

基于遗传算法和最小二乘支持向量机可靠性分配*

张根保, 刘 佳, 王国强, 任显林
(重庆大学 机械工程学院, 重庆 400030)

摘 要: 为了提高系统可靠性的精确快速分配, 采用支持向量机对系统可靠性进行建模, 采用逆向思维对系统可靠性进行分配; 为了提高求解速度和鲁棒性, 用最小二乘法对支持向量机进行算法优化, 并用遗传算法对最小二乘支持向量机进行参数优化; 为了提高分配精度, 用三角模糊数进行模糊处理; 最后针对某系统的可靠性, 采用遗传算法优化和模糊处理的最小二乘支持向量机进行分配, 并与神经网络和普通遗传算法优化的最小二乘支持向量机进行对比。结果表明, 用遗传算法优化和模糊数处理的最小二乘支持向量机具有分配精度高, 泛化能力强等优点。

关键词: 可靠性分配; 遗传算法; 最小二乘支持向量机; 逆向思维; 三角模糊数

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2010)09-3300-03

doi: 10.3969/j.issn.1001-3695.2010.09.026

Reliability allocation based on genetic algorithm and LS-SVM

ZHANG Gen-bao, LIU Jia, WANG Guo-qiang, REN Xian-lin
(College of Mechanical Engineering, Chongqing University, Chongqing 400030, China)

Abstract: For improving precise and rapid system reliability allocation, made the model by support vector machines, used the reliability by reverse thinking. In order to improve the solution speed and robustness, allocated least square method to optimization. At the same time used genetic algorithm for parameter optimization in least squares support vector machines. Used the triangle fuzzy number for improving distribution accuracy. At last allocated some system reliability by using least squares support vector machines which was optimized by genetic algorithm and triangle fuzzy number. Compared with neural network, the results show the least squares support vector machines which was optimized by genetic algorithm and triangle fuzzy number has the advantages include such as high accuracy and strong generalization ability.

Key words: reliability allocation; genetic algorithm; least squares SVM; reverse thinking; triangle fuzzy number

可靠性分配问题(reliability allocation problem, RAP)已成为产品可靠性设计的重要问题,传统的等分配法、AGREE法和比例分配法等已不能满足复杂系统可靠性准确分配的要求,因此复杂系统的可靠性分配与优化已成为可靠性设计中亟需解决的问题。Suman和Liang等人分别用模拟退火算法和蚁群算法对可靠性分配进行优化^[1,2],Tian和Liu等人分别用故障树和Monte Carlo与遗传算法相结合的方式对可靠性进行分配^[3,4]。Ramirez-Marquez和贺星采用了神经网络对系统的可靠性进行分配^[5,6]。但是这些方法所用的数据大多是依靠专家的经验,缺乏真实性,不能对复杂系统实现可靠性的精确分配^[7]。因此可以在可靠性预测的基础上,通过各子系统约束条件变化引起系统可靠度的变化程度来确定子系统在可靠性分配中所占的权重,采用逆向思维的方式对系统可靠性进行分配^[7]。由于神经网络存在局部最优解和泛化能力弱等缺点,由Vapnik^[8]提出的支持向量机(support vector machines, SVM)具有严格的数学基础,很好的泛化能力,并且能够得到全局最优解, SVM在模式识别、函数逼近和预测建模等领域应用广泛。Suykens等人^[9]提出的最小二乘支持向量机(least squares

support vector machine, LS-SVM)采用最小二乘线性系统误差平方和作为损失函数,将不等式约束改为等式约束,将二次规划问题转换为线性方程组求解,降低了计算复杂性,因此基于该算法的预测方法将具有更快的求解速度和更好的鲁棒性^[10]。由于最小二乘支持向量机的参数选择将影响最后的精度,遗传算法以其强大的全局搜索能力、并行性、高效性的优点得到了广泛的应用^[11]。为了提高系统可靠性的精确性,本文采用遗传算法优化最小二乘支持向量机的逆向思维方式对系统可靠性分配方法进行研究。

1 系统可靠性分配约束与算法基础分析

1.1 可靠性分配的约束条件

可靠性的分配问题实质是在某种约束条件下的目标优化问题,因此在进行可靠性分配时,必须明确目标函数和约束条件^[12]。本文采用逆向思维对系统可靠性进行分配,因此分配应从子系统本身来考虑,采用子系统自身的约束条件如重要度、复杂度、工作时间、工作环境、费用、技术水平等各个方面的

