基于和声搜索优化支持向量回归的软件可靠性预测

汪顺和

（安徽开放大学 学习资源中心，合肥 230022）

摘要：为了提高软件可靠性预测的精确率，采用支持向量机理论对软件可靠性建模，并对支持向量回归中参数优化难的问题，使用和声搜索算法优化支持向量回归中的参数，提出了一种基于和声搜索优化支持向量回归的软件可靠性预测模型。使用两组真实数据对提出的模型进行实验，并将实验结果与经典软件可靠性模型（G-O 模型和 M-O 模型）进行比较，结果表明：基于和声搜索优化支持向量回归的软件可靠性预测模型的预测精度更高。

关键词：软件可靠性；和声搜索；支持向量机

中图分类号：TP301.6　　**文献标识码**：A　　**文章编号**：2097-0625(2022)03-0082-05

一、引言

软件可靠性（Software Reliability）是软件产品在规定的条件下和规定的时间区间完成规定功能的能力。规定的条件是指直接与软件运行相关的使用该软件的计算机系统的状态和软件的输入条件，或统称为软件运行时的外部输入条件；规定的时间区间是指软件的实际运行时间区间；规定功能是指为提供给定的服务，软件产品所必须具备的功能。软件可靠性不但与软件存在的缺陷和（或）差错有关，而且与系统输入和系统使用有关。

软件可靠性模型就是根据已发生的软件失效数据，通过统计方法计算出软件的可靠性估计值或预测值[1]。它是评估和预测软件可靠性的重要工具，对于软件可靠性的评估起着核心作用，从而对软件质量的保证有着重要的意义，也为改善软件质量提供了指南。张坤等人使用神经网络对软件可靠性建模[2]，张婷婷等人[3]建立了贝叶斯组合模型用来提高模型的预测精度和模型的适应性，李思雨等人[3]利用极限学习机对软件可靠性建模，将经典软件可靠性模型和人工智能算法有机结合。

和声搜索（Harmony Search，简称 HS）算法[5]是 Greem 提出的一种新型启发式优化算法。类似于遗传算法对生物进化的模仿、模拟退火算法对物理退火的模拟以及粒子群优化算法对鸟群的模仿等。和声算法模拟了音乐演奏的原理，在音乐演奏中，乐师们凭借自己的记忆，通过反复调整乐队中各乐器的音调，最终达到一个美妙的和声状态。该算法简单，易于与其他算法混合，构造出具有更优性能的算法[4]，在参数寻优问题上有很大的优势。

支持向量机[6]（Support Vector Machine，简称 SVM）由 Vapnik 首先提出，它是一种监督式学习的方法，广泛地应用于统计分类以及回归分析中。该方法理论基础是统计学习理论，可用于模式识别和回归问题，能提供很好的全局最优性和泛化能力。SVM 的关键在于核函数的选择、SVM 中参数和核函数中参数难以确定。传统的方法有：实验法、经验选择法、交叉验证法等，其中交叉验证法使用较多。近年来，启发式算法已被成功地应用到 SVM 参数优化中来，如遗传算法、粒子群算法、模拟退火算法。本文利用和声搜索算法参数寻优的优点，用来优化支持向量回

收稿日期：2022-04-01

基金项目：安徽省高校优秀青年人才支持计划一般项目（项目编号：gxyq2017162）

作者简介：汪顺和（1982—），男，安徽枞阳人，讲师，硕士。研究方向：计算机控制、算法优化。

归中的参数，并将之用于软件可靠性预测，提出了一种基于和声搜索优化支持向量回归的软件可靠性预测模型，并通过实验证明该方法的可行性和有效性。

二、算法原理

（一）支持向量回归（Support Vector Regression，SVR）

SVM 作为一种监督式学习的方法，其理论基础是统计学习理论，既可以用于模式识别，又可以用于回归问题。这两方面上本质是相同的，都有一个可以是属性矩阵或者是自变量的输入 x，也都有一个输出 y。模式识别输出是分类标签的，回归输出是因变量，即相当于一个函数映射 $y=f(x)$。利用训练集中已知数据 (x,y) 来建立模型，再利用这个模型去对测试集进行分类或者回归。研究表明，SVM 在回归问题上也具有极好的性能。

