

# 基于 PSOABC-SVM 的软件可靠性预测模型<sup>①</sup>

贾冀婷

(西安邮电大学 计算机学院, 西安 710121)

**摘要:** 软件可靠性预测是指在软件开发初期对软件中各模块出错的可能性进行预测, 对提高软件的可信性具有重要意义. 提出了一种基于粒子群与人工蜂群优化支持向量机的软件可靠性预测模型, 将粒子群优化算法与人工蜂群算法相结合的混合算法引入到支持向量机的参数选择中, 提高软件可靠性预测的效果. 实验结果表明, 该模型比 BP 网络预测模型、粒子群优化支持向量机等预测模型收敛速度更快、预测精度更高, 能更好的进行软件可靠性预测.

**关键词:** 软件可靠性预测; 粒子群优化算法; 人工蜂群算法; 支持向量机

## Software Reliability Prediction Based on PSOABC-SVM Model

JIA Ji-Ting

(Department of Computer, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

**Abstract:** Software reliability prediction can predict the fault-prone modules at the early age of software development. And it is important to improve the credibility of the software. In order to improve the effect of software reliability prediction, this paper proposes a PSOABC-SVM model to predict software reliability, and puts forward a model of predicting the software reliability based on PSOABC-SVM. The experimental results show that this model can achieve more precise prediction results than other prediction models such as BP neural network and PSO-SVM. The PSOABC-SVM model is more applicable for software reliability prediction.

**Key words:** software reliability prediction; particle swarm optimization; artificial bee colony algorithm; support vector machine

### 1 引言

软件可靠性预测是软件可靠性工程中的一个重要组成部分, 它在软件开发初期对软件中各模块的出错可能性进行预测, 从而使软件开发者能更加合理的对资源进行分配, 对提高软件的可信性和降低软件的开发成本都具有非常重要的意义. 目前, 常用的软件可靠性预测方法有线性判别分析、分类回归树、人工神经网络的聚类分析等.

支持向量机<sup>[1,2]</sup>(Support Vector Machine, SVM)是在有限样本统计学习理论基础上建立起来的一种通用学习方法, 具有良好的全局最优性和泛化能力, 在解决小样本、高维数、非线性等学习问题上取得了较好

的效果.

粒子群优化算法<sup>[3-5]</sup>(Particle Swarm Optimization, PSO)是一种基于群体智能的全局随机优化算法, 它通过种群中各个粒子之间的协作和信息共享来实现最优解的搜索. 该算法具有参数少、实现简单、收敛速度快等优点, 在解决不连续、不可微的非线性和组合优化等问题上具有较好的效果.

人工蜂群算法<sup>[6][7][8]</sup>(Artificial Bee Colony, ABC)是一种基于群体智能的随机优化算法, 该算法通过模拟蜜蜂群体的采蜜行为来实现最优解的搜索. 由于该算法具有控制参数少、易于实现、计算简单等优点, 已被越来越多的学者所关注.

<sup>①</sup> 收稿时间:2013-11-20;收到修改稿时间:2013-12-23

粒子群优化算法和人工蜂群算法虽然有许多优点,但是在求解最优问题时都存在容易陷入局部最优的缺陷,为了克服此缺陷、提高算法寻优的效率,本文将这两种算法相结合,提出了一种新的混合算法.支持向量机性能的好坏主要取决于对核函数和参数的选择,所以如何根据具体的模型来选择合适的核函数和参数成为了关键问题.本文利用粒子群优化算法与人工蜂群算法相结合的混合算法来进行支持向量机的参数选择,建立一种基于粒子群算法与人工蜂群算法优化支持向量机的软件可靠性预测模型,并通过仿真实验与 BP 网络预测模型、粒子群优化支持向量机等预测模型进行比较.实验证明,该模型具有训练速度快、预测精度高等优点.

## 2 算法原理

### 2.1 支持向量机

支持向量机是在有限样本统计学习理论上建立起来的一种通用学习方法,它针对二分类问题提出了最大间隔算法.支持向量机在分类问题上的基本思想是将线性不可分样本经过非线性变换映射到另一个高维空间中,在转换后的空间中找到一个最优的分界面,使其推广能力更优.

