

组合核相关向量机的软件可靠性预测模型

梁宏涛^{1,2} 徐建良¹ 许 可³

(中国海洋大学信息科学与工程学院 青岛 266100)¹ (青岛工学院信息工程学院 青岛 266300)²

(北京航空航天大学计算机学院 北京 100191)³

摘 要 可靠性作为衡量软件质量的一种重要特性,对软件管理具有重要的意义。针对单一核函数的缺陷,提出一种组合核函数相关向量机的软件可靠性预测模型。首先对当前软件可靠性研究现状进行分析,然后采用组合核函数相关向量机对训练集进行学习和建模,最后通过具体实例对模型的预测性能进行分析。结果表明,本模型获得了理想的软件可靠性预测结果,且其预测性能要优于单一核函数模型,在软件可靠性预测中有重要的应用价值。

关键词 软件可靠性,预测模型,组合核函数,相关向量机

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.11.050

Software Reliability Prediction Model Based on Hybrid Kernels RVM

LIANG Hong-tao^{1,2} XU Jian-liang¹ XU Ke³

(School of Information Science and Engineering, Ocean University of China, Qingdao 266100, China)¹

(School of Information Science and Engineering, Qingdao University of Technology, Qingdao 266300, China)²

(School of Computer Science and Engineering, Beihang University, Beijing 100191, China)³

Abstract Reliability as an important measure of software quality, is of great significance to software management. A software reliability prediction model by using hybrid kernel function relevance vector machine was proposed in this paper to solve defects of single kernel function. Firstly, research status of software reliability is analyzed. Secondly, training samples are studied and modeled by using hybrid kernel function relevance vector machine. Finally, the prediction performance is analyzed with some specific examples. The results show that the proposed model can obtain ideal software reliability prediction results, and the performance is better than the single kernel function models, so it has important practical application value in software reliability prediction.

Keywords Software reliability, Prediction model, Hybrid kernels function, Relevance vector machine

1 引言

随着软件规模越来越庞大,软件质量管理难度相应增加,软件质量控制就显得尤为重要^[1]。可靠性是评价软件质量的重要指标之一,而软件可靠性预测根据软件失效数据,采用统计理论作为工具建立相应的预测模型,然后根据预测模型对软件可靠性进行评估,因此软件可靠性预测具有重要的实践意义,如何建立高精度的软件可靠性预测模型引起了人们的广泛关注^[2-4]。

针对软件可靠性预测问题,从 20 世纪 70 年代起,许多学者和专家投入了大量的时间和精力,对其进行了深入研究,提出了许多有效的对软件可靠性进行预测的模型^[5]。软件可靠性模型分为分析模型和数据驱动模型两类,其中分析模型主要有 Markov 模型等,它们需要做一些假设而且它们没有自适应性,在实际应用中有一些假设难以得到满足,因此该类分析模型的应用范围受到了限制^[6]。由于软件自身复杂性以及失效行为的不确定性,目前采用数据驱动模型对软件可靠性进行分析,它们采用机器学习算法对软件可靠性数据进行分析,挖掘其蕴含的规律,实现软件可靠性预测。数据驱动模型

主要有:ARIMA、神经网络、支持向量机等^[8-11],它们具有很好的自适应性与自学习能力,因此也成为了软件可靠性预测的主要研究方向^[7]。在实际应用中,数据驱动模型也存在局限,如神经网络易得到“过拟合”结果、支持向量机训练费时等。相关向量机(Extreme Learning Machine, RVM)是一种稀疏贝叶斯学习模型,不存在神经网络“过拟合”缺陷,且训练速度快,具有支持向量机的泛化性能,被学者引入到软件可靠性预测中,得到了更优的预测结果^[12]。

