

基于鸽群改进 RBF 网络的软件质量预测方法

田川 王闯 刘灿 郭强

(中航机载系统共性技术有限公司 江苏 扬州 225000)

摘要: 日益显现的软件质量问题给高安全、高可靠装备的研制提出严峻挑战,而精确、可信的软件质量预测技术是降低软件测试成本、提升软件质量的有效策略。针对装备大规模、高综合化软件质量与度量之间的强非线性关系,提出一种基于鸽群改进 RBF 网络的软件质量预测方法。考虑一种软件质量评价体系,以评价软件质量;对于样本提出一种基于主成分分析的数据预处理方法;在所提出的软件质量度量模型中,利用鸽群优化算法改进 RBF 网络的权值和阈值,以优化训练过程;利用机载软件的质量预测预评价实例,验证了方法的有效性。

关键词: 软件质量; 鸽群优化算法; 径向基函数神经网络; 主成分分析法

中图分类号: TP311 **文献标识码:** A **文章编号:** 1671-654X(2021)05-0024-05

Software Quality Prediction Based on PIO – RBF

TIAN Chuan ,WANG Chuang ,LIU Can ,GUO Qiang

(AVICAS Generic Technology Co. Ltd. ,Yangzhou 225000 ,China)

Abstract: The increasingly emerging software quality problems pose serious challenges to the development of high safety and high reliability equipment design. It is noted that accurate and reliable software quality prediction technology is an effective strategy to reduce the software testing cost and improve the software quality. Aiming at analyzing the strong nonlinear relationship between the quality and measurement of large-scale and highly integrated software ,a software quality prediction method based on a pigeon inspired optimization improved RBF network proposed in this paper. Firstly ,a software quality evaluation systems is considered to evaluate the software quality. Secondly ,a data preprocessing method based on principal component analysis is proposed for samples. Then ,in the proposed software quality measurement model ,the pigeons swarm optimization algorithm was used to improve the weights and thresholds of the RBF network to optimize the training process. Futhermore ,an example of quality prediction and pre-evaluation of airborne software is applied to verify the effectiveness of the proposed method.

Key words: software quality; pigeon-inspired optimization; RBF network; principal components analysis

引言

近年来,随着装备信息化和现代化的快速发展和推进,其软件的规模、应用范围和综合化程度都伴随功能的多样化和高精化呈指数增长^[1]。但随之而来的是更为明显的软件质量问题,其始终是制约装备安全性和可靠性保障和提升的瓶颈问题^[2]。以高安全和可靠需求的机载系统为例,在追求先进性能和复杂功能的同时,机载软件质量问题导致的缺陷和失效问题,会严重影响了航空器的安全飞行,甚至可能导致机毁人亡的重大事故^[3-5]。因此,关于软件质量的研究十分必要,逐渐引起了研究人员的广泛关注,并在软件质量评

价^[6-7]、软件质量预测^[8-9]和软件质量控制等方面的技术均取得了实质的进展^[10]。

在软件预测方面,良好的预测方法可大幅降低测试成本,提升软件质量^[11]。对于软件预测建模技术,国内外主要围绕基于软件错误测试报告或软件质量度量来预测软件的缺陷数^[11]、可靠性^[12]和质量评分^[13]等方法的研究展开。相关的质量预测模型研究工作包括采用传统的线性与非线性回归及回归树方法,依据软件复杂度、规模、需求评审数据和测试数据等建立质量预测模型,对软件缺陷数和综合质量进行预测^[14]。为解决传统方法中的概率假设先验条件问题,文

收稿日期:2021-07-08 修订日期:2021-08-27

基金项目:装备技术基础科研项目资助(2022ZX31006)

作者简介:田川(1991-),男,辽宁沈阳人,工程师,博士。

献[15]采用模糊变量,建立了模糊软件可靠性预测模型。

为进一步处理大规模软件质量建模中的非线性关系,集群智能和人工智能算法也逐步应用于软件质量预测建模技术中。如运用BP等人工神经网络的软件质量预测模型,另外,为解决BP等神经网络收敛速度慢和已陷入局部最优的问题,引入了神经网络训练算法^[16-17]。此外,优化软件质量预测模型,具有更快收敛速度且在理论上可逼近任意的RBF神经网络也应用于预测模型构建中^[18]。尽管上述方法对神经网络的训练速度和收敛性问题有一定的改善,但其仍旧是制约网络模型性能的重要挑战。

