

文章编号:1001-9081(2010)06-1648-03

基于 LSSVRM 和 SA 算法的软件失效模型

晁冰,徐仁佐

(武汉大学 计算机工程国家重点实验室, 武汉 430072)

(qpdee@163.com)

摘要:针对软件失效数据的特点,提出将最小二乘支持向量回归机(LSSVRM)用于软件可靠性失效模型的构造。利用LSSVRM对小样本失效数据构造拟合模型,使用模拟退火(SA)算法对LSSVRM的参数进行优化,得到基于SA的LSSVRM,实现对拟合模型的进一步改进。通过与常用的NHPP类软件失效模型对比表明,利用LSSVRM和SA算法可以构造拟合精度更高的失效模型,并可简化对模型参数的调整

关键词:软件失效模型;软件可靠性;最小二乘支持向量回归机;模拟退火算法

中图分类号: TP311 **文献标志码:** A

Failure model of software reliability based on LSSVRM and simulated annealing algorithm

CHAO Bing, XU Ren-zuo

(State Key Laboratory of Software Engineering, Wuhan University, Wuhan Hubei 430072, China)

Abstract: According to the characteristics of the software failure data, the Least Square Support Vector Regression Machine (LSSVRM) was applied to build the software failure model. The LSSVRM was used to construct a fitting model according to the failure data that is a small sample and the parameters in LSSVRM were optimized by the simulated annealing algorithm to get the SA-LSSVRM and further improve the fitting model. Compared with the failure mode of NonHomogeneous Poisson Process (NHPP), the failure model based on the LSSVRM and SA algorithm has the higher fitting precision and the adjustments of the model parameters could be simplified.

Key words: software failure model; software reliability; Least Square Support Vector Regression Machine (LSSVRM); Simulated Annealing (SA) algorithm

0 引言

软件可靠性指标的评估与软件失效函数有着密切的联系。软件失效函数是软件故障数据的统计模型,目前常用的失效模型是基于时间单位的模型。这类模型以时间为基准,研究软件的可靠性特征随时间变化的规律。常用的G-O模型、S-W模型、S-曲线模型及指数模型等都是此类模型的代表。根据此类模型对数据的要求,又可分为失效时间间隔模型和失效计数模型两个子类^[1]。前者的分析方法建立在以失效时间间隔服从特定概率分布的基础上,后者的分析方法大多建立在Poisson过程理论的基础上。在应用时,对失效数据集进行适当转换,就可以同时应用这两类模型对其进行分析。这类模型的缺点是其假设的前提条件较高,很难完全满足,不具有普适性。此外,模型中含有未知参数,在故障数据较少时,对模型参数的估计会有较大的误差,这都影响了模型的准确性和易用性。但是这类模型数量众多,选择余地大,所以在实际应用中大多还是选用此类模型。另外,诸如混沌模型^[2]、时间序列模型、灰色模型^[3]及组合模型^[4]等虽然建模方法不同,但也都是以时间尺度作为参考,故也可纳入此类模型中。其他可用于软件可靠性分析的模型还包括基于输入数据的模型、基于错误数的模型、基于故障树的模型、基于神经网络的模型、基于测试覆盖的模型^[5]和基于体系结构的模

型^[6]等。这些模型有的无法反映软件可靠性与时间尺度的关系,有的需要大量工程数据的支持,有的在结构实现及应用上比较繁琐,因此在实用中都受到较大限制。

最小二乘支持向量回归机(Least Square Support Vector Regression Machine, LSSVRM)^[7-8]是适用于小样本数据集的拟合方法,使用LSSVRM构造的失效模型能够准确地反映软件失效与时间尺度的关系,不需估计参数,具有较强的泛化能力,适用于多数的失效数据集。另外,随着软件故障数据的逐渐积累,只需适当调整模型参数即可对更新后的故障数据集进行拟合,而使用一般的失效模型,需要重新估计模型参数值,当模型无法满足精度要求时,则需要另外更换模型。

1 基于 LSSVRM 构造失效模型

软件失效数据集 T 可表示为:

$$T = \{(t_i, m_i) \mid i = 1, \dots, n\}$$

其中: t_i 表示第 i 次失效发生的时间值, m_i 表示软件在 t_i 时总的失效数。

规定 t_i 为失效模型的输入, m_i 为失效模型的输出。根据LSSVRM算法,通过求解下述最优化问题得到构造失效模型所需的参数。

$$\lim_{\omega, b, \xi} \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^n \xi_i^2$$
$$\text{s. t. } \omega^T \varphi(t_i) + b + \xi_i - m_i = 0; i = 1, \dots, n$$

