(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112579463 A
(43)申请公布日 2021.03.30

(21)申请号 202011562073.1

(22)申请日 2020.12.25

(71)申请人 北京信息科技大学
地址 100192 北京市海淀区清河小营东路12号

(72)发明人 杨慧文 崔展齐 贾明华 刘秀磊 刘建宾 郑丽伟

(74)专利代理机构 北京睿智保诚专利代理事务所(普通合伙) 11732
代理人 周新楣

(51)Int.Cl.
G06F 11/36(2006.01)

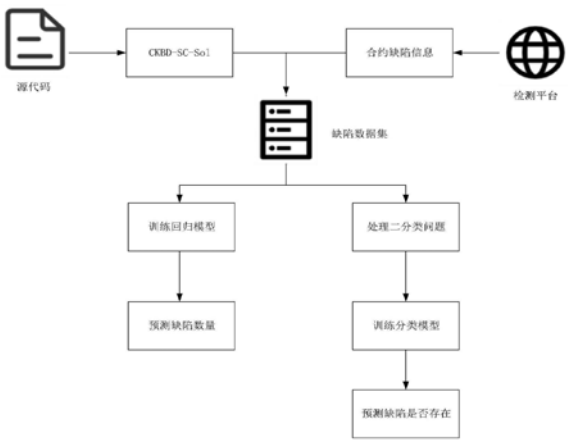
权利要求书1页 说明书4页 附图1页

(54)发明名称

面向Solidity智能合约的缺陷预测方法

(57)摘要

本发明公开了一种面向Solidity智能合约的缺陷预测方法，应用于软件缺陷预测技术领域，首先从Solidity源码中提取代码模块的度量元，并为每个代码模块标记缺陷数量，从而构建缺陷预测数据集；然后针对Solidity缺陷预测数据集中的类不平衡问题，采用过采样方法进行数据预处理；最后分别构建缺陷数量预测模型和缺陷倾向性预测模型，并评估模型的性能。本发明将度量元集与Solidity智能合约缺陷检测结果结合，构建了Solidity智能合约缺陷预测数据集，能够更好地描述Solidity智能合约的特征，基于以上数据集，分别验证了缺陷数量预测和缺陷倾向性预测问题中，不同模型的性能差异。



CN 112579463 A

1.一种面向Solidity智能合约的缺陷预测方法,其特征在于,具体步骤如下:从源代码中提取度量元信息,并获得缺陷信息,结合构成缺陷数据集;

利用回归模型和分类模型对预测Solidity智能合约的进行预测。

2.根据权利要求1所述的一种面向Solidity智能合约的缺陷预测方法,其特征在于,度量元信息包括:面向对象特征、代码复杂度、Solidity智能合约的函数,方法,变量类型,属性以及Solidity语言的限制。

3.根据权利要求1所述的一种面向Solidity智能合约的缺陷预测方法,其特征在于,提取度量元信息的具体步骤:将针对面向对象特征和代码复杂度的CKBD度量元信息与针对Solidity智能合约的函数、方法、变量类型、属性以及Solidity语言的限制的SC-Sol度量元信息进行组合得到CKBD-SC-Sol度量元集。

4.根据权利要求1所述的一种面向Solidity智能合约的缺陷预测方法,其特征在于,获得缺陷信息的具体步骤:根据Solidity缺陷检测信息,按照缺陷类型整理为每个contract/library含有的不同类型缺陷的缺陷数量;对于缺陷数量数据集,每个contract/library的缺陷数量即为各个类型缺陷数量之和;对于缺陷倾向性数据集,将各个类型的缺陷数量二值化,即缺陷数量大于1的contract/library的标签标记为1,否则标记为0。

5.根据权利要求1所述的一种面向Solidity智能合约的缺陷预测方法,其特征在于,利用回归模型预测Solidity智能合约的缺陷数量;其中所述回归模型为线性回归、贝叶斯岭、决策树回归、随机森林回归、K邻近回归中的一种。

6.根据权利要求1所述的一种面向Solidity智能合约的缺陷预测方法,其特征在于,利用分类模型预测Solidity智能合约的缺陷倾向性;其中,所述分类模型为伯努利贝叶斯分类器、高斯贝叶斯分类器、K邻近分类器、决策树分类器、随机森林分类器和支持向量机分类器中的一种。

面向Solidity智能合约的缺陷预测方法

技术领域

[0001] 本发明涉及软件缺陷预测技术领域，更具体的说是涉及一种面向Solidity智能合约的缺陷预测方法。

背景技术

[0002] 区块链是以比特币为代表的数字加密货币体系的核心支撑技术。区块链技术的核心优势是去中心化，为解决中心化机构存在的高成本、低效率以及数据存储不安全等问题提供了解决方案。区块链技术的研究与应用呈现爆发式增长态势，政府部门、金融机构、科技企业和资本市场均等均在探索利用区块链技术解决实际问题的方法。

[0003] 智能合约是区块链的核心构成要素，是一种用算法和程序来编制合同条款、运行在区块链且可按照规则自动执行的数字化协议。智能合约最早于1994年提出，随着区块链技术的出现受到广泛关注。通过编写智能合约可以实现更复杂的应用，从而拓展了区块链的功能。目前，智能合约在传统金融资产，以及社会系统中的资产管理、合同管理等方面发挥作用，如股权众筹，又或基于智能合约制定投票协议等。

[0004] 智能合约在拓展区块链功能的同时，也带来了潜在的安全风险，智能合约的缺陷会对财产造成了巨大的损失，如:2017年11月Parity钱包遭到攻击，导致2.85亿美元的以太币被冻结;2016年6月最大众筹项目TheDAO的300多万以太币被非法转移等，并且与传统软件不同，智能合约在部署后进行补丁修复十分困难，因此智能合约的质量保证技术引起了工业界和学术界的广泛关注。

