}_i + \text{rand} \cdot \text{step} \frac{\mathbf{x}_{\text{max}} - \mathbf{x}_i}{d_{i \text{ max}}}. \quad (14)$$

式中: \mathbf{x}_{i1} 为第 i 条鱼 (抗体 i) 当前的位置; $\mathbf{x}_{\text{inext}}$ 为该

人工鱼下一步游动到的位置, 处于该位置的人工鱼即为产生的新抗体.

5) 抗体记忆细胞的分化和更新.

自然免疫系统在第 2 次遇到某抗原时, 免疫反应能够迅速有效排除抗原, 免疫学称为二次应答. 自然免疫系统甚至能对以前从未遇到的抗原产生二次应答, 二次应答具有时间短、高速率及合成抗体持续时间长等特点. 二次应答反映了免疫具有系统的记忆功能. 这里用记忆库存储高亲和度的抗体细胞来模拟自然免疫的记忆功能. 记忆库是适时更新的, 用每次产生的具有更高亲和度的抗体细胞取代记忆库中亲和度低的记忆细胞.

6) 输出 Pareto 近似解集.

采用抗体的平均浓度稳定来结束迭代. 即:

$$\left| \frac{C_{\text{avg}}^{k+1} - C_{\text{avg}}^k}{C_{\text{avg}}^{k+1}} \right| < \varepsilon. \quad (15)$$

式中: C_{avg}^{k+1} 和 C_{avg}^k 分别表示第 $k+1$ 和第 k 代抗体种群的平均浓度, ε 的取值通常位于 $[0.001, 0.01]$. 本文 ε 取值为 0.005.

3.2 优化结果及算法性能评价

本文采用 SVM 回归算法对非线性动态系统数据进行预报, 以训练速度和训练精度为共同目标, 采用改进的免疫鱼群算法对这个 MOO 问题进行优化. 所选取的非线性动态系统由式(16)所描述^[8].

$$y(t+1) = \frac{y(t)y(t-1)[y(t)+1.5]}{1+y^2(t)+y^2(t-1)} + u(t). \quad (16)$$

对上述系统模型, 进行仿真以收集数据样本来训练网络. 在系统的输入端加上一个随机信号得到一组 2 000 个输入—输出训练数据对, 取前 1 000 个作为训练样本, 后 1 000 个作为测试样本, 采样周期 0.1 s, 结果曲线如图 3 和图 4 所示.

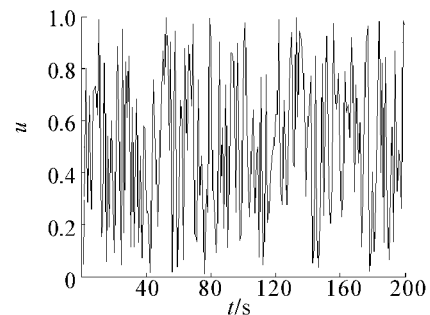


图3 输入数据曲线

Fig. 3 Curve of input data

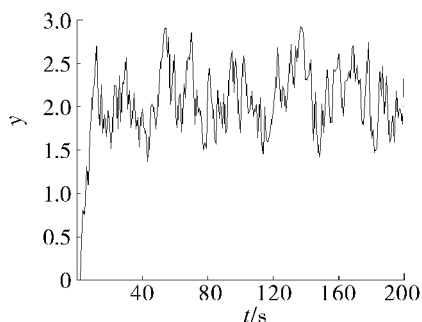


图4 输出数据曲线

Fig.4 Curve of output data

采用改进的免疫鱼群算法求解训练精度和学习速度,同时优化这一多目标优化问题的 Pareto 近似解集.图5是采用图形化方法表示所求得训练精度和训练精度的 Pareto 近似解集.

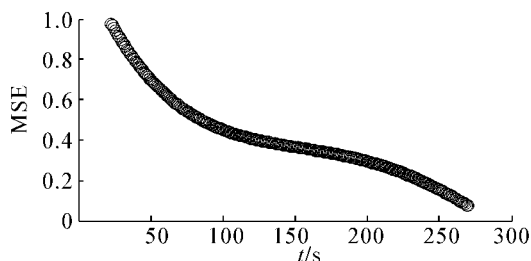


图5 Pareto 近似解集结果

Fig.5 Results of Pareto approximate solution set

下面从以下3个方面定量地对算法所求得的 Pareto 近似解集进行评估和比较:

1) 能否得到分布性能最好(尽量为均匀分布)的近似集,采用均匀分布(uniform distribution, UD)作为评价指标.

2) 能否进化得到最接近 Pareto 最优前端的近似集,采用代平均距离(generational distance, GD)作为评价指标.

3) 能否获得在目标空间分布范围最广的近似集,即进化产生的非支配解集在每维目标空间能够覆盖尽可能宽的区域范围,采用前端范围(front spread, FS)作为评价指标.

其中,均匀分布 UD、代平均距离 GD、前端范围 FS 的求解方法见文献[9-11].

为了验证免疫鱼群算法对 SVM 性能多目标优化的有效性和优越性,同时将遗传算法和基本鱼群算法对 SVM 性能多目标优化的结果与之对比,表1列出了以上3种方法所求得的 Pareto 近似解集的代平均距离 GD、均匀分布 UD、前端范围 FS.

表1 Pareto 近似解集的性能比较

Table 1 Performance comparison of Pareto approximate solution set

优化方法	代平均距离 GD	均匀分布 UD	前端范围 FS
遗传算法	0.401	0.647	1 287.6
基本鱼群算法	0.320	0.591	1 342.7
免疫鱼群算法	0.259	0.536	1 394.2

从表1中可以看出,免疫鱼群算法相对于遗传算法和基本鱼群算法所求得的 Pareto 近似解集与 Pareto 最优前端距离最近(代平均距离 GD 最小),分布性能最好(均匀分布 UD 最小),分布范围也最广(前端范围 FS 最大).这说明免疫鱼群算法在 SVM 性能多目标优化问题的求解中表现出更大的优越性.