本文提出一种基于鸽群改进RBF网络的软件质量预测方法。为评价软件质量,考虑一种基于软件度量的软件质量评价体系,从可靠性、易用性和效率等多方面综合评价软件质量。为提升样本质量,提出了一种基于主成分分析的数据预处理方法。并利用鸽群优化算法^[19]改进RBF神经网络的训练过程,设计一种PIO-RBF的软件质量预测模型。最后,利用机载软件质量预测与评价仿真实例验证了所提方法的有效性。

1 基于软件度量的软件质量评价体系

本文首先提出一种基于软件度量的软件质量评价体系,并给出软件度量数据的预处理方法,为软件预测模型的构建建立基础。

1.1 软件质量度量体系

参考ISO/IEC9126软件质量度量标准,以及GJB5000A基本过程和DO-178C适航过程的质量特性,建立如图1所示的软件质量度量评价体系。度量元详情如表1所示。

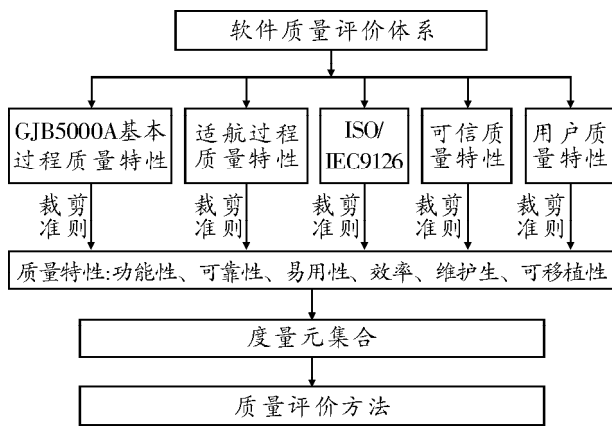


图1 软件质量度量评价体系

在该软件质量度量体系中将质量的结果分为优秀、良好、合格、不合格4种类型。

表1 软件质量评价体系

Index	评价指标	Index	评价指标
Index1	软件规模	Index9	项目成本
Index2	模块数	Index10	文档审查问题数
Index3	代码行数	Index11	代码审查问题数
Index4	有效代码行数	Index12	静态分析问题数
Index5	注释行数	Index13	动态测试问题数
Index6	行数超过200模块	Index14	首轮动态测试用例
Index7	圈复杂度	Index15	回归测试次数
Index8	扇入扇出模块数	Index16	回归测试问题总数

1.2 软件度量数据预处理方法

对于软件质量度量体系下大量的度量数据,本文采用主成分分析法(Principal Components Analysis, PCA)对数据进行预处理。PCA作为一种多元统计分析方法,可有效降低样本数据的维数,剔除非相关度量数据,构建合适的软件质量预测数据样本和验证数据样本,以提升预测的效率。

具体地,考虑n个样本的s个变量的原始数据矩阵X^o,其元素为x_{iτ}^o,进行如下的数据预处理。

1) 对原始样本数据矩阵作标准化操作,并获取标准化数据矩阵X^c,其元素为:

$$x_{i\tau}^c = \frac{x_{i\tau}^o - \bar{x}_{\tau}^o}{\varepsilon} \quad (1)$$

式中 \bar{x}_{τ}^o 为第τ列原始数据样本均值,ε为x_{iτ}^o的最大和最小值之差。

2) 依据标准化数据矩阵X^o,建立s个变量相关系数矩阵X^R形式为:

$$X^R = \begin{bmatrix} x_{11}^R & x_{12}^R & \cdots & x_{1s}^R \\ x_{21}^R & x_{22}^R & \cdots & x_{2s}^R \\ \vdots & \vdots & \cdots & \vdots \\ x_{s1}^R & x_{s2}^R & \cdots & x_{ss}^R \end{bmatrix} \quad (2)$$

式中 $x_{i\tau}^c = \frac{(X^o)^T X^o}{n-1}$

3) 计算X^R的特征值,并以大小为依据进行排序,得到排序后的特征值为λ₁ ≥ λ₂ ≥ ... ≥ λ_s。

4) 计算贡献率和累计贡献率。则前μ个主成分的贡献率和累计贡献率为:

$$z_{\mu} = \frac{\lambda_{\mu}}{\sum_{i=1}^s \lambda_i} \quad (3)$$

$$Z_{\mu} = \frac{\sum_{i=1}^{\mu} z_{\mu}}{\sum_{i=1}^s \lambda_i} \quad (4)$$

本文按照累计贡献率大于等于90%为基准确定软件度量元主成分个数,并将其作为预测输入因子,以

达到消除信息重叠、提升神经网络预测能力的目的。

2 鸽群改进 RBF 网络的软件综合预测方法

本节提出一种鸽群改进的 RBF 神经网络(PIO-RBF)算法,用以构建软件质量综合预测评价的计算模型。其中利用鸽群算法对 RBF 神经网络的权重等结构参数进行优化。