RVM 核函数及参数直接影响软件可靠性预测性能,单一核函数存在不同的缺陷,组合核函数却可以利用局部核函数和全局核函数的优点克服各自的不足。为了提高软件可靠性预测精度,本文提出一种组合核函数相关向量机的软件可靠性预测模型(HK-RVM)。实验结果表明,本文模型获得了理想的软件可靠性预测结果,而且其预测性能要优于其它模型,具有重要的实际应用价值。

2 HK-RVM

2.1 相关向量机

设训练集为 $\{x_i, t_i\}_{i=1}^N, t = [t_1, t_2, \dots, t_N]^T$ 为期望输出,

到稿日期:2015-10-06 返修日期:2016-01-25 本文受山东省高等学校科技计划项目(J13LN77, J13LN78)资助。

梁宏涛(1979—),男,博士生,副教授,主要研究领域为软件体系结构、数据挖掘、机器博弈;徐建良(1969—),男,教授,博士生导师,主要研究领域为数据挖掘、语义网络;许可(1971—),男,教授,博士生导师,主要研究领域为社交网络、空间计算。

$x_i \in R^d, t_i \in R$, 期望输出含有噪声 ϵ_i , 且 $\epsilon_i \sim N(0, \sigma^2)$, 那么 RVM 的输出为:

$$\begin{cases} f(x_i) = \sum_{i=1}^N w_i k(x, x_i) + w_0 \\ t_i = f(x_i) + \epsilon_i \end{cases} \quad (1)$$

其中, $k(x, x_i)$ 为核函数; $w = (w_0, w_1, \dots, w_N)^T$.

由于 $p(t_i | x) = N(t_i | f(x_i), \sigma^2)$, 那么训练集的似然函数为:

$$p(t | w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \|t - \Phi w\|^2\right\} \quad (2)$$

其中, $\Phi = [\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)]^T$ 为一个 $N \times (N+1)$ 的矩阵, $\Phi(x_i) = (1, k(x_i, x_1), k(x_i, x_2), \dots, k(x_i, x_N))^T$.

设 w_i 服从均值为 0、方差为 a_i^{-1} 的先验高斯正态分布, 那么有:

$$p(w | a) = \prod_{i=0}^N N(w_i | 0, a_i^{-1}) \quad (3)$$

其中, $a = (a_0, a_1, \dots, a_N)^T$ 为 w_i 的超参数。

在先验分布和训练集已知的情况下, w 的后验分布为:

$$p(w | t, a, \sigma^2) = \frac{p(t | w, \sigma^2) p(w | a)}{p(t | a, \sigma^2)} = N(w | \mu, \Sigma) \quad (4)$$

其中, $\mu = \sigma^{-2} \Sigma \Phi^T t$, $\Sigma = (A + \sigma^{-2} \Phi^T \Phi)^{-1}$, $A = \text{diag}(a_0, a_1, \dots, a_N)$.

为使 w 的后验分布最大化, 首先要优化 a 和 σ^2 , 它们的边缘似然函数为:

$$\begin{aligned} p(t | a, \sigma^2) &= \int p(t | w, \sigma^2) p(w | a) dw \\ &= N(t | 0, \sigma^2 I + \Phi A^{-1} \Phi^T) \\ &= (2\pi)^{-\frac{N}{2}} |\sigma^2 I + \Phi A^{-1} \Phi^T|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} t^T \right. \\ &\quad \left. (\sigma^2 I + \Phi A^{-1} \Phi^T)^{-1} t\right) \end{aligned} \quad (5)$$

采用该最大化似然函数不断重复迭代, 得到 a_{MP} 和 σ_{MP}^2 的值, a 中一些元素值均向无穷大变化, 使得 $p(w_i | t, a, \sigma^2)$ 向 0 趋近, 一些非 0 权值的训练样本为相关向量。

给定新的输入样本 x_* , 其估计值 t_* 的预测分布为:

$$\begin{aligned} p(t_* | t) &= \int p(t_* | w, \sigma_{MP}^2) p(w | t, a_{MP}, \sigma_{MP}^2) dw \\ &= N(t_* | y_*, \sigma_*^2) \end{aligned} \quad (6)$$

其中, $y_* = \mu^T \Phi(x_*)$, $\sigma_*^2 = \sigma_{MP}^2 + \Phi(x_*)^T \Sigma \Phi(x_*)$.