收稿日期:2009-12-10;修回日期:2010-02-21。

作者简介:晁冰(1972-),男,安徽合肥人,讲师,硕士,主要研究方向:软件可靠性分析、软件测试;徐仁佐(1949-),男,湖北武汉人,教授,博士生导师,主要研究方向:软件工程、系统可靠性工程。

其中: $\varphi(t_i)$ 是向高维特征空间所做的映射变换, $\xi = (\xi_1, \dots, \xi_n)$ 为降低误差所引入的松弛变量, C 为惩罚参数。

采用拉格朗日乘子法求解上述最优化问题:

$$L(\omega, b, \xi, \alpha) = \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i (\omega^T \varphi(t_i) + b + \xi_i - m_i)$$

根据 KKT 优化条件 $\partial L / \partial \omega_i = 0, \partial L / \partial b = 0, \partial L / \partial \xi_i = 0, \partial L / \partial \alpha_i = 0$, 求得拉格朗日乘子 α_i 以及 b 的值。定义核函数 $K(t_i, t) = \varphi(t_i) \cdot \varphi(t)$, 最后可得失效模型:

$$m(t) = \sum_{i=1}^n \alpha_i K(t_i, t) + b$$

引入软件失效数据集 Naval Tactical Data Systems (NTDS)、Tandem Computers Software Data Project (TCSDP) 和 AT&T System T Project (ATTSTP)^[9]。利用 LSSVRM 构造失效模型, 核函数为 Gaussian 径向基函数 $\exp(-\|x - x'\|^2 / \sigma^2)$, 拟合结果分别如图 1 所示。

2 与其他失效模型的比较

软件失效模型之间的比较有许多不同的参考标准, 这需

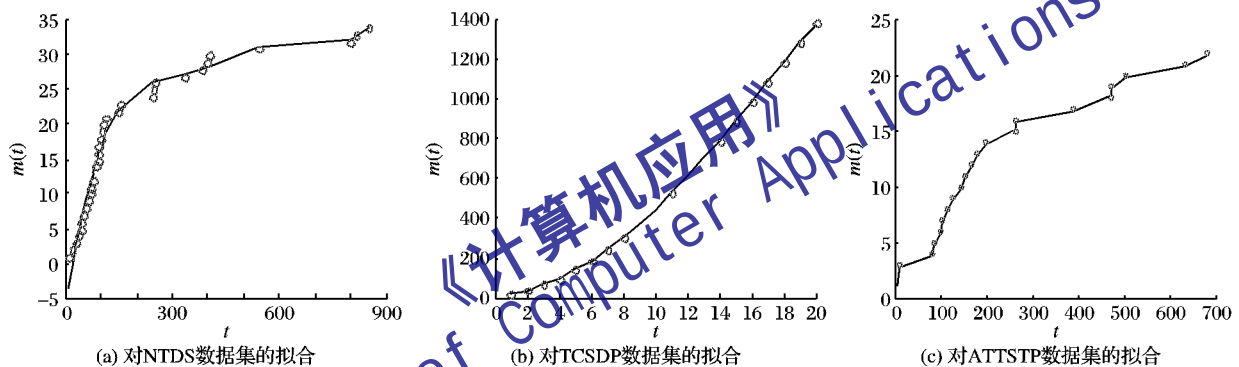


图1 由 LSSVRM 构造的失效模型对故障数据的拟合

表1 不同软件失效模型的 SSE 比较

模型名称	模型表达式	SSE		
		NTDS	TCSDP	ATTSTP
G-O	$m(t) = a(1 - e^{-bt})$	208	104 870	28.97
Delay S-shaped	$m(t) = a(1 - (1 + bt)e^{-bt})$	174	35 624	31.50
Inflection S-shaped	$m(t) = a(1 - e^{-bt}) / (1 + \beta e^{-bt})$	187	30 015	28.97
Yamada exponential	$m(t) = a(1 - \exp(\gamma \alpha (1 - e^{(-\beta t)})$	208	110 351	29.02
Yamada Rayleigh	$m(t) = a(1 - \exp(\gamma \alpha (1 - \exp(-\beta t^2/2)))$	174	43 626	41.27
Imperfect debugging model(1)	$m(t) = ab(e^{\alpha t} - e^{-bt}) / (\alpha + b)$	1 034	77 506	31.35
Imperfect debugging model(2)	$m(t) = a(1 - e^{-bt})(1 - \alpha/b) + \alpha at$	2 125	91 579	30.71
PNZ	$m(t) = (a(1 - e^{-bt})(1 - \alpha/b) + \alpha at) / (1 + \beta e^{-bt})$	71	31 595	22.29
基于 LSSVRM 的模型	$\sum_{i=1}^n \alpha_i K(t_i, t) + b$	67	536	14.90