2.1 RBF 神经网络

RBF 神经网络是一种可以任意精度逼近任意的非线性函数、具有全局逼近能力的 3 层前馈型神经网络。其具体原理拓扑结构如图 2 所示。

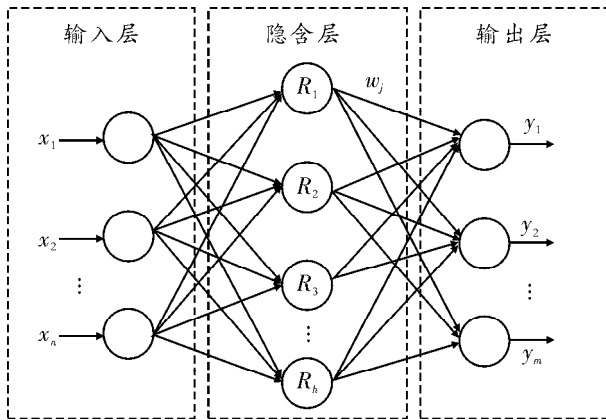


图 2 RBF 神经网络结构示意图

RBF 神经网络包括输入层、单隐层和输出层 3 层,其采用径向基函数神经网络作为隐层神经元的激活函数,其从输入层到隐含层空间的变换是非线性的,而从隐含层空间到输出层的变换是线型的。

令 RBF 神经网络以高斯函数作为激活函数,表示为:

$$R_i(X - c_i) = \exp\left(-\frac{1}{2\sigma_i^2} \|X - c_i\|^2\right) \quad (5)$$

式中,向量 \$X = [x_1, x_2, \dots, x_n]\$ 为 \$n\$ 维软件度量输入样本数据, \$n\$ 为输入层的节点数; \$c_i\$ 为第一个隐含层神经元的中心向量,其与输入向量 \$X\$ 维数相同; \$\sigma_i\$ 为径向基函数的标准差;而 \$\|X - c_i\|\$ 表示 \$X\$ 与 \$c_i\$ 之间的欧氏距离。

此外,图 2 中的输出函数可表示为:

$$y_j = \sum_{i=1}^m \omega_{ij} R_i(X - c_i) \quad (6)$$

式中 \$y_j\$ 为第 \$j\$ 个输出层神经元的输出值; \$w_{ij}\$ 为第 \$i\$ 个隐含层神经元到第 \$j\$ 个输出神经元的连接权值。那么输出值 \$y_j\$ 即为计算的软件质量评价预测值。但需要指出,经典的 RBF 神经网络预测方法的预测精度严重依赖于中心向量参数和标准差等参数的选取。非最优的参数选择往往导致预测效果不佳等问题。

2.2 鸽群优化算法

针对上述问题,本文采用鸽群算法对 RBF 神经网络的参数进行更新和优化。鸽群优化算法是学习和模拟鸽群自主归巢行为的一种集群智能优化算法。相较于粒子群等模拟鸟类的仿生优化算法,鸽群算法具有更快的收敛速度,且一定程度上缓解了局部最优问题,其可以视作对粒子群算法的一种改进。

鸽群算法包含了 2 个仿生工具数学:地图和指南针算子以及地标算子,是对利用太阳、地球磁场和地标寻巢的高度总结,其原理如图 3 所示。

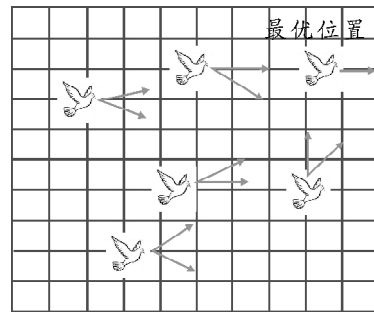


图 3 鸽群算法原理示意图

地图和指南针算子运用算子形式进行每个鸽子的速度和位置更新,公式的数学表述为:

$$v_k^l = v_k^{l-1} e^{-\gamma \times l} + rand(p_{gbest} - p_i^{l-1}), \quad (7)$$