设训练集  $D = \{x_i, y_i\}_{i=1}^N$ , 输入向量  $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)}) \in R^n$ , 目标向量为  $\{-1, 1\}$ . 支持向量机分类器在非线性的情况下满足以下条件:

$$\begin{cases} \text{Min} \frac{1}{2} w^T w \\ \text{s.t. } y_i (w^T \phi(x_i) + b) \geq 1 \quad i=1,2,\dots,N \end{cases} \quad (1)$$

其中,  $w$  表示权重向量,  $b$  表示偏项.

$\phi(\cdot): R^n \rightarrow R^m$  是经过非线性变换将输入向量映射到另一个高维空间中.因此,公式(1)变为:

$$\begin{aligned} & \text{Min}_{w,b} \left( \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \right) \\ & \text{s.t.} \begin{cases} y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \end{aligned} \quad (2)$$

其中,  $\xi_i$  表示允许误分类,  $C$  表示惩罚参数.

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:

$$y(x) = \text{sgn} \left( \sum_{i=1}^N \alpha_i y_i k(x, x_i) + b \right) \quad (3)$$

其中,核函数为  $k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ .

在支持向量机中,核函数参数和惩罚参数的选择非常重要,只有选择合适的模型参数,支持向量机性能的优越性才能更好的发挥出来.于是,本文针对目前使用最广泛的径向基核函数支持向量机进行优化,采用粒子群优化算法与人工蜂群算法相结合的混合算法对其参数进行选择.

### 2.2 粒子群优化算法

粒子群优化算法最早是由 Eberhart 博士和 Kennedy 博士提出,于 1995 年发展起来的一种全局随机优化算法.该算法的基本思想是通过群体中各个粒子之间的协作和信息共享来实现最优解的搜寻.它包含进化计算和群体智能二者的特点,具有参数少、实现简单、收敛速度快等优点,并且已在神经网络训练、目标函数优化、模糊控制系统等领域被广泛应用.

粒子群优化算法是通过种群中粒子之间的协作和信息共享来搜索最优解.算法的具体步骤为:首先随机生成初始种群,并为种群中每个粒子都确定一个适应值.在每一代搜索最优值的过程中,每个粒子都会追随两个极值,一个是粒子自身迄今为止找到的最优解,另一个是整个种群迄今为止找到的最优解.经过迭代搜索最终得到最优解.

粒子群优化算法的数学表示为:设在一个  $n$  维的空间中,有一个种群包含  $m$  个粒子,即

$X = \{X_1, \dots, X_i, \dots, X_m\}$ , 第  $i$  个粒子位置表示为  $X_i = \{X_{i1},$

$X_{i2}, \dots, X_{in}\}^T$ , 其速度表示为  $V_i = \{V_{i1}, V_{i2}, \dots, V_{in}\}^T$ . 第  $i$

个粒子的个体极值表示为  $P_i = \{P_{i1}, P_{i2}, \dots, P_{in}\}^T$ , 种群的

全局极值表示为  $P_g = \{P_{g1}, P_{g2}, \dots, P_{gn}\}^T$ , 根据粒子群算

法的基本原理,第  $i$  个粒子可根据公式(4)、(5)改变其速度和位置:

$$v_{id}^{(t+1)} = v_{id}^{(t)} + c_1 r_1 (p_{id}^{(t)} - x_{id}^{(t)}) + c_2 r_2 (p_{gd}^{(t)} - x_{id}^{(t)}) \quad (4)$$

$$x_{id}^{(t+1)} = x_{id}^{(t)} + v_{id}^{(t+1)} \quad (5)$$

其中,  $d \in \{1, 2, \dots, n\}$ ,  $i \in \{1, 2, \dots, m\}$ ,  $t$  表示当前的进化代数,  $r_1$ 、 $r_2$  是分布在  $[0, 1]$  之间的随机数;  $c_1$  和  $c_2$  是学习因子.