收稿日期: 2010-03-15; **修回日期:** 2010-04-22 **基金项目:** 国家“863”计划资助项目(2009AA04Z119); 国家自然科学基金资助项目(50835008); 华中科技大学数字制造国家重点实验室开放基金资助项目

作者简介: 张根保(1953-),男,山西新绛人,教授,博导,博士,主要研究方向为数字化可重构制造系统及设备、机床可靠性和质量管理(gen.bao.zhang@263.net); 刘佳(1985-),男,四川遂宁人,硕士研究生,主要研究方向为可靠性; 王国强(1979-),男,山西新绛人,博士研究生,主要研究方向为可靠性; 任显林(1978-),男,山西祁县人,博士研究生,主要研究方向为免废质量。

数据对系统的可靠性进行分配^[7]。在对系统可靠性分配之前需要从各个方面对子系统的约束条件进行综合充分考虑,收集各子系统的约束条件和数据,为模型的建立做好前期准备。

1.2 最小二乘支持向量机

训练数据的样本可以表示为 (x_i, y_i) $x_i \in R^n, y_i \in R$ 。其中: y_i 是输出向量, x_i 是输入向量, 利用结构风险最小化原则 (structural risk minimization, SRM) 进行风险最小化, 则函数逼近的最小二乘支持向量机的优化问题为 $\min_{\omega, b, e} J(\omega, e) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^l e_i^2$ 。约束条件为 $y_i = \varphi(x_i) \times \omega + b + e_i, i = 1, \dots, l$ 。其中: ω 为权向量; γ 为正则化参数; e_i 为误差向量; b 为偏置量。

最小二乘支持向量机优化问题对应的拉格朗日函数为:

$$L(\omega, b, e, \alpha) = J(\omega, e) = \sum_{i=1}^l \alpha_i (\omega^T \varphi(x_i) + b + e_i - y_i) \quad (1)$$

其中: $\alpha_i (i = 1, \dots, l)$ 为拉格朗日乘子。

由于在任何约束条件下, 所有的最优化问题都必须满足 KKT (Karush-Kuhn-Tucker) 条件, 从而可得 $\frac{\partial L}{\partial \omega} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial e_i} = 0, \frac{\partial L}{\partial \alpha_i} = 0$ 。消去 ω 和 e 可得 $\begin{bmatrix} 0 & I \\ I & ZZ^T + \gamma^{-1} E \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}$ 。其中: $y = [y_1, y_2, \dots, y_l]^T, I = [1, \dots, 1]^T, \alpha = [\alpha_1, \alpha_2, \dots, \alpha_l]^T, E$ 为 $l \times l$ 阶的单位矩阵, $Z = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_l)]^T$; 核函数矩阵 $\Omega = ZZ^T$; 则 $\Omega_i = K(x_k, x_i) = \varphi(x_k)^T \varphi(x_i)$; 核函数 $K(x_k, x_i)$ 是满足 Mercer 条件的对称函数。

最小二乘支持向量机的预测函数为

$$y(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b \quad (2)$$

径向基 (RBF) 核函数是最常用的核函数 $K(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2)$ 。其中: σ 为径向基函数的宽度。

1.3 遗传算法优化最小二乘支持向量机的参数选择

对于采用径向基核函数的 LS-SVM, 主要参数是正则化参数 γ 和核函数宽度 σ , 这两个参数在很大程度上决定了 LS-SVM 的学习和泛化能力^[13]。因此选择径向基核函数的问题可以简化为寻找参数 γ 和 σ 的最佳组合, 使 LS-SVM 具有最好的性能。目前对于支持向量机参数选择应用较多的是采用网格搜索与交叉验证相结合的优化算法, 然而此算法计算量较大, 尤其是训练样本较大时搜索过程非常费时^[11], 因此该方法具有一定局限性。而遗传算法因为具有强大的全局搜索能力、并行性、高效性使之成为 LS-SVM 参数优化算法的较好选择。