2.2 组合核函数的构造

在基于 RVM 的软件可靠性预测的建模过程中, 核函数及参数的选择直接决定了软件可靠性预测结果的好坏^[13], 通常 RVM 主要采用单核函数, 其存在映射方式单一、无法准确描述软件可靠性变化特点等不足, 而组合核函数可以将不同核函数进行组合, 映射能力更强, 组合核函数表示形式为:

$$k(x_i, x_j) = \sum_{m=1}^M c_m k_m(x_i, x_j) \quad (7)$$

其中, c_m 为权重系数, M 表示核函数数量。

核函数分为局部核函数和全局核函数两类, 它们各有不足, 本文利用两者优势同时去避免它们的不足, 将二者进行线性组合, 建立一种组合核函数的 RVM(HK-RVM), 组合核函数为:

$$k(x_i, x_j) = m k_{local}(x_i, x_j) + (1-m) k_{global}(x_i, x_j), 0 \leq m \leq 1 \quad (8)$$

采用高斯核函数和多项式核函数进行组合, 那么组合核函数具体为:

$$k(x_i, x_j) = m \exp(-\gamma \|x_i - x_j\|^2) + (1-m)(x_i^T x_j + 1)^d, 0 \leq m \leq 1, \gamma > 0 \quad (9)$$

组合核函数则兼具了高斯核函数和多项式核函数的优点,

学习、泛化能力均得到改善, 组合核函数的 RVM 输出变为:

$$\begin{cases} y(x; w) = \sum_{i=1}^N w_i [m k_{local}(x, x_i) + (1-m) k_{global}(x, x_i)] + w_0 \\ t_i = y(x_i; w) + \epsilon_i \end{cases} \quad (10)$$

为了测试 HK-RVM 预测性能, 采用标准测试函数进行模拟仿真实验, 测试函数定义如下:

$$y = \sin|x|/|x| + 0.2x, x \in [-10, 10] \quad (11)$$

随机产生 500 个数据点, 其中 300 个数据点作为训练集训练 HK-RVM, 其它数据点对 HK-RVM 的有效性进行测试, 并与单核 RVM 进行对照分析, 性能对比结果如表 1 所列。从表 1 可知, 相对于单核 RVM, KF-RVM 的预测误差大幅度降低, 有效提高了预测精度, 而且向量数目也减少, 预测性得到了显著改变, 此结果证明了组合核函数的优越性。

表 1 预测模型的性能对比

模型	误差(%)	向量数目(个)
RBF 核	14.878	42
多项式核	26.579	66
组合核	4.610	37

3 HK-RVM 在软件可靠性预测中的应用

3.1 数据来源

为了分析 HK-RVM 的软件可靠性预测的有效性, 选择一个大型通信软件系统的失效数据进行仿真实验, 样本数据具体如图 1 所示, 前 20 个样本数据作为训练集, 其它作为测试集。为防止原始数据值波动太大对 RVM 训练过程造成影响, 所以对其进行归一化处理, 具体为:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)}, i = 1, 2, \dots, n \quad (12)$$

其中, x_i 与 x_i' 分别为原始和归一化后的软件失效数据^[14]。

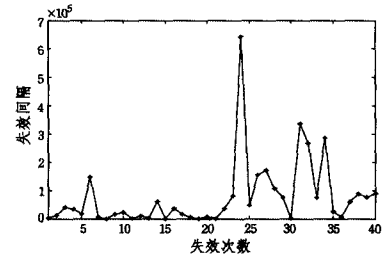


图 1 软件失效数据

3.2 结果与分析

3.2.1 HK-RVM 的拟合效果

前 20 个作为训练集对, HK-RVM 得到的拟合结果如图 2 所示。从图 2 可知, 拟合结果表示 HK-RVM 可以很好地描述软件可靠性变化特点, 实验结果初步证明了将 HK-RVM 用于软件可靠性预测的有效性。