设给定训练集 $\{(x_i,y_i)\}\in R^n\times R, i=1,2,\cdots,l$。对于线性回归，采用 $f(x)=w\cdot x+b$ 作为回归函数，其中 w 表示权重向量，b 表示偏项，目标就是寻找合适的 w 和 b，使得 $f(x_i)$ 估计 y_i 时的估计误差最小，即回归风险最小。采用 ε － 不敏感损失函数，回归问题转化为：

$$\begin{aligned}\min_{w,b,\xi,\xi^*}\ & \frac{1}{2}w^Tw+C\sum_{i=1}^{l}\xi_i+C\sum_{i=1}^{l}\xi_i^*\\ \text{s.t.}\quad & w^Tx_i+b-y_i\leqslant\varepsilon+\xi_i\\ & y_i-(w^Tx_i+b)\leqslant\varepsilon+\xi_i^*\\ & \xi_i,\xi_i^*\geqslant 0, i=1,2,\cdots,l\end{aligned}\tag{1}$$

其中，ξ_i 和 ξ_i^* 为松弛变量，C 为惩罚参数。

引入 Lagrange 函数，公式(1)可以转化为其对偶问题：

$$\begin{aligned}\min_{\alpha^{(*)}\in R^{2l}}\ & \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)+\varepsilon\sum_{i=1}^{l}y_i(\alpha_i^*-\alpha_i)\\ & +\varepsilon\sum_{i=1}^{l}y_i(\alpha_i^*-\alpha_i)\\ \text{s.t.}\quad & \sum_{i=1}^{l}(\alpha_i^*-\alpha_i)=0\\ & 0\leqslant\alpha_i,\alpha_i^*\leqslant C, i=1,2,\cdots,l\end{aligned}\tag{3}$$

在线性不可分的情况下，将最优化问题转化为二重 QP 问题，在原空间得到如下非线性判定函数：

$$f(x)=\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)K(x_i,x)+b\tag{4}$$

（二）标准和声搜索算法（Harmony Search，HS）

HS 算法模拟音乐家创作的过程：HS 算法中将每次演奏的和声类比于每次迭代的解向量；和声中的音调类比解向量中的分量；美学评价类比目标函数；最佳的和声类比全局最优。首先，算法产生 N 个初始解（和声）放入和声记忆库 HM 内，以概率 $HMCR$ 在 HM 内搜索新解，以概率 1-$HMCR$ 在 HM 外变量可能值域中搜索。然后，算法以概率 PAR 对新解产生局部扰动，判断新解目标函数值是否优于 HM 内的最差解，若是，则替换之；再不断迭代，直至达到预定迭代次数 N_i 为止。HS 算法中参数有：决策变量的个数 N，各个决策变量的取值范围 $[L_i,U_i]$，和声记忆库大小 HMS，和声记忆库取值概率 $HMCR$，音调微调概率 PAR，音调微调带宽 BW，最大迭代次数 N_i。HS 具体的算法步骤如下：

Step 1 确定目标函数和初始化参数。

Step2 初始化和声记忆库，并计算目标函数值。

在每个决策变量取值范围内，随机生成 HMS 个解向量放入和声记忆库 HM 中，每个决策变量按照以下公式生成：

$$x_i^k=L_i+(U_i-L_i)*\text{rand}(0,1)\tag{5}$$

其中 $i=1,2,\cdots,N, k=1,2,\cdots,HMS$；

$$HM=\begin{bmatrix}x_1^1 & x_2^1 & \cdots & x_N^1\\ x_1^2 & x_2^2 & \cdots & x_N^2\\ \cdots & \cdots & \cdots & \cdots\\ x_1^{HMS} & x_2^{HMS} & \cdots & x_N^{HMS}\end{bmatrix}$$