3 利用模拟退火算法优化模型参数

观察 LSSVRM 可以发现, 当训练集和松弛变量确定后, 参数 C 就决定了最优化问题的解。为了能获得符合要求的拟合精度, 需要对 C 的值进行反复调整。另外, 在 LSSVRM 中, 核函数是很重要的组成部分, 而核函数参数的调整对优化问题的求解也有很大影响。由于这些参数设置的不确定性, 导致由 LSSVRM 所构造的失效模型的拟合精度无法得到保证。为克服参数调整的不确定性, 引入模拟退火 (Simulated

要根据模型的应用环境及目的进行合理地选择。为便于与基于 LSSVRM 所构建的失效模型进行比较, 以时间模型中常用的非齐次泊松过程 (Nonhomogeneous Poisson Process, NHPP) 类失效模型为参考, 仍然以 IBM、TCSDP 和 ATTNMS 数据集为样本, 对这些模型的拟合效果进行定量分析。以常用的方差和 (Sum of Squared Errors, SSE) 作为比较标准, 其计算公式为:

$$SSE = \sum_{i=1}^n [y_i - m(t_i)]^2$$

其中: y_i 是在第 i 个时间间隔 t_i 内总的失效数的观察值, $m(t_i)$ 是在第 i 个时间间隔 t_i 内总的失效数的估计值。

使用不同的失效模型分别对 3 类故障数据拟合后的 SSE 值见表 1。

从表 1 中可以看出, 基于 LSSVRM 构造的失效模型对失效数据拟合的 SSE 值明显小于其他模型。另外, 从软件可靠性的一般计算公式 $R(t) = \exp(-m(t))$ 中还可看出, 随着对失效数据拟合误差的减小, 即 $m(t)$ 计算精度的增加, 软件可靠性的计算精度也相应得到提高。

Annealing, SA) 算法^[10]对 LSSVRM 中的参数进行自动寻优, 以获得优化的 SA-LSSVRM。

使用 Gaussian 核函数的 LSSVRM, 需要调整 C 和 σ 两个参数, 其中 σ 为核函数的参数。利用 SA 算法对这两个参数进行优化, 选择 SSE 作为目标函数, 参数 C 和 σ 作为其决策变量。在通过 SA 算法求得参数 C 和 σ 的最优值同时, 也获得了最佳的失效拟合模型。

设置算法运行的初始条件, 两个参数的取值范围为 $100 \leq C \leq 10\,000, 1 \leq \sigma \leq 1\,000$ 。退火计算模型采用

Boltzmann 机,退火控制间隔为 100,退火初始温度为 10 000,温度更新函数采用指数函数。以目标函数值的最终变化小于 $1\text{E}-6$ 作为算法的终止条件。

参数 C 和 σ 优化前后的值见表 2。利用参数优化后的 LSSVRM 重新构造失效模型并对数据集 NTDS、TCDSP 和 ATTSTP 进行拟合,结果如图 2 所示。模型在优化前及优化后对 3 类数据集拟合的 SSE 值见表 3。由表 3 可以看出,使用 SA 算法优化后的 LSSVRM 所构造的失效模型可进一步降

低对故障数据的拟合误差,从而为软件可靠性的评估提供更优化的分析模型。

表 2 参数 C 和 σ 优化前后的取值

数据集	基于 LSSVRM 的模型		基于 SA-LSSVRM 的模型	
	C	σ	C	σ
NTDS	4 000	3.750	6 827.057	1.004
TCDSP	5 000	6.000	9 286.324	4.190
ATTSTP	9 000	2.500	4 047.917	86.686

