

文章编号: 1000-582X(2011)09-021-05

## 装配故障率的灰色最小二乘支持向量机分析

张根保, 刘 佳, 王国强, 范秀君

(重庆大学 机械工程学院, 重庆 400044)

**摘 要:** 为了对装配故障率进行定量研究, 用最小二乘支持向量机(LSSVM)对装配故障率与属性之间的关系进行了建模。在该模型中对影响故障率的 5M1E(Man, Machine, Material, Method, Measurement and Environment)因素用装配可靠性评价方法(Assembly Reliability Evaluation Method, AREM)提取的装配故障率属性进行了改进, 建立了装配故障率的全属性模型; 为提高求解效率以及使装配可靠性控制更具有目的性, 用灰色关联分析对装配故障率的属性进行提取, 得到了主要属性, 并用遗传算法对主要属性建立的装配故障率模型进行参数优化。用灰色关联分析提取的主要属性的 LSSVM 模型与全部属性建立的 LSSVM 模型和主要属性建立的 BP 神经网络模型的装配故障率预测进行比较, 结果表明用灰色关联分析的 LSSVM 故障率模型不仅建模简单而且还具有预测精度高等优点。

**关键词:** 装配故障率; 支持向量机; 5M1E; 灰色关联分析; 遗传算法

**中图分类号:** TH161.7 TB114.3

**文献标志码:** A

## Assembly fault rate analysis using grey relation and least squares support vector machines

ZHANG Gen-bao, LIU Jia, WANG Guo-qiang, FAN Xiu-jun

(College of Mechanical Engineering, Chongqing University, Chongqing 400044, P. R. China)

**Abstract:** To get the relationship between assembly fault rate and its attributes, least squares support vector machine (LSSVM) is introduced to quantitatively study assembly fault rate. Aiming at the drawbacks of assembly reliability evaluation method (AREM), the attributes of assembly-fault-rate-affecting 5M1E (Man, Machine, Material, Method, Measurement and Environment) factors obtained by AREM are improved, hence the LSSVM model with all attributes is established. To reduce the time of calculating the assembly fault rate and provide the priority for assembly reliability improvement, grey relation analysis is applied to extracting the main attributes, at the same time genetic algorithm (GA) is used for parameter optimization in LSSVM. The assembly fault rate analysis results show that the method using grey relation analysis and least square support vector machine is simpler and more accurate compared with other methods such as LSSVM model using all attributes and BP neural network using main attributes.

**Key words:** assembly fault rate; support vector machines; 5M1E; grey relation analysis; genetic algorithm

**收稿日期:** 2011-04-10

**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); 国家“高档数控机床与基础制造装备”科技重大专项(2009ZX04014-016; 2009ZX04001-013; 2009ZX04001-023; 2010ZX04014-015); 数字制造装备与技术国家重点实验室(华中科技大学)开放基金资助

**作者简介:** 张根保(1953-), 男, 重庆大学教授, 博士生导师, 主要研究方向为现代质量工程、先进制造技术和数控机床可靠性等, (E-mail) gen. bao. zhang@263. net。

为了提高产品的可靠性,国内制造厂家(特别是高档数控机床制造商)普遍采用购买国外高质量的零部件,尽管这些零部件本身的可靠性水平很高,但装配完成后的产品可靠性却远远达不到国外水平,可见装配过程中的可靠性控制对提高产品整机的可靠性举足轻重。到目前,国内外关于装配可靠性的研究工作很多,Chumakov<sup>[1]</sup>用神经网络对装配故障进行诊断,余锋杰等<sup>[2]</sup>对飞机大部件自动化对接装配的故障进行了维修决策的优化,Delvecchio 等<sup>[3]</sup>用信号处理工具对装配故障进行了监控,然而对装配故障率的研究却很少;Suzuki 等<sup>[4]</sup>用装配可靠性评价方法(AREM)对装配故障率进行了研究,提出装配故障率由设计属性和车间属性决定,车间属性可以由 BFRS(the basic fault rate of shop)表示,然而 BFRS 一般是随着工序和工人变化的<sup>[5]</sup>,设计属性和车间属性之间一般有一定的关联,在 AREM 中却把设计属性和车间属性分开研究,并且子属性对装配故障率的影响规律一般是未知的;由 Vapnik<sup>[6]</sup>提出的支持向量机(Support Vector Machines, SVM)具有严格的数学基础,有很好的泛化能力,并且能够得到全局最优解,在模式识别、函数逼近和预测建模等领域应用广泛。Suykens 等<sup>[7]</sup>提出的最小二乘支持向量机(Least Squares Support Vector Machines, LSSVM)具有更快的求解速度和更好的鲁棒性<sup>[8]</sup>。笔者采用 LSSVM 对装配故障率的属性进行建模,在装配可靠性评价方法(AREM)的基础上针对 5M1E 影响因素对装配故障率的全属性进行分析,然后采用灰色关联分析得到影响装配故障率的主要属性,为以后装配可靠性控制提供目标依据,并对 LSSVM 模型进行预测和误差分析。