Step3 产生一个新和声（即新解）。

新和声 $x'=(x'_1,x'_2,\cdots,x'_N)$ 中任一音调（即变量）x'_i 按照如下规则产生：首先产生一个 0 到 1 之间的随机数 rand1，如果 rand1 小于和声记忆库取值概率 $HMCR$，则在和声记忆库中 HM 个第 i 维变量中随机选择一个，然后产生一个 0 到 1 之间的随机数

rand2，如果 rand2 小于音调微调概率 PAR，按照公式(6)进行局部干扰；如果 rand1 大于和声记忆库取值概率 $HMCR$，则按照公式(5)随机产生一个新解。

$$x'_i = x'_i + \text{rand2} * BW \tag{6}$$

Step4 若 Step3 中的新解优于 HM 中的最差和声，则将新解 x' 替换 HM 中当前最差和声，更新和声记忆库 HM。

Step5 判断算法终止条件，若满足，则停止迭代，输出最优解；否则重复步骤 Step3 和 Step4。

（三）基于和声搜索优化支持向量回归（HS-SVR）的软件可靠性预测

构造出一个具有良好性能的 SVM，核函数的选择是关键。核函数的选择包括两部分工作：一是核函数类型的选择，二是确定核函数类型后相关参数的选择。常用的核函数有线性核函数，多项式核函数，径向基核函数，Sigmoid 核函数。

径向基核函数也叫高斯核函数，是一种局部性强的核函数，其可以将一个样本映射到一个更高维的空间内，是应用最广的一个核函数，无论大样本还是小样本都有比较好的性能，而且其相对于多项式核函数参数要少。本文基于 HS-SVM 的软件可靠性预测模型选取径向基核函数作为核函数，公式(7)是其表达式，将公式(3)中的 $K(x_i, x_j)$ 替换成公式(7)。模型中需要优化的参数有惩罚因子 C、核函数参数 σ 及损失函数中的 ε，采用和声搜索算法来优化参数 C、σ 和 ε。

$$K(x_i, x_j) = \exp\left(-\frac{\| x_i - x_j \|^2}{2\sigma^2}\right) \tag{7}$$

其中，σ 是高斯核的宽度，其值大于 0。

和声搜索算法优化支持向量回归模型中参数时，将参数 (C, σ, ε) 看作为和声，即问题的解，回归指标均方差作为 HS 算法中的目标函数，当目标函数最小时所对应的解 (C, σ, ε) 是最优的。基于和声搜索算法的支持向量回归（HS-SVR）的软件可靠性预测建立的步骤如下：

(1)对选定的失效数据进行归一化处理，确定训练样本数据和预测样本数据；

(2)选择 SVM 模型中的回归指标均方差作为 HS 算法中的目标函数；

(3)初始化和声算法的参数和 C、σ 和 ε 的取值范围；

(4)将训练样本数据代入目标函数，得到 C、σ 和 ε 的最佳组合值 $(C_O, \sigma_O, \varepsilon_O)$；

(5)将最佳组合值 $(C_O, \sigma_O, \varepsilon_O)$ 代入 SVM 模型中，预测测试样本值。

三、仿真实验及结果分析

（一）仿真实验

为了验证本文提出模型的可行性和有效性，实验选择两组真实故障数据 SYS1 和 SYS2 作为测试数据，两组数据均来自 Muse 数据集[7]。SY1 数据中有 136 条记录，SYS2 数据中有 86 条记录。两组数据中每一条记录由两列构成，第一列是故障编号，第二列是当前故障与上次故障发生的时间间隔。

实验前两组数据中的第二列数据转化为累计时间，然后对故障数和连续时间进行归一化处理，SYS1 中取前 90 条数据作为训练样本，SY2 中取前 56 条数据作为训练样本，两组数据中剩下部分作为预测样本。实验平台采用 Matlab2013，实验中和声搜索算法参数设置为 $HMS = 5$，$HMCR = 0.95$，$PAR = 0.5$，$BW = 0.01$，$N_i = 200$。算法独立运行 10 次，取均方差最小所对应的参数为最终参数。均方差值越小，说明参数估计的精确度越高。将得到的最佳参数代入模型，对测试集进行预测，再与真实值进行比较。仿真实验结果如表 1 所示，表 1、表 2 和表 3 中的 G-O 和 M-O 模型的数据来自参考文献[8]。