遗传算法^[14]将问题的求解表示成染色体 (用编码表示字符串)。该算法从一群染色体串出发, 根据适者生存的原则, 从中选择出适应度高的染色体进行复制, 通过交叉、变异两种基因操作产生出新一代的更适应环境的染色体种群。随着遗传算法一代一代地进行, 那些适应度高的模式将在后代中呈指数增长, 最终得到适应度最高的染色体, 即优化问题的最优解。用均方相对误差 D_r 作为模型最终性能评价指标:

$$D_r = \sqrt{\frac{1}{l} \sum_{i=1}^l \left(\frac{y_{ii} - y_{\hat{p}i}}{y_{ii}} \right)^2} \quad (3)$$

其中: y_{ii} 表示实际值, $y_{\hat{p}i}$ 表示预测值。

遗传算法设计如下:

a) 编码方式。由于 LS-SVM 模型需要优化两个参数 γ 和 σ , 采用二进制编码。

b) 适应度函数为 $f(\gamma, \sigma) = 1/D_r$ 。

c) 遗传算子。按适应度值大小选择个体, 然后运用交叉和变异算子形成新个体。

d) 运行参数。种群规模 M 和终止进化代数 T 需要提前设定, 交叉概率 P_c 和变异概率 P_m 采用自适应遗传算法^[15], 即

$$P_c = \begin{cases} \frac{k_1 (f_{\max} - f')}{f_{\max} - f_{\text{avg}}} & f' \geq f_{\text{avg}} \\ k_2 & f' < f_{\text{avg}} \end{cases}$$

$$P_m = \begin{cases} \frac{k_3 (f_{\max} - f)}{f_{\max} - f_{\text{avg}}} & f \geq f_{\text{avg}} \\ k_4 & f < f_{\text{avg}} \end{cases}$$

其中: f_{\max} 是每代群体中个体的最大适应度值; f_{avg} 为平均适应度值; f' 是交叉两个个体较大的适应度值; f 是变异个体的适应度值; k_1, k_2, k_3, k_4 为自适应控制参数 $\in (0, 1)$, 只要选定值, 就可以自适应调整。

2 系统可靠性分配方法设计

系统可靠性分配模型的本质就是充分利用 LS-SVM 具有强大的处理非线性多因素系统能力, 建立系统可靠性与子系统约束条件之间复杂的非线性关系, 挖掘内部规律, 从而实现利用子系统约束条件对系统可靠性的精确分配。

首先需要确定分配模型的输入和输出向量。由于本文的基本思想是在可靠性预测的基础上, 通过各子系统约束条件变化引起系统可靠度的变化程度来确定子系统在可靠性分配中所占的权重。设输入向量为: 系统可靠度和各个子系统所有的自身约束条件, 则输出向量为各个子系统可靠度的比值。对应的各个子系统的约束条件必须具有一致性。由于采集的各数据单位不一致, 在数据输入之前需进行归一化处理, 而在输出后需要对数据进行反归一化处理。

为了提高分配精度, 对可靠性预测结果进行模糊处理, 设三角模糊数为 $\tilde{p}^{(k)} = (l_{ij}^{(k)}, m_{ij}^{(k)}, u_{ij}^{(k)})$, $l_{ij}^{(k)} = m_{ij}^{(k)} - 0.1\lambda$, $u_{ij}^{(k)} = m_{ij}^{(k)} + 0.1\lambda$ 。其中: λ 为模糊度等级。具体步骤如下:

a) 对收集的数据进行前期处理, 如归一化处理。

b) 遗传算法初始化, 种群规模 $M = 100$, 终止进化代数 $T = 500$, 设定正则化参数 γ 和核函数宽度 σ 的初始值为 $(1, 0.01)$, 自适应参数 $k_1 = k_2 = 0.5, k_3 = k_4 = 0.02$ 。

c) 根据归一化后的训练数据建立 LS-SVM 模型, 并根据运行参数用遗传算法对 LS-SVM 模型参数进行优化。判断是否满足终止条件, 不满足则进行交叉、变异产生新个体, 然后重新编码产生初始种群再判断; 否则终止得到参数的最优解。

d) 根据优化后的参数建立分配模型, 对所分配的样本进行计算, 反归一化, 得到预测结果。

e) 对预测结果进行模糊处理, 取模糊度等级 $\lambda = 1$, 预测结果转换为三角模糊数, 由 $\tilde{u}_i = (\sum_{j=1}^n l_{ij}, \sum_{j=1}^n m_{ij}, \sum_{j=1}^n u_{ij}) \otimes (\sum_{i=1}^n \sum_{j=1}^n l_{ij},$

$$\sum_{i=1}^n \sum_{j=1}^n m_{ij}, \sum_{i=1}^n \sum_{j=1}^n u_{ij})^{-1} \approx \left[\frac{\sum_{j=1}^n l_{ij}}{\sum_{i=1}^n \sum_{j=1}^n m_{ij}}, \frac{\sum_{j=1}^n m_{ij}}{\sum_{i=1}^n \sum_{j=1}^n m_{ij}}, \frac{\sum_{j=1}^n u_{ij}}{\sum_{i=1}^n \sum_{j=1}^n u_{ij}} \right]$$