## 1 基于 AREM 装配故障率的属性分析

### 1.1 基于 AREM 方法的装配故障率

由 Suzuki 等<sup>[4]</sup>提出的 AREM 被用来提高车间运行的可靠性和选择可靠性水平更好的车间,并且通过与真实的故障数据进行比较证明这种方法是有效的。AREM 将装配故障率分为设计属性和车间属性,设计属性包括工艺属性和零件属性,工艺属性通过具体装配操作来衡量,如“向下装配”、“水平装配”、“回转装配”等;零件属性则由装配零件本身来决定,例如将不平整的零件装配到平整的基座上比平整零件装配到平整基座上发生故障的可能性就要大得多,工艺属性和零件属性可以通过 AREM 符号<sup>[4]</sup>来定量计算,如表 1 所示。

表 1 AREM 符号

工艺属性			零件属性		
装配操作	符号	系数	装配表面	符号	系数
向下	↓	1	标准	—	1
水平	→	2	美观的	#3	2
斜向下	↘	2	不稳定的	u1	3
回转	∪	3	多配合	h9	配合数
...	...	...	...	...	...

在 AREM 中,车间属性可以由 BFRS 来定量研究,虽然 BFRS 可以通过  $C_{FC}$  (Fault generation coefficient,故障发生系数)、 $C_{FO}$  (Fault overlook coefficient,故障忽视系数)和  $C_{RF}$  (Response failure coefficient,响应失效系数)计算得到<sup>[4]</sup>,然而通过调查表得到的 3 个参数值缺乏一定的准确性,并且这 3 个参数对于车间的装配可靠性控制的具体因素并不明确,不能为以后的可靠性改进提供具体的措施。

### 1.2 装配故障率的全属性分析

质量管理学的研究发现:5M1E (Man, Machine, Material, Method, Measurement and Environment)6 个因素对企业的产品质量和工序的控制能力有很大的影响,同样对产品的可靠性和装配过程可靠性也有很大影响。为了能够对装配可靠性的改进提供主要的属性控制,从 5M1E 的角度出发对装配故障率的子属性进行分析,通过对生产过程中人、机、料、法、环、测(5M1E)等 6 个主要生产要素进行详细分析得到装配故障率的属性。因此对于全面提取装配故障率的属性具有重要意义。

人(Man):装配操作工人,他的技能和熟练程度等都会影响装配故障率。

机器设备(Machine):包括装配机械、工装夹具和检测仪器等,主要是车间设备情况。

材料(Material):指材料的成分、物理化学性能等,在这里指装配零部件的表面性能等对于装配故障率有重要影响。

方法(Method):包括加工工艺和操作规程等。

测量(Measurement):测量时采取的方法是否标准正确。

环境(Environment):车间的温度、湿度、照明和清洁条件等。

与 AREM 相比较可知其中材料和方法属性可以与 AREM 的零件属性和工艺属性相对应,因此材料和方法属性的值可以通过表 1 的 AREM 符合定量得到,而人、机器设备、测量和环境属性可以对应于车间属性的子属性。