表 1　两组数据对应的模型中的参数

数据 \ 模型	G-O	M-O	HS-SVR
	a，b	λ，θ	C，σ，p
SYS1	$a=124.466\,467, b=0.000\,051$	$\lambda=0.010\,34, \theta=0.022\,703$	$C=205.5701, \sigma=107.090\,564\,4, p=0.002\,868\,61$
SYS2	$a=94.149\,339, b=0.000\,019$	$\lambda=0.002\,23, \theta=0.020\,108$	$C=180.377\,8, \sigma=1.965\,591, p=0.000\,452\,872$

（二）结果分析

均方差（Mean squared error，MSE）和平方相关系数（Squared correlation coefficient，SCC 或 R^2）是回归问题中的两个重要指标。本文使用 MSE 和 R^2 对软件可靠性模型进行评价。令 y_i 是观测数据，$\overline{y}$ 是 n 个观察值的平均数，f_i 是模型数据，n 为数据个数。

均方差是误差平方的平均数，则均方差定义为：

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2 \tag{8}$$

平方相关系数，以 R^2 表示。R^2 提供了吻合度相关信息，用来评价预测模型的好坏。R^2 的值表示模型预测值真实值相吻合的程度，值越大吻合度越高。平方相关系数定义为：

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-f_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2} \tag{9}$$

由表 1 可以得出，用和声搜索算法优化支持向量回归中的参数方法效果很好，没有拟合不出来的情况，说明本文提出的软件可靠性预测模型是可行的。

对于 G-O 模型和 M-O 模型，以故障数据向量和失效累计时间向量作为输入量，利用和声搜索算法进行参数估计，参数估计结果见表 1。利用三种模型分别对相同的训练数据和测试数据进行分析计算，计算结果见表 2 和表 3。可以看出，HS-SVR 模型与 G-O 模型和 M-O 模型相比，HS-SVR 模型的均方差最小，平方相关系数最大，这说明模型的预测能力和吻合度非常好。

G-O 模型的均值函数：

$$\mu(t)=a(1-e^{-bt}) \tag{10}$$

M-O 模型的均值函数：

$$\mu(t)=\ln(\lambda\theta t+1)/\theta \tag{11}$$

其中，$\mu(t)$ 表示截止到 t 时刻检测到错误数的期望值，t 表示错误发现的时刻，a、b、λ、θ 是未知参数。

表 2　SYS1 的训练误差及预测误差

模型	训练误差		预测误差	
	MSE	R^2	MSE	R^2
G-O	0.004 631	92.706 7%	0.008 544	94.029 4%
M-O	0.000 380 677	99.400 5%	0.000 379 931	99.734 5%
HS-SVM	0.000 146 52	99.887 6%	0.000 298 27	99.797 6%

表 3　SYS2 的训练误差及预测误差

模型	训练误差		预测误差	
	MSE	R^2	MSE	R^2
G-O	0.000 42	98.987 1%	0.000 819	89.478 5%
M-O	0.000 163 856	99.605 271%	0.000 183 928	97.637 546 2%
HS-SVM	0.000 153 32	99.680 1%	7.900 01E-05	99.185 6%

将本文提出的 HS-SVR 模型、G-O 模型、M-O 模型各自生成的曲线和原始数据点进行对比，如图 1、图 2 所示，可以发现，模型曲线和原始观测数据点非常接近，仅有少量的偏离，说明模型的吻合度非常好，预测能力非常强，优于其他两种模型。