$$p_k^l = p_k^{N_a-1} + v_k^l, \quad (8)$$

式中 \$v_k\$ 和 \$p_k\$ 分别表示第 \$k\$ 个鸽子的位置和速度; \$l\$ 为更新的代数;向量 \$p_k = [p_{k1}, p_{k2}, \dots, p_{kM}]\$, \$v_k = [v_{k1}, v_{k2}, \dots, v_{kM}]\$; \$r\$ 表示地图和指南针因数,其取值范围为 \$0 \sim 1\$; \$rand\$ 函数表示 \$0 \sim 1\$ 的随机数; \$p_{gbest}\$ 是 \$N_a\$ 次迭代循环结束后通过比较所有鸽子的位置得到的鸽群全局最优位置。当该更新的循环次数达到所要求的最大迭代次数后,或预设指标小于某阈值时,更新结束,再利用地标算子继续进行计算。

地标算子模拟鸽群接近鸽巢时的生物学特征,在每一次更新运算中,鸽子数 \$N_d^l\$ 都减少一半(初值为 \$M\$),具体的位置更新公式为:

$$N_d^l = \frac{N_d^{l-1}}{2} \quad (9)$$

$$p_c^{l-1} = \frac{\sum_{q=1}^{N_d^{l-1}} p_q^{l-1} f(p_q^{l-1})}{N_d^{l-1} \sum_{q=1}^{N_d^{l-1}} f(p_q^{l-1})} \quad (10)$$

$$p_k = p_q^{l-1} + rand(p_c^{l-1} - p_q^{l-1}) \quad (11)$$

式中 \$f(\cdot)\$ 表示适应度函数; \$p_c\$ 表示剩下鸽子的中心位置,即地标。该位置更新的停止条件为循环至最大预设代数或 \$f(\cdot)\$ 小于预设阈值。

2.3 鸽群改进RBF网络算法设计

基于上述方法,本文提出采用鸽群优化算法改进RBF网络的权重、中心位置和标准差 σ_i 。将RBF网络预测模型的权重、中心位置和标准差 σ_i 值作为鸽群算法的拟优化鸽群位置向量 p_k ,令鸽群算法的适应度函数为如下形式:

$$f = \frac{\lambda_f}{m} \sum_{j=1}^m (Y_j - y_j)^2 \quad (12)$$

式中 Y_j 为第 j 个理想输出值,而 y_j 为第 j 个理想输出值实际输出值, λ_f 为常数增益。

具体的神经网络训练步骤为:

Step 1: RBF网络初始化。初始化RBF网络结构,选取 n 个软件质量度量数据样本作为输入数据,初始化隐含层神经元个数 h 、RBF的中心 c_i 、标准差以及各权重值。

Step 2: 数据预处理。按照PCA方法对软件质量度量元数据做主要成分分析,并通过模糊聚类的方法进行软件质量度量元和度量元数据的筛选预处理。

Step 3: 计算标准差。按下式计算RBF的标准差:

$$\sigma_i = \frac{c_{\max}}{\sqrt{2h}} \quad i = 1, 2, \dots, h \quad (13)$$

式中 c_{\max} 表示所选取各基函数中心点间的最大距离。

Step 4: 计算隐含层与输出层间的连接权值。按如下公式计算:

$$w_{ij} = \exp\left(\frac{h}{c_{\max}^2} \|X - c_i\|^2\right) \quad i = 1, 2, \dots, n \quad (14)$$

Step 5: 利用鸽群算法优化RBF网络参数。利用式(3)~式(8)优化RBF神经网络的权重、中心位置和标准差参数值。

Step 6: 获取神经网络优化后参数,计算软件质量预测值。把Step 5中优化后的权值、赋予RBF神经网络,然后可通过式(2)计算软件质量预测值。

Step 7: 更新RBF网络的参数。按照下列式(15)~式(16)调整和更新RBF网络的参数,包括:RBF的中心 c_i 、标准差 σ_i 、隐含层与输出层间的连接权值 w_{ij} 。具体更新方式为:

$$\begin{cases} \Delta w_{ij}(t) = \eta_1 (Y_{ij} - y_{ij}) R_i \\ \Delta \sigma_i(t) = \eta_1 \frac{w_{ij}(t)}{\sigma_i^3} \sum_{j=1}^m (Y_{ij} - y_{ij}) \|X - c_i\|^2 R_i \\ \Delta c_i(t) = \eta_1 \frac{w_{ij}(t)}{\sigma_i^2} \sum_{j=1}^m (Y_{ij} - y_{ij}) \|X - c_i\|^2 R_i \end{cases} \quad (15)$$

$$\begin{cases} w_{ij}(t+1) = w_{ij}(t) + \eta_2 \Delta w_{ij}(t) \\ \Delta \sigma_i(t) = \sigma_i(t) + \eta_2 \Delta \sigma_i(t) \\ c_i(t+1) = c_i(t) + \eta_2 \Delta c_i(t) \end{cases} \quad (16)$$