## 2 装配故障率的建模与算法分析

### 2.1 装配故障率的 LSSVM 模型建立

为了明确装配故障率的全属性模型对装配故障率的具体影响规律,需要建立装配故障率的全属性模型对装配故障率的影响模型。但是由于采用神经网络的方法需要提前对网络进行优化,并确定隐含层和训练函数<sup>[9]</sup>,而支持向量机能够避免这种缺陷,因此对装配故障率建立 LSSVM 模型。模型的输入为装配故障率的属性用  $x_i$  表示,输出为装配故障率用  $y_i$  表示,利用结构风险最小化原则 (Structural Risk Minimization, SRM) 进行风险最小化,则最小二乘支持向量机的优化问题为

$$\min_{\omega, b, e} J(\omega, e) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^l e_i^2, \quad (1)$$

约束条件为

$$y_i = \varphi(x_i) \times \omega + b + e_i, i = 1, 2, \dots, l。$$

式中:  $\omega$  为权向量;  $\gamma$  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的预测函数为

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

式中:  $\alpha_i$  为拉格朗日乘子;  $K(x, x_i)$  是满足 Mercer 条件的对称核函数。

径向基(RBF)核函数是最常用的核函数  $K(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2)$ ,  $\sigma$  为径向基函数的宽度。对于采用径向基核函数的 LSSVM, 正则化参数  $\gamma$  和核函数宽度  $\sigma$  在很大程度上决定了 LSSVM 的学习和泛化能力<sup>[10]</sup>。因此选择径向基核函数的问题可以简化为寻找参数  $\gamma$  和  $\sigma$  的最佳组合, 使 LSSVM 具有最好的性能, 在参数优化中遗传算法成为了较好选择<sup>[11-13]</sup>。

### 2.2 装配故障率主要属性的灰色关联分析

由于装配故障率的属性涉及的影响因素太多, 为了有重点的控制装配可靠性和减少数据输入提高预测效率, 需要提取装配故障率的主要属性。由于装配故障率的属性对装配故障率的影响程度或影响规律缺乏, 因此装配故障率的属性对装配故障率构成了一个灰色系统。灰色系统可以采用灰色关联分析方法, 它用关联度作为事物之间和因素之间的动态发展态势及相似程度<sup>[14]</sup>, 对于样本量的多少和样本有无规律都同样适用, 比传统的数理统计方法更具优势, 并且分析结果较准确。因此采用灰色关联度排序的方法对属性进行分析。

由于影响装配故障率的属性含义和单位不同, 如果两个序列间的数据在大小上相差很大, 则小数

值序列的作用容易被大数值序列掩盖, 为了便于比较, 保证各因素间具有等效性和同序性, 需对原始数据进行归一化处理<sup>[15]</sup>。

设系统的参考序列为

$$X_0 = (x_0(1), x_0(2), \dots, x_0(n)),$$

比较序列为

$$X_i = (x_i(1), x_i(2), \dots, x_i(n)) (i = 1, 2, \dots, m),$$

则属性之间的灰色关联系数为

$$\gamma(x_0(k), x_i(k)) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \xi \max_i \max_k |x_0(k) - x_i(k)|}{\xi \max_i \max_k |x_0(k) - x_i(k)|}, \quad (3)$$

式中  $\xi$  为分辨系数,  $\xi \in (0, 1)$ , 通常取  $\xi = 0.5$ 。

比较序列  $X_i$  与参考序列  $X_0$  的灰色关联度为

$$\gamma_{0i} = \frac{1}{n} \sum_{k=1}^n \gamma(x_0(k), x_i(k)) (i = 1, 2, \dots, m)。 \quad (4)$$

根据求出的灰色关联度排序就可以提取出影响装配故障率的主要属性。

## 3 实例分析

为了证明装配故障率的灰色最小二乘支持向量机分析方法的有效性和准确性, 对某小批量生产模式的企业装配车间进行了实例分析。

### 3.1 装配故障率的前期数据处理

首先通过 5M1E 分析得到装配故障率的所有属性, 其中材料和方法属性可以由表 1 的 AREM 符号得到, 而装配故障率的其他属性由于定量化研究比较困难, 并且会随材料和方法属性变化, 因此可由专家评分法得到, 为了减少主观判断的误差, 用三角模糊数对评分结果进行处理。为了对装配故障进行全面的分析, 分别对材料和方法属性的所有值进行评分, 由于不同专家得到的数据不同, 为了综合所有专家的意见, 经过三角模糊处理得到属性的期望值, 作为 LSSVM 的输入向量。