### 2.3 人工蜂群算法

人工蜂群算法最早是由 Karaboga 于 2005 年提出

的一种基于群体智能的随机优化算法。该算法模拟自然界蜜蜂群体的采蜜行为,蜜蜂根据各自不同的分工来完成采蜜过程中相应的任务,并实现蜜蜂之间信息的交流与共享,最终得到问题的最优解。人工蜂群算法由于其控制参数少、易于实现、计算简单等优点,已被越来越多的学者所关注,并在解决函数的数值优化、人工神经网络训练、机器人路径规划等领域上得到广泛的应用。

人工蜂群算法中,蜜蜂根据不同的分工可分为采蜜蜂、观察蜂和侦察蜂三种,每个蜜源的位置代表一个可能解,蜜源的花蜜量对应着每个可能解的适应值。算法的具体步骤为:首先,随机产生初始种群,即  $N$  个初始解。每个解  $x_i$  ( $i=1,2,\dots,N$ ) 表示一个  $D$  维的向量,  $D$  为优化参数的个数。接着,三种蜜蜂分别开始对所有的初始解进行循环搜索。采蜜蜂记住自身历史的最优解,在附近邻域进行搜索,其搜索公式为:

$$x'_{ij} = x_{ij} + \phi(x_{ij} - x_{kj}) \quad (6)$$

其中,  $k \in \{1,2,\dots,N\}$ ,  $j \in \{1,2,\dots,D\}$ , 且  $k \neq i$ ,  $\phi$  是  $[-1,1]$  之间的随机数。

采蜜蜂比较自身历史最优解和邻域搜索解,当邻域搜索解优于自身历史最优解时,替换自身历史最优解;反之,则保持不变。当所有的采蜜蜂搜索完成后,将蜜源信息通过舞蹈区与观察蜂共享。观察蜂按照与花蜜量相关的概率选择一个蜜源位置,花蜜量大的采蜜蜂吸引观察蜂的概率大于花蜜量小的采蜜蜂。观察蜂像采蜜蜂那样对历史最优的位置进行更新,并检查新位置的花蜜量。若新位置优于历史最优的位置,则用新位置替换历史最优位置;反之,则保持不变。观察蜂选择某个蜜源的概率  $q_i$  的计算公式为:

$$q_i = \frac{f_i}{\sum_{n=1}^N f_n} \quad (7)$$

其中,  $f_i$  表示第  $i$  个解的适应值,  $N$  表示蜜源数量。

如果一个蜜源经过限定的循环次数之后没有发生变化,则该蜜源处的采蜜蜂变为侦察蜂,该蜜源位置将会被侦察蜂找到的随机新位置替换。假设被放弃的位置为  $x_i$ , 则侦察蜂替换  $x_i$  的公式如下:

$$x'_{ij} = x_{\min}^j + \text{rand}(0,1)(x_{\max}^j - x_{\min}^j) \quad (8)$$

### 3 基于 PSOABC-SVM 软件可靠性预测模型

#### 3.1 粒子群优化算法与人工蜂群算法相结合的混合算法 PSOABC

粒子群优化算法在寻优过程中存在容易陷入局部最优的缺陷,即算法没有较强的搜索能力,而人工蜂群算法中的搜索算子侧重于提高搜索能力。因此将这两种算法相结合,提出一种新的混合优化算法-PSOABC,用来提高算法寻优的能力。为了进一步提高算法的搜索能力,PSOABC 将初始种群随机分为两组,其中一组中的个体按照粒子群优化算法进行寻优,而另一组中的个体按照人工蜂群算法进行寻优。在寻优过程中,每隔一定的迭代次数对两组种群产生的最优个体进行比较,找到相对更优的个体,使两组避免陷入局部最优。具体的 PSOABC 算法的步骤如下:

1) 设置参数,初始化种群;

2) 将种群随机分为两组,第一组中的个体按照粒子群算法进行寻优,第二组中的个体按照人工蜂群算法进行寻优;

3) 当迭代次数达到指定次数时,对两组产生的最优个体进行比较,将较优个体的值赋给公式(4)中的  $p_g$ , 并从第一组中随机选取值赋给公式(6)中  $x_k$ ;

4) 两组分别通过公式(4)、(5)、(6)产生新解,并计算其适应值;

5) 第一组更新  $p_i$  和  $p_g$ ;

6) 第二组中如果新解的适应值优于  $x_i$ , 则替换  $x_i$ , 否则保持不变;计算  $x_i$  的适应值,并根据公式(7)计算概率  $q_i$ ;观察蜂根据  $q_i$  选择蜜源,并根据公式(6)产生新解,如果新解的适应值优于  $x_i$ , 则替换  $x_i$ , 否则保持不变;

7) 判断是否有需要放弃的解,若有,则按照公式(8)产生新解;