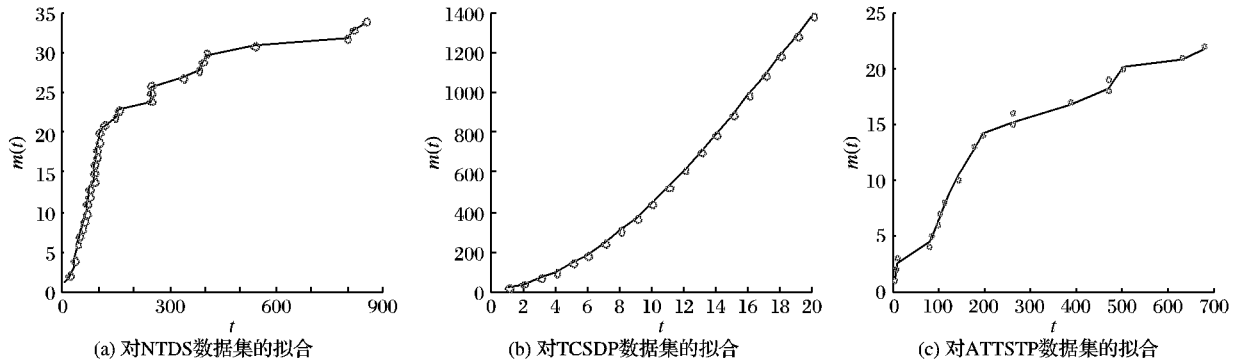


图 2 由 SA-LSSVRM 构造的 $m(t)$ 对故障数据的拟合情况

表 3 优化前后失效模型拟合故障数据的 SSE 比较

模型名称	SSE		
	NTDS	TCDSP	ATTSTP
基于 LSSVRM 的模型	67	536	14.9
基于 SA-LSSVRM 的模型	2	17	8.5

4 结语

本文提出了基于 LSSVRM 构造软件失效模型的方法。通过对比可以看出,由 LSSVRM 构造的失效模型对失效数据的拟合效果要优于常用的 NHPP 类失效模型。但是在利用 LSSVRM 构造失效模型时,存在模型参数需人工调整的问题,因此引入 SA 算法对 LSSVRM 中的参数进行自动优化,由此所构造的失效模型,可以做到对故障数据的最佳逼近,进一步提高软件可靠性指标的估算精度。

参考文献:

- [1] 徐仁佐,谢旻,郑人杰. 软件可靠性模型及应用[M]. 北京:清华大学出版社,1994:46-123.
- [2] 邹丰忠,李传湘. 软件可靠性混沌模型[J]. 计算机学报,2001,

24(3):281-291.

- [3] 李海峰,陆民燕,王智新. 基于灰色系统理论的软件可靠性综合评价框架[J]. 北京航空航天大学学报,2008,34(11):1261-1265.
- [4] 滕云龙,师尧兵,康荣雷. 软件可靠性组合预测模型研究[J]. 计算机应用,2008,28(12):3092-3094.
- [5] 徐仁佐. 软件可靠性工程[M]. 北京:清华大学出版社,2007:181-204.
- [6] 普杰,罗云峰,责可荣. 基于体系结构的软件可靠性模型综述[J]. 计算机与数字工程,2009,37(4):59-63.
- [7] 邓乃扬,田英杰. 支持向量机:理论、算法与拓展[M]. 北京:科学出版社,2009:182-188.
- [8] 宋海鹰,桂卫华,阳春华. 稀疏最小二乘支持向量机及其应用研究[J]. 信息与控制,2008,37(3):334-338.
- [9] PHAM H. System software reliability [M]. Berlin: Springer-Verlag,2006:137-149.
- [10] 康立山,冯康. 非数值并行算法(第一册):模拟退火算法[M]. 北京:科学出版社,1994:29-55.

(上接第 1637 页)

3 结语

YH-RTSIM 软件很好地解决了 YHSIM 仿真时帧时间偶尔漂移的问题,实时性更好(与 YHSIM 相比较),已经在一些单位得到了应用,用户反应良好。但是随着航空航天应用的发展,会出现一些超级大的模型,因此为了保证半实物仿真的实时性,还需要进一步研究实时操作系统和多处理机、多核环境下的并行处理技术,使得模型计算和 I/O 通信能获得更大的并行度。

参考文献:

- [1] 姚益平,蒋志文,鄢来斌,等. YH-AStar 高性能实时仿真平台[J]. 计算机仿真,2003,20(z1):219-220.
- [2] 夏红伟,凌明祥,曾庆双,等. 基于 RTX 的卫星姿轨控系统地面

实时仿真系统[J]. 计算机仿真,2006,23(9):40-42.

- [3] 刘晓川,樊子明. Windows 2000 (XP) + RTX 的实时性分析与测试[J]. 舰船电子工程,2007,27(6):135-138.
- [4] 王礼,凌明祥,曾庆双,等. RTX 在激光制导炸弹半实物仿真中的应用[J]. 红外与激光工程,2006,35(1):78-81.
- [5] 赵正伟,杜承烈. 基于实时网和 RTX 的分布虚拟实验系统研究[J]. 计算机仿真,2007,24(10):209-211.
- [6] 王普,张蕾,都立伟. 基于 RTX 的全软件数控系统的研究[J]. 燕山大学学报,2007,31(6):513-516.
- [7] 宫厚良,陈曾汉. 基于 RTX 和 LabVIEW 的多任务实时测控系统[J]. 计算机应用,2007,27(6):1551-1553.
- [8] 张广春,符文星,闫杰. RTX 在负载模拟器控制软件中的应用[J]. 计算机仿真,2009,26(8):287-290.