设三角模糊数为  $p^{(k)} = (l_{ij}^{(k)}, m_{ij}^{(k)}, u_{ij}^{(k)})$ , 经过模糊处理的评价结果转化为三角模糊数, 由

$$u = \left( \sum_{j=1}^n l_{ij}, \sum_{j=1}^n m_{ij}, \sum_{j=1}^n u_{ij} \right) \otimes \left( \sum_{i=1}^n \sum_{j=1}^n l_{ij}, \sum_{i=1}^n \sum_{j=1}^n m_{ij}, \sum_{i=1}^n \sum_{j=1}^n u_{ij} \right)^{-1} \approx \left[ \frac{\sum_{j=1}^n l_{ij}}{\sum_{i=1}^n \sum_{j=1}^n l_{ij}}, \frac{\sum_{j=1}^n m_{ij}}{\sum_{i=1}^n \sum_{j=1}^n m_{ij}}, \frac{\sum_{j=1}^n u_{ij}}{\sum_{i=1}^n \sum_{j=1}^n u_{ij}} \right], \quad (5)$$

得到模糊评价值,则  $u_i$  的期望值为

$$I(u_i) = \eta I_L(u_i) + (1 - \eta) I_R(u_i), \quad (6)$$

式中:  $I_L(u_i)$  为左期望;  $I_R(u_i)$  为右期望。

对每个材料和方法属性进行评分(由于条件有限仅对部分数据进行了分析和统计),并且对装配车间相应的真实故障率进行统计,结果如表 2 所示。

表 2 装配故障率与属性值

装配故障率 /‰	装配故障率的属性值					
	材料	方法	人	机器设备	测量	环境
25	1	1	0.315 053	0.214 756	0.202 042	0.268 436
46	1	2	0.319 568	0.210 670	0.200 640	0.269 418
64	1	3	0.320 841	0.209 068	0.200 359	0.270 035
48	2	1	0.319 206	0.208 895	0.200 700	0.273 854
86	2	2	0.320 756	0.210 020	0.200 047	0.267 850
112	2	3	0.320 956	0.209 042	0.200 320	0.272 450
66	3	1	0.320 041	0.207 548	0.200 480	0.271 140
117	3	2	0.321 011	0.208 140	0.200 154	0.271 024
156	3	3	0.325 741	0.207 486	0.201 047	0.268 957

### 3.2 基于灰色关联分析的装配故障率属性分析

将装配故障率作为参考序列,装配故障率的属性作为比较序列,为避免采用一种数据规范化方法进行分析的片面性<sup>[15]</sup>,分别建立起 3 个灰关联序:初值化变换、均值化变换和极差化变换各比较序列与参考序列的关联度。根据 3 种排列中都靠前的属性作为装配故障率的主要属性。

初值化变换

$$\begin{aligned} \gamma_{01} &= 0.717\ 643, \gamma_{02} = 0.709\ 959, \\ \gamma_{03} &= 0.607\ 125, \gamma_{04} = 0.602\ 356, \\ \gamma_{05} &= 0.604\ 401, \gamma_{06} = 0.606\ 247. \end{aligned}$$

均值化变换

$$\begin{aligned} \gamma_{01} &= 0.720\ 420, \gamma_{02} = 0.694\ 494, \\ \gamma_{03} &= 0.622\ 647, \gamma_{04} = 0.620\ 508, \\ \gamma_{05} &= 0.619\ 772, \gamma_{06} = 0.618\ 822. \end{aligned}$$

极差化变换

$$\begin{aligned} \gamma_{01} &= 0.747\ 456, \gamma_{02} = 0.734\ 701, \\ \gamma_{03} &= 0.807\ 485, \gamma_{04} = 0.615\ 990, \\ \gamma_{05} &= 0.604\ 955, \gamma_{06} = 0.675\ 586. \end{aligned}$$