8) 如果当前群体的最优个体满足精度要求或进化已达到最大迭代次数,则终止算法,否则跳转至 3)。

#### 3.2 基于 PSOABC-SVM 的软件可靠性预测模型

本文采用粒子群算法与人工蜂群算法相结合的混合算法优化支持向量机的模型参数和核参数,通过搜索并使用参数的最优值来提高支持向量机的软件可靠性预测精度。支持向量机选用了目前使用最广泛的径向基核函数,其中包括参数  $C$  和  $\sigma$ 。

基于 PSOABC-SVM 进行软件可靠性预测的步骤如下:

1)初始化支持向量机参数  $C$  和  $\sigma$ , 确定支持向量机参数寻优的终止条件;

2)利用 PSOABC 算法来搜索  $C$  和  $\sigma$  的值;

3)用支持向量机对训练样本进行训练, 以获得  $C$  和  $\sigma$  的最佳组合, 将获得的  $C$  和  $\sigma$  作为预测模型参数.

#### 4 实验及结果分析

本文以一个具体的应用软件为例, 通过对软件系统中的部分模块的规模、扇入、扇出、控制流路径及缺陷数等相关数据进行仿真实验, 来评价本文所提出的预测模型的预测精度. 表 1 中列举了该软件系统中的 10 个模块的相关信息.

表 1 实验数据

序号	规模	扇出	扇入	控制流路径	缺陷数
1	25	3	1	3	0
2	26	3	1	3	1
3	30	2	2	2	2
4	35	6	2	4	2
5	37	7	6	15	5
6	40	7	12	8	4
7	46	5	3	13	5
8	51	6	3	21	7
9	55	4	2	12	5
10	62	8	3	18	8

仿真实验中将表 1 的前 7 个数据样本作为训练样本来建立预测模型, 后 3 个数据样本作为测试样本来评价预测模型的精度. 将训练样本和测试样本进行标准化处理后, 分别输入 BP 网络预测模型、PSO-SVM 预测模型和 PSOABC-SVM 预测模型中进行训练. 训练结束后, 用三种预测方法分别得到适用于数据样本的预测模型, 将测试样本输入预测模型中进行预测. 预测结果及误差如表 2、3 所示.

表 2 预测结果

测试集序号	真实值	BP 网络预测模型	PSO-SVM 预测模型	PSOABC-SVM 预测模型
1	7	6.6703	7.2954	7.1063
2	5	4.5329	5.4534	5.0932
3	8	7.5531	8.3223	8.1376

表 3 预测误差(%)

测试集序号	BP 网络预测模型	PSO-SVM 预测模型	PSOABC-SVM 预测模型
1	4.71	4.22	1.52
2	9.34	9.07	1.86
3	5.59	4.03	1.72
MAPE	6.55	5.77	1.70

#### 5 结语

实验结果表明, PSOABC-SVM 预测模型使用优化后的模型参数和核参数, 预测精度明显优于 PSO-SVM 预测模型和 BP 网络预测模型. 由此可见, 本文的 PSOABC-SVM 预测模型获得了较好的预测效果并具有一定的实用性.

#### 参考文献

- 1 Vapnik V. Statistical Learning Theory. New York: Wiley, 1998.
- 2 Elish KO, Elish MO. Predicting defect-prone software modules using support vector machine. Journal of Systems and Software, 2007, 81(5): 649-660.
- 3 Kennedy J, Eberhart RC. Particle Swarm Optimization. Proc. of IEEE conf on Neural Network, Perth. IEEE Piscataway. 1995. 1942-1948.
- 4 Eberhart RC, Kennedy J. A new optimizer using particle swarm theory. Proc. of the Sixth International Symposium on Micro Machine and Human Science. Nizgoya, Japan. 1995. 39-43.
- 5 Kennedy J, Eberhart RC. Swarm Intelligence. San Francisco, CA. Morgan Kaufmann Publishers Inc. 2001.
- 6 Karaboga D. An idea based on honey bee swarm for numerical Optimization. Ercyes University, Engineering Faculty, Computer Engineering Department, 2005.
- 7 Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony algorithm. Journal of Global Optimization, 2007, 39(3): 459-471.
- 8 Karaboga D, Basturk B. A comparative study of artificial bee colony algorithm. Applied Mathematics and Computation, 2009, 214(1): 108-132.