得模糊预测值 \tilde{u}_i , 取乐观—悲观系数 $\eta = 0.5$, 则 \tilde{u}_i 的期望值为 $I(\tilde{u}_i) = (l_i + 2m_i + u_i) / 4$ 。

f) 由模糊预测值的期望 $I(\tilde{u}_i)$ 可得子系统的权重 $\omega_i = I(\tilde{u}_i) / \sum_{i=1}^n \tilde{u}_i$ 。

g) 根据可靠度分配式和所得权重可计算出各子系统的可靠度。满足要求时则终止, 否则重新分配。

3 实例与比较

为了与文献[7]的神经网络模型进行对比, 采用此文献的数据建立分配模型, 同样将可靠度为 0.75 和 0.92 的作为训练样本, 可靠度为 0.95 的作为测试样本。首先应对数据进行归一化处理, 然后用遗传算法优化最小二乘支持向量机, 由初始化的遗传算法参数, 根据平均相对误差 D_i 是否最小或者满足终止进化代数 T 为终止条件, 不断循环可得到最优参数 $\gamma = 841.0017, \sigma^2 = 13.6999$ 。应用上面得到的最优参数对可靠度为 0.95 的样本进行测试, 结果如表 1 所示。

表 1 子系统可靠度比值测试结果

比值	真实值	GA-LS-SVM	神经网络	
			BP	RBF
R_1/R_2	1.002 915	1.002 729	1.027 2	1.022 3
R_1/R_3	1.019 415	1.018 328	1.062 1	1.069 2
R_1/R_4	1.011 765	1.010 388	1.045 2	1.051 0
R_1/R_5	1.005 341	1.002 191	0.984 0	0.992 9
R_2/R_3	1.016 452	1.015 406	1.055 6	1.074 5
R_2/R_4	1.008 824	1.006 856	1.038 4	1.051 3
R_2/R_5	1.002 419	0.998 255	0.994 7	0.989 0
R_3/R_4	0.992 495	0.992 003	1.000 6	0.955 3
R_3/R_5	0.986 194	0.983 825	0.951 5	0.929 0
R_4/R_5	0.993 651	0.992 591	0.974 6	0.977 5
平均绝对误差/%		0.169	2.6004	3.4533

从表 1 可知用遗传算法优化的 LS-SVM 测试精度比 BP 和 RBF 神经网络高很多, 主要是支持向量机本身就具有全局最优解和预测精度高等优点。

将预测结果转换为三角模糊数, 由可靠性分配步骤 e) f) 可得各子系统权重

$$\omega = (0.198\ 67, 0.199\ 294, 0.202\ 309, 0.200\ 656, 0.19\ 907)$$

由可靠性分配公式 $R_i = R^{\omega_i}$, ω_i 即为子系统的权重。从而可得各子系统分配的可靠度如表 2 所示。

表 2 子系统可靠度分配结果

真实值	BP	RBF	GA-LS-SVM	F-GA-LS-SVM
$R_1 = 0.997\ 6$	0.989 6	0.989 5	0.989 726	0.989 861
$R_2 = 0.994\ 7$	0.989 7	0.989 6	0.989 758	0.98 983
$R_3 = 0.978\ 6$	0.990 1	0.990 3	0.989 911	0.989 677
$R_4 = 0.986$	0.99	0.989 9	0.989 828	0.98 976
$R_5 = 0.992\ 3$	0.989 6	0.989 5	0.989 747	0.989 841
平均绝对误差/%	0.624	0.632	0.610 2	0.598 1