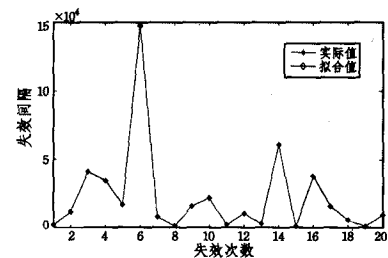


图 2 HK-RVM 的拟合结果

3.2.2 HK-RVM 的泛化效果

软件可靠性预测的建模目的主要是对软件可靠性将来变

化的趋势进行分析,而拟合结果只能描述软件系统过去的可靠性状态,使得拟合结果没有什么实际应用价值,为此采用测试样本集对 HK-RVM 的泛化效果进行分析,实验结果如图 3 所示。从图 3 可知, HK-RVM 的软件可靠性预测精度高,预测误差相当小,预测值与真实值很接近,具有较好的泛化推广性能,实验结果再一次证明了将 HK-RVM 用于软件可靠性预测的有效性,预测结果是可靠的。

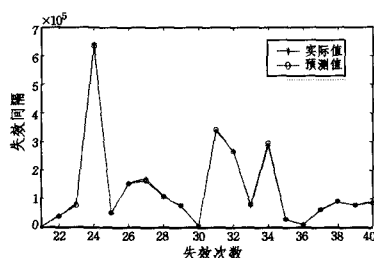


图 3 HK-RVM 的预测结果

3.2.3 与模型的性能对比

选择 RBF 核函数 RVM(RBF-RVM)、多项式核函数 RVM(poly-RVM)和文献[15]模型进行对照实验,并选择平均相对误差(MAPE)作为评价指标:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (13)$$

其中, x_i 和 $\hat{x}_i$ 分别为软件可靠性的真实值和预测值, n 表示样本数量。

HK-RVM 与对比软件可靠性预测模型的 MAPE 值如表 2 所列,从中可得到如下结论:

(1) 相对于 poly-RVM 和 RBF-RVM, HK-RVM 的 MAPE 值更低,这主要是由于组合核函数避免了单核函数映射特性单一的缺陷,可以从不同角度对软件可靠性变化特性进行刻画,学习能力和泛化性能都得到了不同程度的改善,有效提高了软件可靠性的预测精度。

(2) 相对于文献[15]的软件可靠性预测模型, HK-RVM 的 MAPE 也相对更低,获得了更加理想的软件可靠性预测结果。

表 2 MAPE 的统计结果

模型	训练集的 MAPE(%)	预测集的 MAPE(%)
poly-RVM	6.84	10.94
RBF-RVM	6.37	11.26
文献[15]	3.87	6.25
HK-RVM	2.95	5.31

结束语 随着软件应用范围的不断扩大,软件失效带来的经济损失越来越严重。通过研究软件可靠性预测可以降低软件维护的费用,对提高软件质量具有重要的实际意义。针对 RVM 的核函数构建问题以及单核 RVM 的不足,提出一种组合核函数相关向量机的软件可靠性预测模型,仿真实验结果表明,该模型提高了软件可靠性预测精度,预测结果的可靠性要优于其它模型,可以为软件研发人员提供有用信息,具有广泛的应用前景。