4 结束语

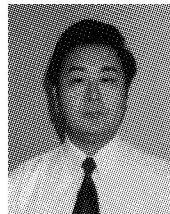
综合考虑训练精度和学习速度2个目标变量,从多目标优化的角度建立了 SVM 性能的多目标优化模型,并针对基本鱼群算法的不足,将免疫原理中的浓度机制引入多目标优化的鱼群算法适应度函数设计中,形成了一种改进的免疫鱼群算法.以文献[8]非线性动态系统仿真数据为样本数据,采用免疫鱼群算法优化 SVM 性能的多目标优化模型,求得了训练精度和学习速度的 Pareto 近似解集.最后,将遗传算法、基本鱼群算法和免疫遗传算法优化所得的 Pareto 近似解集的代平均距离 GD、均匀分布 UD、前端范围 FS 进行比较,仿真结果表明,虽然免疫鱼群算法较之基本鱼群算法增加了算法复杂度,但所求得的 Pareto 近似解集与 Pareto 最优前端距离最近,分布性能最好,分布范围也最广,在 SVM 性能多目标优化问题的求解中具有更大的优越性.

参考文献:

- [1] 刘 胜,李妍妍.基于支持向量机的锅炉过热系统建模研究[J].热能动力工程,2007,22(1):38-41.
LIU Sheng, LI Yanyan. A study of the modeling of a boiler superheating system based on a supportive vector machine [J]. Journal of Engineering for Thermal Energy and Power, 2007,22(1):38-41.
- [2] APNIK V N. The nature of statistical learning theory[M]. New York: Springer-Verlag,1995:24-35.
- [3] SHELOKAR P S, ADHIKARI S, VAKIL R, JAYARAMAN V K. Multi-objective ant algorithm: combination of strength

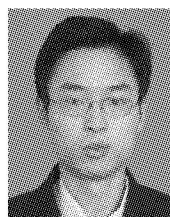
- Pareto fitness assignment and thermodynamic clustering[J]. Foundations of Computing and Decision Sciences, 2000, 25(4): 213-229.
- [4] COELLO C A C, PULIDO G T, LECHUGA M S. Handling multiple objectives with particle swarm optimization[J]. IEEE Transaction on Evolutionary Computation, 2004, 8(3): 256-279.
- [5] 李晓磊. 一种新型的智能优化方法-人工鱼群算法[D]. 杭州: 浙江大学, 2003.
- LI Xiaolei. A new intelligent optimization method-artificial fish school algorithm[D]. Hangzhou: Zhejiang University, 2003.
- [6] 张梅凤, 邵 诚, 甘 勇. 基于变异算子与模拟退火混合的人工鱼群优化算法[J]. 电子学报, 2006, 34(8): 1381-1385.
- ZHANG Meifeng, SHAO Cheng, GAN Yong. Hybrid artificial fish swarm optimization algorithm based on mutation operator and simulated annealing[J]. Acta Electronica Sinica, 2006, 34(8): 1381-1385.
- [7] 莫宏伟. 人工免疫系统原理与应用[M]. 哈尔滨: 哈尔滨工业大学出版社, 2002: 27-58.
- [8] 张浩然, 韩正之, 李昌刚. 基于支持向量机的未知非线性系统辨识与控制[J]. 上海交通大学学报, 2003, 37(6): 927-930.
- ZHANG Haoran, HAN Zhenghi, LI Changgang. Support vector machine based unknown nonlinear systems identification and control[J]. Journal of Shanghai Jiaotong University, 2003, 37(6): 927-930.
- [9] VELDHUIZEN D A V, LAMONT G B. On measuring multiobjective. evolutionary algorithm performance[C]//Proceedings of the 200 Congress on Evolutionary Computation. California, CA, USA, 2000: 204-211.
- [10] TAN K, LEE T, KHOR E. Evolutionary algorithms with dynamic population size and local exploration for multiobjective optimization[J]. IEEE Transactions on Evolutionary Computation, 2001, 7(3): 565-588.
- [11] KAPUR J, KESAVAN H. Entropy optimization principles with applications[M]. San Diego: Academic Press, 1992: 214-258.

作者简介:



刘 胜,男,1957年生,教授,博士生导师,黑龙江省教学名师,黑龙江省重点一级学科"控制科学与工程"学科负责人.兼任教育部工程研究中心"船舶控制工程研究中心"主任,中国造船

学会仪器仪表学术委员会副主任,黑龙江省自动化学会副理事长.主要研究方向为智能控制、鲁棒控制、船舶航行与姿态控制.目前承担国家"973"计划项目,国防基础研究基金项目、国防预研项目4项,省部级项目6项.曾获黑龙江省优秀教学工作者、中国船舶工业总公司优秀青年科技工作者、获省部级科学技术奖7项、获省教学成果奖一等奖2项,二等奖2项、获省教育科学研究成果一等奖4项、获省部级自然科学学术成果奖8项.发表学术论文150余篇,被SCI、EI、STP检索70余篇,出版学术著作3部.



李高云,男,1981年生,博士研究生,主要研究方向为智能控制、故障诊断与容错控制、船舶航行与姿态控制.参与科研项目3项,获黑龙江省科学技术二等奖1项,黑龙江省高校科学技术一等奖1项,发表学术论文近10篇,被EI、ISTP检索4篇.



江 娜,女,1981年生,博士研究生,主要研究方向为故障诊断与故障预报.参与科研项目2项,发表学术论文近10篇,被EI、ISTP检索5篇.