Step 8: 学习终止条件。判断式(8)的误差适应度是否符合预设阈值。如符合则停止RBF网络学习,若不符合则跳转至Step 2。

通过上述学习和优化步骤,可得软件质量预测的最优神经网络计算模型。

3 软件研制质量预测实例

本部分对实践中的某项目软件质量进行预测,以验证本文所提出方法的有效性。

3.1 机载软件质量评价样本与参数设置

本文选取某项目机载系统研制中的软件数据,用以鸽群改进RBF的软件质量预测实验。根据软件研制过程,选取第1部分表1中的软件度量元,并将软件最终的质量评价结果划分为V1(优秀)、V2(良好)、V3(合格)、V4(不合格)。其中,软件质量综合评价结果的4个等级采取打分原则进行划分,划分标准为:

优秀:3.01~4.00;良好:2.50~3.00;一般:2.00~2.49;差:0.00~1.99。

本文收集了关于某项目机载系统研制任务中软件质量标准化自评和用户正式评价的质量分数及相关数据共56组,将其中的40组作为训练样本,用以改进的RBF神经网络模型的训练,而另外的16组数据作为验证神经网络模型的测试样本。

在实例验证中,对鸽群改进RBF网络的参数设置如下:设置鸽群算法的鸽群数 $M=320$,最大迭代次数 l_{\max} 为500,地图和指南针因子 $r=0.6$ 。

此外,本文在仿真中还采用了软件质量预测和评价中的传统RBF神经网络模型、粒子群改进BP神经网络模型以及线性回归模型作为对比,进行试验,以验证所提出方法的有效性。

3.2 结果分析

仿真试验结果如图4~图6所示。图4显示了各种方法对软件质量预测结果与真实软件质量评价结果的对比情况,尽管各类方法从总体趋势上都能对软件质量作出符合的预测,实际的评价结果和改进算法的预测结果一致。但本文所提出的鸽群改进RBF网络模型的预测结果与真实软件质量评价结果更为贴合。图5反映了各种方法的软件质量预测结果与真实质量评价结果之差,相较于其余3种方法,本文提出的改进RBF软件质量预测模型具有更精准的预测结果。图6为传统RBF网络、粒子群改进BP网络和本文所提出鸽群改进RBF网络训练情况曲线。从图中可以看出,相较于其他两种神经网络,本文所提出的算法及模型具有更快的收敛速度和逼近精度,在这两个指标方面分别有超过74%和65%改进和提升。