参考文献

[1] Zheng Yan-yan, Guo Wei, Xu Ren-zuo. Overview of software reliability engineering [J]. Computer Science, 2009, 36(2): 20-25 (in Chinese)
郑艳艳,郭伟,徐仁佐. 软件可靠性工程学综述[J]. 计算机科学, 2009, 36(2): 20-25
[2] Yang B, Li X, Xie M, et al. A generic data-driven software reliability

model with model mining technique[J]. Reliability Engineering and System Safety, 2010, 95(6): 671-678
[3] Xiao X, Dohi T. Wavelet Shrinkage Estimation for Non-Homogeneous Poisson Process Based Software Reliability Models [J]. IEEE Trans on Reliability, 2013, 62(1): 211-225
[4] Jia Zhi-yu, Kang Rui. Forecasting method of software reliability based on ARIMA model [J]. Computer Engineering and Application, 2008, 44(3): 17-19 (in Chinese)
贾治宇,康锐. 软件可靠性预测的 ARIMA 方法研究[J]. 计算机工程与应用, 2008, 44(3): 17-19
[5] Amin A, Grunske L, Colmn A. An approach to software reliability prediction based on time series modeling [J]. The Journal of Systems and Software, 2013, 86(7): 1923-1932
[6] Zhang Ke, Zhang De-ping, Wang Shuai. Chaotic neural network model for software reliability [J]. Computer Science, 2014, 44(4): 172-1177 (in Chinese)
张柯,张德平,汪帅. 软件可靠性混沌神经网络模型[J]. 计算机科学, 2014, 44(4): 172-1177
[7] Li H, Zeng M, Lu M, et al. Adaboosting-based dynamic weighted combination of software reliability growth models [J]. Quality and Reliability Engineering International, 2012, 28(1): 67-84
[8] Mohanty R, Ravi V, Patra M R. Hybrid intelligent systems for predicting software reliability [J]. Applied Soft Computing, 2013, 13(1): 189-200
[9] Park J, Baik J. Improving software reliability prediction through multi-criteria based on dynamic model selection and combination [J]. The Journal of Systems and Software, 2015, 101(1): 236-244
[10] Roy P, Mahapatra G S, Rani P, et al. Robust feed forward and recurrent neural network based dynamic weighted combination models for software reliability prediction [J]. Applied Soft Computing, 2014, 22: 629-637
[11] Zhang Ting-ting, Zhang De-ping, Liu Guo-qiang. Enhanced Bayesian combinational model for software reliability prediction [J]. Application Research of Computers, 2015, 33(9): 5-10 (in Chinese)
张婷婷,张德平,刘国强. 软件可靠性预测的增强贝叶斯组合模型[J]. 计算机应用研究, 2015, 33(9): 5-10
[12] Lou Jun-gang, Jiang Jian-hui, Shen Zhang-guo, et al. Software reliability prediction modeling with relevance vector machine [J]. Journal of Computer Research and Development, 2013, 50(7): 1542-1550 (in Chinese)
楼俊钢,江建慧,沈张果,等. 软件可靠性预测的相关向量机模型[J]. 计算机研究与发展, 2013, 50(7): 1542-1550
[13] Ye Qiao-lin, Ye Ning, Zhang Xun-hua. Extreme Decomposition based mixtures of Kernels and its improvement [J]. Pattern Recognition and Artificial Intelligence, 2009, 8(3): 366-373 (in Chinese)
业巧林,业宁,张训华. 基于极分解下的混合核函数及改进[J]. 模式识别与人工智能, 2009, 8(3): 366-373
[14] Qiao Hui, Zhou Yan-zhou, Shao Nan, et al. Research of software reliability prediction model based on AGA-LVQ [J]. Computer Science, 2013, 40(1): 179-182 (in Chinese)
乔辉,周雁舟,邵楠,等. 基于 AGA-LVQ 神经网络的软件可靠性预测模型研究[J]. 计算机科学, 2013, 40(1): 179-182
[15] Jia Ji-ting. Software reliability prediction based on PSOABC-SVM model [J]. Computer Systems & Applications, 2014, 23(7): 161-164 (in Chinese)
贾冀婷. 基于 PSOABC-SVM 的软件可靠性预测模型[J]. 计算机系统应用, 2014, 23(7): 161-164