根据以上结果对装配故障率的 6 个属性进行排序如表 3 所示。

表 3 灰色关联度排序

变换方法	初值化	均值化	极差化
关联度	$\gamma_{01} > \gamma_{02} > \gamma_{03} >$	$\gamma_{01} > \gamma_{02} > \gamma_{03} >$	$\gamma_{03} > \gamma_{01} > \gamma_{02} >$
排序	$\gamma_{06} > \gamma_{05} > \gamma_{04}$	$\gamma_{04} > \gamma_{05} > \gamma_{06}$	$\gamma_{06} > \gamma_{04} > \gamma_{05}$

根据表 3 将排在前 3 的属性作为装配故障率的主要属性即材料、方法和人属性,对主要属性进行深入分析可知,影响装配故障率的主要因素为装配零

件表面特征、装配工艺或工艺规程、装配工人的熟练程度等等。因此对于装配可靠性的控制可以从上述因素重点考虑。

### 3.3 LSSVM 的预测和误差分析

由灰色关联分析提取的主要属性作为 LSSVM 模型的输入向量,装配故障率作为 LSSVM 模型的输出向量,将收集到的 9 组数据同时用来训练和预测。首先应对数据进行前期处理,由于最小二乘支持向量机采用了径向基核函数,因此用遗传算法对 LSSVM 进行参数优化,设置种群规模  $M=100$ ,终止进化代数  $T=500$ ,适应度函数  $f(\gamma, \sigma)$ ,交叉概率  $P_c$  和变异概率  $P_m$ ,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,不断循环可得到最优参数。根据最优参数对数据进行预测,为了进行比较,对未提取主要属性的数据也进行了 LSSVM 建模和预测,同时用 BP 神经网络对用灰色关联分析提取的主要属性进行了训练和预测作如表 4 所示。

表 4 误差分析

装配故障率 真实值/‰	装配故障率预测值/‰		
	G-LSSVM	LSSVM	G-BP
25	25.025 35	25.278 61	25.002 30
46	46.017 63	45.281 11	47.321 41
64	64.155 77	66.563 42	63.825 70
48	47.718 34	48.125 81	48.493 44
86	86.201 68	85.678 45	86.817 21
112	111.878 70	110.696 30	111.909 10
66	66.475 32	68.326 39	65.959 86
117	116.554 20	114.971 10	117.434 90
156	155.972 90	155.078 80	155.936 90
平均绝对 误差/‰	0.194 611	1.176 494	0.381 963

由表4的误差分析结果可知,用灰色关联分析得到的G-LSSVM模型的平均绝对误差比普通LSSVM模型小得多,同时用灰色关联分析的G-LSSVM模型比G-BP模型的平均绝对误差也要小。并且G-LSSVM模型虽然输入向量减少了,但是预测精度却比全属性的预测精度高,可见由灰色关联分析后的LSSVM不仅具有输入数据少收敛速度快的特点,而且对于装配故障率的预测精度比普通LSSVM和G-BP神经网络也要高。

#### 4 结 语

由AREM方法得到的装配故障率属性缺乏对属性之间的关系进行考虑,并且提取出的属性对于装配可靠性控制缺乏实际应用,因此在AREM基础上对影响装配故障率的5M1E因素进行分析得到相关属性,为了提高装配可靠性控制的目的性,采用灰色关联分析对属性进行关联度分析得到装配故障率的主要属性,用遗传算法优化的LSSVM对主要属性进行建模并对装配故障率进行预测,与未用灰色关联分析的LSSVM模型和BP神经网络的预测结果进行比较,结果表明装配故障率的灰色LSSVM分析方法不仅预测精度高而且能为装配可靠性的重点控制提供依据,同时可以通过对主要属性深入分析得到可靠性控制更具体的措施。由于条件有限因此收集到的装配故障率缺乏一定的准确性,这也是误差所在的原因,需要采取更准确的方法收集更多的数据。