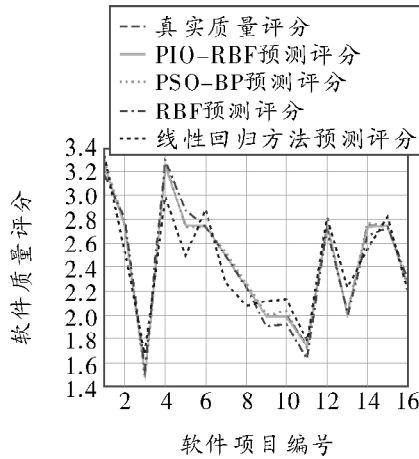


图 4 软件质量评分预测验证

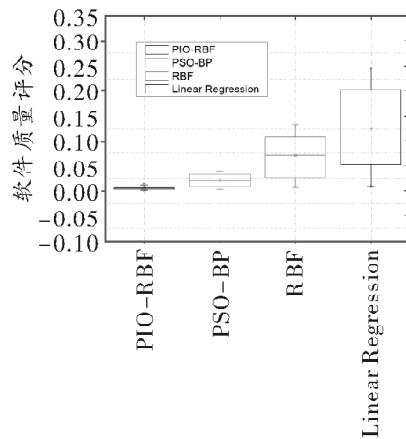


图 5 运用各模型预测软件质量评分误差

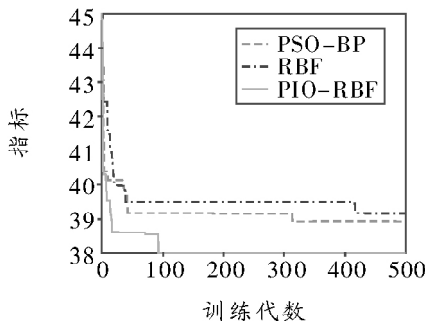


图 6 训练情况曲线

4 结束语

本文针对软件质量预测问题,提出了一种基于鸽群改进 RBF 网络的软件质量预测方法。首先,本文考虑一种软件质量评价体系。对于原始数据,采用基于 PCA 主成分分析的预处理方法筛选样本。之后,采用鸽群算法改进 RBF 网络,改进网络模型逼近精度,提升收敛速度,并给出整套算法的设计和实现流程。最后,本文进行了机载软件质量预测与评价的实例验证,仿真显示相较于传统的软件质量预测模型,所提出方

法具有更好的预测性能。

参考文献:

[1] Moises ,Rodriguez ,Mario ,et al. Software Verification and Validation Technologies and Tools [J]. IEEE Software , 2019 ,36(2) : 13 - 24.

[2] Klotins E ,Unterkalmsteiner M ,Gorschek T. Software Engineering Anti - Patterns in Start - Ups [J]. IEEE Software , 2019 ,36(2) : 118 - 126.

[3] 郝继锋,叶宏,任晓瑞. DO - 333 标准形式化方法研究 [J]. 航空计算技术 ,2020 ,50(1) :128 - 133.

[4] 罗玲. 采用随机 Petri 网的嵌入式机载软件可靠性检测 [J]. 计算机工程与应用 ,2019 ,55(1) :233 - 240.

[5] 莊露,陆中,张子文. 基于随机 Petri 网的机载系统动态可靠性建模 [J]. 西北工业大学学报 ,2020 ,38 ,184(4) :169 - 177.

[6] 岳川,彭小红. 带有区间信息的软件质量评价模型 [J]. 计算机科学 ,2019 ,46(10) :209 - 214.

[7] 涂俊翔,邱亮,李志伟. 面向广义质量体系的软件质量贝叶斯评估模型 [J]. 计算机系统应用 ,2019 ,28(4) :242 - 246.

[8] 宫丽娜,姜淑娟,姜丽. 软件缺陷预测技术研究进展 [J]. 软件学报 ,2019 ,30(10) :170 - 194.

[9] 柳溪,刘嘉,刘畅,等. 基于系统理论事故模型的雷达系统软件缺陷预测应用研究 [J]. 现代雷达 ,2020 ,42(11) :91 - 96.

[10] 余红英,徐凯健. 航空发动机控制软件质量量化控制的应用 [J]. 航空发动机 ,2018 ,44(5) :102 - 106.

[11] 陈翔,顾庆,刘望舒,等. 静态软件缺陷预测方法研究 [J]. 软件学报 ,2016 ,27(1) :1 - 25.

[12] 楼俊钢,蒋云良,申倩,等. 软件可靠性预测中不同核函数的预测能力评估 [J]. 计算机学报 ,2013 ,36(6) :1303 - 1311.

[13] 李虎,史晓华,杨海燕,等. 软件质量评价技术 [J]. 计算机研究与发展 ,2002 ,39(1) :61 - 67.

[14] 王明珠. 基于 Gompertz 模型的软件质量与测试过程评估分析 [J]. 信息系统工程 ,2020 ,322(10) :32 - 33.

[15] Yue C. An Intuitionistic Fuzzy Projection - based Approach and Application to Software Quality Evaluation [J]. Soft Computing ,2020 ,24(1) :429 - 443.

[16] 宫丽娜,马怀志. 粒子群算法优化的 BP 网络预测软件质量 [J]. 计算机工程与应用 ,2014 ,50(23) :65 - 68.

[17] 郑鹏. 基于 LM - BP 神经网络的软件质量综合评价 [J]. 山东理工大学学报(自然科学版) ,2016 ,30(3) :78 - 82.

[18] 俞华锋. 神经网络在软件可靠性预测中的应用研究 [J]. 计算机仿真 ,2011 ,28(4) :203 - 207.

[19] Duan H ,Huo M ,Shi Y. Limit Cycle Based Mutant Multi - Objective Pigeon Inspired Optimization [J]. IEEE Transactions on Evolutionary Computation ,2020 ,24(5) :948 - 959.