为了比较对预测结果不进行模糊处理直接建立判断矩阵求各子系统的权重, 分配结果见表 2。

文献[7]采用的 BP 和 RBF 神经网络分配结果见表 2。

从表 2 可知, 遗传算法优化的 LS-SVM 模型和 F-GA-LS-SVM 模型的平均绝对误差小于 BP 和 RBF 神经网络, 并且经过模糊处理的 GA-LS-SVM 模型高于普通 GA-LS-SVM 模型。因此用遗传算法优化和三角模糊数处理的 LS-SVM 系统可靠性分配模型不仅训练精度比神经网络建立的模型高很多, 而且分配精度也高, 并且泛化能力强, 具有全局最优解, 对于可靠性设计具有一定的指导作用。

4 结束语

本文采用最小二乘支持向量机对系统可靠性进行建模, 采

用逆向思维进行分配。将各子系统的约束条件和系统可靠度作为输入向量, 各子系统的可靠度比值作为输出向量, 通过各子系统约束条件变化引起系统可靠度的变化程度来确定子系统在可靠性分配中所占的权重, 从而实现可靠性的精确分配。为了提高模型的学习和泛化能力, 用遗传算法对模型的两个参数 γ 和 σ 进行优化。为了提高分配精度, 用三角模糊数对预测结果进行模糊处理, 对某系统可靠性的分配结果与神经网络模型和普通 GA-LS-SVM 模型相对比可知, 用遗传算法优化和模糊处理的 LS-SVM 可靠性分配模型比其他模型分配精度高, 并且泛化能力强。但是子系统的约束条件多种多样, 影响规律复杂多变, 并且训练样本较少, 这是建立的某系统可靠性分配模型的误差所在, 因此对于采用逆向思维方式和遗传算法优化的 LS-SVM 模型应全面考虑各个子系统的约束条件, 并且对于各子系统的约束条件应有较丰富的样本数据和较为准确的定量研究, 才能进一步提高系统可靠性分配的精度。

参考文献:

- [1] SUMAN B. Simulated annealing-based multiobjective algorithm and their application for system reliability [J]. *Engineering Optimization*, 2003, 35(4): 391-416.
- [2] LIANG Y C, SMITH A E. An ant colony optimization algorithm for the redundancy allocation problem[J]. *IEEE Trans on Reliability*, 2004, 53(3): 417-423.
- [3] TIAN Pei, WANG Jian-cheng, ZHANG Wei, et al. A fault tree analysis based software system reliability allocation using genetic algorithm optimization [C]//Proc of World Congress on Software Engineering. Xiamen: [s. n.], 2009:194-198.
- [4] LIU Xing-hua, XU Gui-hong, MA Chao-chen, et al. Reliability allocation and optimization of engine system by using genetic algorithm and Monte Carlo method[J]. *Journal of Beijing Institute of Technology*, 2007, 16(3):268-273.
- [5] RAMIREZ-MARQUEZ J E, ROCCO C M. All-terminal network reliability optimization via probabilistic solution discovery reliability [J]. *Engineering & System Safety*. 2008, 93(11):1689-1697.
- [6] 贺星, 孙丰瑞, 刘永葆, 等. 基于 Hopfield 神经网络的燃气轮机可靠性分配[J]. *华中科技大学学报: 自然科学版*, 2009, 37(6):48-51.
- [7] 张宏斌, 贾志新. 基于逆向思维的系统可靠性精确分配模型[J]. *计算机应用*, 2009, 29(9):2468-2470.
- [8] VAPNIK V N. The nature of statistical learning theory [M]. New York: Spring-Verlag, 1999.
- [9] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [10] 吴德会. 基于最小二乘支持向量机的铣削加工表面粗糙度预测模型[J]. *中国机械工程*, 2007, 18(7):838-841.
- [11] 王克奇, 杨少春, 戴天虹, 等. 采用遗传算法优化最小二乘支持向量机参数的方法[J]. *计算机应用与软件*, 2009, 26(7):109-111.
- [12] 刘惟信. 机械可靠性设计 [M]. 北京: 清华大学出版社, 1996.
- [13] 郭辉, 刘贺平, 王玲. 最小二乘支持向量机参数选择方法及其研究 [J]. *系统仿真学报*, 2006, 18(7):2033-2036.
- [14] 张豪, 罗亦泳, 张立亭, 等. 基于遗传算法最小二乘支持向量机的耕地变化预测[J]. *农业工程学报*, 2009, 25(7):226-231.
- [15] KEETHI S S, LIN C J. Asymptotic behaviors of support vector machines with Gaussian kernel [J]. *Neural Computation*, 2003, 15(7):1667-1689.