#### 参考文献:

- [1] CHUMAKOV R. An artificial neural network for fault detection in the assembly of thread-forming screws[J]. *Journal of Intelligent Manufacturing*, 2008, 19(3): 327-333.
- [2] 余锋杰,柯映林,应征. 飞机自动化对接装配系统的故障维修决策[J]. *计算机集成制造系统*, 2009, 15(9): 1823-1830.  
YU FENG-JIE, KE YING-LIN, YING ZHENG. Decision on failure maintenance for aircraft automatic join-assembly system [J]. *Computer Integrated Manufacturing Systems*, 2009, 15(9): 1823-1830.
- [3] DELVECCHIO S, D'ELIA G, MUCCHI E, et al. Advanced signal processing tools for the vibratory surveillance of assembly faults in diesel engine cold tests[J]. *Journal of Vibration and Acoustics*, 2010, 132(2): 21008-21010.
- [4] SUZUKI T, OHASHI T, ASANO M, et al. AREM shop evaluation method [J]. *CIRP Annals-Manufacturing Technology*, 2004, 53(1): 43-46.
- [5] SUZUKI T, OHASHI T, ASANO M, et al. Assembly reliability evaluation method (AREM)[C]// *The IEEE International Symposium on Assembly and Task Planning 2001*, May 28-29, 2001, Fukuoka, Japan. [S. l.]: IEEE, 2002: 294-299.
- [6] VAPNIK V N. *The nature of statistical learning theory*[M]. New York: Spring-Verlag, 1999.
- [7] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers [J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [8] 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.  
WU DE-HUI. A prediction model for surface roughness in milling based on least square support vector machine [J]. *China Mechanical Engineering*, 2007, 18(7): 838-841.
- [9] JUHOS I, MAKRA L, TÓTH B. Forecasting of traffic origin NO and NO<sub>2</sub> concentrations by support vector machines and neural networks using principal component analysis [J]. *Simulation Modelling Practice and Theory*, 2008, 16(9): 1488-1502.
- [10] 郭辉,刘贺平,王玲. 最小二乘支持向量机参数选择方法及其应用研究. [J]. *系统仿真学报*. 2006, 18(7): 2033-2036.  
GUO HUI, LIU HE-PING, WANG LING. Method for selecting parameters of least squares support vector machines and application [J]. *Journal of System Simulation*, 2006, 18(7): 2033-2036.
- [11] CABALLERO J, FERNANDEZ L, GARRIGA M, et al. Proteomic study of ghrelin receptor function variations upon mutations using amino acid sequence autocorrelation vectors and genetic algorithm-based least square support vector machines [J]. *Journal of Molecular Graphics and Modelling*, 2007, 26(1): 166-178.
- [12] WU C, TZENG G, GOO Y, et al. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy [J]. *Expert Systems with Applications*, 2007, 32(2): 397-408.
- [13] 李锋,汤宝平,刘文艺. 遗传算法优化最小二乘支持向量机的故障诊断[J]. *重庆大学学报*, 2010, 33(12): 14-20.  
LI FENG, TANG BAO-PING, LIU WEN-YI. Fault diagnosis based on least square support vector machine optimized by genetic algorithm [J]. *Journal of Chongqing University*, 2010, 33(12): 14-20.
- [14] 李中凯,冯毅雄,谭建荣,等. 基于灰色系统理论的质量屋中动态需求的分析与预测[J]. *计算机集成制造系统*, 2009, 15(11): 2272-2279.  
LI ZHONG-KAI, FENG YI-XIONG, TAN JIAN-RONG, et al. Analysis and prediction for dynamic requirements in house of quality based on grey theory [J]. *Computer Integrated Manufacturing Systems*, 2009, 15(11): 2272-2279.
- [15] 贾振元,马建伟,王福吉,等. 多零件几何要素影响下的装配产品特性预测方法[J]. *机械工程学报*, 2009, 45(7): 168-173.  
JIA ZHEN-YUAN, MA JIAN-WEI, WANG FU-JI, et al. Characteristics forecasting method of assembled product based on multiple part geometric elements [J]. *Journal of Mechanical Engineering*, 2009, 45(7): 168-173.