四、结论

软件可靠性模型是评估和预测软件可靠性的重要工具，支持向量回归在小样本数据预测中具有较为突出优势。本文为了提高软件可靠性预测的精确度，提出了一种基于和声搜索优化支持向量回归的软件可靠性预测模型，并通过两组数据进行实验，证实了模型的有效性。将实验结果与两个经典软件可靠性模型做比较，结果表明，该模型具有较好的预测效果。文中的方法原理简单，容易实现，也可以运用到其他应用中，具有一定的通用性。

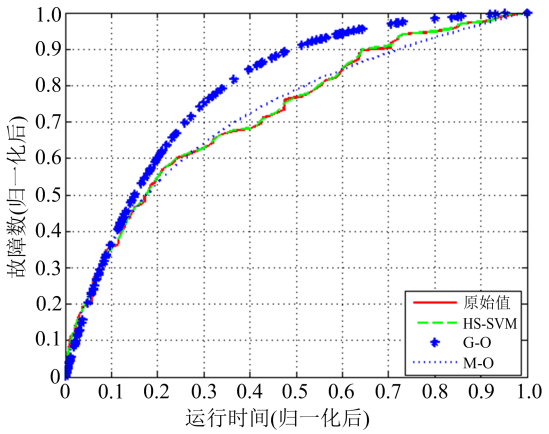

图 1　SYS1 的三种模型曲线及原始数据

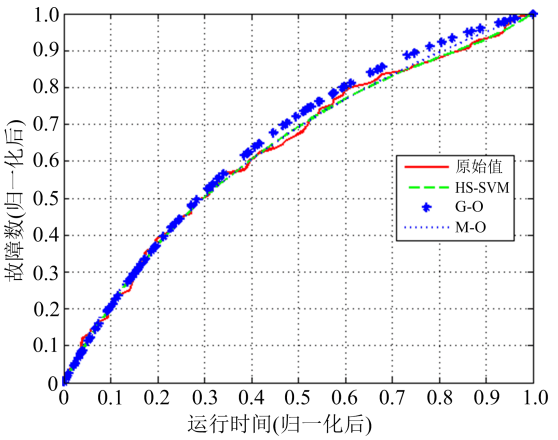

图 2　SYS2 的三种模型曲线及原始数据

参考文献：

[1] 陆民演. 软件可靠性工程[M]. 北京：国防工程出版社，2011：226.

[2] 张坤，李彤，郁湧. 基于灰色神经网络对软件可靠性模型的研究[J]. 计算机应用与软件，2009，26(12)：34-36.

[3] 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，2016，33(4)：1096-1101.

[4] 李思雨. 基于极限学习机的组合软件可靠性模型研究[J]. 计算机与现代化，2019(11)：44-48.

[5] 谷培义，高尚. 改进的和声搜索算法求解多目标优化问题[J]. 计算机与数字工程，2021，49(6)：1132-1133.

[6] VAPNIK V N. Statistical Learning Theory[M]. New Nork：Wiley，1998：138.

[7] LYU M R. Handbook of Software Reliability Engineering[EB/OL]. (2019-09-03)[2021-12-16]. http://www.cse.cuhk.edu.hk/～lyu/book/reliability/data.html.

[8] 周园园，钱丽，李敬明. 基于和声搜索算法的软件可靠性模型参数估计方法[J]. 山东理工大学学报(自然科学版)，2017，31(2)：44-48.

Software Reliability Prediction Based on Harmony Search Optimization Support Vector Regression

WANG Shunhe

(Learning Resources Center, Anhui Open University, Hefei 230022, China)

Abstract: In order to improve the accuracy of software reliability prediction, the support vector machine theory is used to model software reliability. And for the difficulty of parameter optimization in support vector regression, the harmony search algorithm is used to optimize the parameters in support vector regression, and a software reliability prediction model based on harmony search optimization support vector is proposed. Two groups of real data are used to test the proposed model, and the experimental results are compared with the classical softwares (G-O model and M-O model). The experimental results show that the software reliability prediction model based on harmony search optimization support vector regression has higher prediction accuracy.

Keywords: software reliability; harmony search; support vector machine

[责任编辑　李潜生]