说 明 书

1. **发明创造名称**

一种基于神经网络和差分进化算法的软件缺陷预测方法

1. **技术领域**

本发明涉及软件工程中的软件分析及缺陷预测领域，特别涉及一种基于神经网络和差分进化算法的软件缺陷预测方法。

1. **背景技术**
2. **发明内容**
3. 本发明的目的在于克服现有技术的缺点与不足，提供一种基于神经网络和差分进化算法的软件缺陷预测方法。
4. 本发明的目的通过下述技术方案实现：
5. 一种基于神经网络和差分进化算法的软件缺陷预测方法所述方法包括下述步骤：

步骤1) 构造数据集：通过版本控制工具收集待预测软件的源代码文件，从中抽取程序模块，通过分析提取度量元，标记不同模块相应的缺陷个数，作为训练数据集D；

步骤2）数据预处理与数据集分割: 对数据进行预处理，包括缺失数据的处理，重复数据删除等，然后将数据集划分成训练集D1和验证集D2。

步骤3）构造初始预测模型结构：构造多层感知机模型作为预测模型，模型包含输入层，隐藏层，及一个输出层，每一层的结点个数仅代表最大结点可能个数；

步骤 4) 求解模型参数: 基于步骤二中分割出来的训练集D1，通过改进的多目标优化方法，在优化模型排序性能，回归性能，以及模型复杂度性能的同时，对步骤二中的多层感知机模型进行剪枝，一方面进行特征选择，一方面选择优化感知机模型结构。优化结果得到一组模型参数。

步骤5）模型选择：基于步骤4中训练得出的模型参数所确定的模型，使用步骤2中分割出的验证集D2进行验证，计算各个模型在验证集上的排序性能，回归性能以及模型复杂度。基于各个模型在训练集和验证集上的排序性能，回归性能及模型复杂度，选取模型。

步骤6）基于步骤6所选取的模型，输入测试数据即待预测模块的各个度量元值，得到相应的模块缺陷个数。

1. 所述步骤2的数据预处理包括：对数据集中缺失值使用中位数进行填充；删除重复的数据；对不一致的样本（样本具有相同度量元值，不同的缺陷个数），如果存在其中一种缺陷个数的数量最多，保留一个该样本，否则删除所有不一致样本。
2. 所述步骤3中，模型图如1，输入层, 其中是程序模块的第个度量元的值，是度量元的个数；隐藏层，其中是最大隐藏节点个数, 第个隐藏结点，其中为度量元的个数, 是待求的模型参数，是隐藏层的激活函数，在本发明中，选取函数作为隐藏层的激活函数，即；输出层即预测的模块缺陷个数，。
3. 所述步骤4中
4. 使用平均缺陷百分比(Fault-Percentile-Average，FPA)衡量模型的排序性能，均方误差 (Mean Squared Error, MSE)衡量模型的回归性能，模型参数的L1范数衡量模型复杂度；
5. 使用改进的多目标优化方法优化模型的模型性能，回归性能及模型复杂度，在本发明中，使用带精英策略的非支配排序的遗传算法（Non-dominated Sorting Genetic Algorithm II，NSGA-II）并对其做出了改进，使其在完成多目标优化功能的同时，也承担着特征选择和模型简化的功能。即在算法的种群迭代过程中，每一次子代染色体生成后，对子代中的每一条染色体，其每一个值以一定的概率被置为0，然后子代染色体再参与到后续的算法流程中。
6. 所述步骤5中，模型选择的方法具体为：对模型，其训练集上的性能为别为, 验证集上的性能为，,其中，是模型评估时训练集性能所占比例, ，分别是模型评估时排序性能，回归性能，模型复杂度所占权重。最后依据选取最优模型。

(3) 本发明相对于现有技术具有如下优点及效果

1. 由于以多层感知机为模型，预测模型搜索的关系域包含任意线性和非线性关系，相对于单纯的线性模型扩宽了预测模型的关系域，提高预测的准确性。

2. 由于使用多目标优化算法确定多层感知机的参数，相较于使用梯度下降法等传统迭代优化算法，一方面避免了陷入局部极值，扩大了参数的搜索范围，更大概率找到更优解，另一方面，多目标优化使参数求解时可以兼顾多个优化维度，在优化模型的回归性能的同时，还可以优化模型的排序性能及模型复杂度性能等。8

3. 由于使用置零法改进多目标优化算法，一方面执行了特征选择的功能，筛选掉不重要的特征，一方面在优化多层感知机的权重和偏移的同时，优化模型结构，以获得更优的模型。

**五、附图说明**

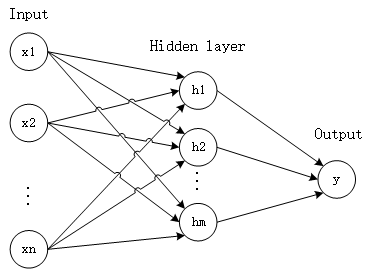


图1 是构造的初始多层感知机模型，包括输入层，隐藏层，输出层共三层。其中是输入层的个输入，是隐藏层的个结点，表示输出层的输出结点。

**六、具体实施方式**

下面结合实施例及附图对本发明4进一步详细的描述，但本发明的实施方式不限于此。

实施例

本实施例中，

除了本实施例提及的方式外，还能做如下变换：多层感知机模型的隐藏层层数增加；多层感知机模型的初始隐藏结点数增加；多层感知机的隐藏层及输出层的激活函数改变；多目标优化方法改变。

上述实施例为本发明较佳的实施方式，但本发明的实施方式并不受上述实施例的限制，其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化、均应为等效的置换方式，都包含在本发明的保护范围之内。

权 利 要 求 书

专利代理公司将会根据说明书完善权利要求书

说 明 书 摘 要

将会根据说明书完善

说 明 书 附 图

（附图可以是工艺流程图、产品结构示意图、选育系谱图等，附图上不能有文字和尺寸、也不要有元件值。

机械类产品或装置类专利应提供该产品或装置装配图（重要的零部件还必须提供详细的零件图）；对机械图，将各部件用引出线通过阿拉伯数字统一标号。

电子类产品或机电相结合的产品必须提供主要的电路图，若在一份图纸难以全面表述，可采用分块表示的形式，或先画出电路原理框图，另外再分别提供详细的电路图；对电路图，在原件旁标注原件符号及顺序标号（如R1、R2）。

软件类技术方案请提供软件的算法流程图，以示出各步骤执行的流程。）

摘 要 附 图

选择说明书附图中最有代表性的一副图置于此