[0005] 软件缺陷预测是缺陷检测技术的有效补充，软件缺陷预测技术通过分析软件代码或开发过程，设计与缺陷相关的度量元，借助机器学习等方法，预测软件模块的缺陷倾向性或缺陷数量，根据预测结果优化缺陷检测资源的分配，或判断系统的测试充分程度，以及作为软件是否可以交付的依据，以促进软件质量的提高。

[0006] 但据我们所知，在智能合约领域还没有缺陷预测的相关研究。将软件缺陷预测技术应用于智能合约领域，面临以下挑战：

[0007] 尚不存在智能合约缺陷预测数据集。

[0008] 现有的度量元集关注代码复杂性以及面向对象程序的特征，而智能合约作为一种新型的、涉及到金额变动的程序，目前缺少有针对性的度量元集描述智能合约的相关特征。

[0009] 因此，如何提供一种面向Solidity智能合约的缺陷预测方法是本领域技术人员亟需解决的问题。

发明内容

[0010] 有鉴于此，本发明提供了一种面向Solidity智能合约的缺陷预测方法，将度量元集和Solidity缺陷检测结果结合，构建了Solidity智能合约缺陷预测数据集，能够更好地描述Solidity智能合约的特征，基于以上数据集，分别验证了缺陷数量预测和缺陷倾向性预测问题中，不同模型的性能差异。对于缺陷倾向性预测问题，进一步分析了使用过采样技

术处理类不平衡数据集是否会提高预测性能。

[0011] 为了实现上述目的,本发明提供如下技术方案:

[0012] 一种面向Solidity智能合约的缺陷预测方法,具体步骤如下:

[0013] 从源代码中提取度量元信息,并对该源代码进行缺陷检测,得到Solidity缺陷检测信息,将两者信息根据contract/library进行对应组合,构成缺陷数据集;

[0014] 利用回归模型和分类模型对预测Solidity智能合约的进行预测。

[0015] 优选的,在上述的一种面向Solidity智能合约的缺陷预测方法中,度量元信息包括:Solidity智能合约的函数,方法,变量类型,属性以及Solidity语言的限制。

[0016] 优选的,在上述的一种面向Solidity智能合约的缺陷预测方法中,提取度量元信息的具体步骤:将针对面向对象特征和代码复杂度的CKBD度量元信息与针对Solidity智能合约的函数、方法、变量类型、属性以及Solidity语言的限制的SC-Sol度量元信息进行组合得到CKBD-SC-Sol度量元集。

[0017] 优选的,在上述的一种面向Solidity智能合约的缺陷预测方法中,获得缺陷信息的具体步骤:根据Solidity智能合约缺陷检测信息,将其整理为每个contract/library含有的不同类型缺陷的缺陷数量;对于缺陷数量数据集,每个contract/library的缺陷数量即为各个类型缺陷数量之和;对于缺陷倾向性数据集,将各个类型的缺陷数量二值化,即缺陷数量大于1的contract/library的标签标记为1,否则标记为0。

[0018] 优选的,在上述的一种面向Solidity智能合约的缺陷预测方法中,利用回归模型预测Solidity智能合约的缺陷数量;其中所述回归模型为线性回归、贝叶斯岭、决策树回归、随机森林回归、K邻近回归、梯度加速回归和支持向量机回归等中的一种。

[0019] 优选的,在上述的一种面向Solidity智能合约的缺陷预测方法中,利用分类模型预测Solidity智能合约的缺陷倾向性;其中,所述分类模型为伯努利贝叶斯分类器、高斯贝叶斯分类器、K邻近分类器、决策树分类器、随机森林分类器和支持向量机分类器等中的一种。

[0020] 经由上述的技术方案可知,与现有技术相比,本发明公开提供了一种面向Solidity智能合约的缺陷预测方法,首先从Solidity源码中提取代码模块的度量元,并为每个代码模块标记缺陷数量,从而构建缺陷预测数据集;然后针对Solidity缺陷预测数据集中的类不平衡问题,采用过采样方法进行数据预处理;最后分别构建缺陷数量预测模型和缺陷倾向性预测模型,并评估模型的性能。本发明结合Solidity缺陷检测信息,构建了Solidity智能合约缺陷预测数据集,能够更好地描述Solidity智能合约的特征,基于以上数据集,分别验证了缺陷数量预测和缺陷倾向性预测问题中,不同模型的性能差异。

附图说明

[0021] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据提供的附图获得其他的附图。

[0022] 图1附图为本发明的方法流程图。

具体实施方式

[0023] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0024] 一种面向Solidity智能合约的缺陷预测方法，如图1所示，具体步骤如下：

[0025] 从源代码中提取度量元信息，并对该源代码进行缺陷检测，获取缺陷检测信息，根据contract/library将两者信息对应组合，构成缺陷预测数据集；

[0026] 利用回归模型和分类模型对预测Solidity智能合约的进行预测。

[0027] 具体地，首先，从源代码中提取度量元信息，即CKBD-SC-Sol，并结合Solidity缺陷检测结果，组成称为Solidity智能合约缺陷数据集。

[0028] 第二步，对于Solidity智能合约的缺陷数量预测问题，应用7种回归模型，即线性回归、贝叶斯岭、决策树回归、随机森林回归、K邻近回归、梯度加速回归和支持向量机回归。

[0029] 第三步，对于Solidity智能合约的缺陷倾向性预测问题，应用6种分类模型，即伯努利贝叶斯分类器、高斯贝叶斯分类器、K邻近分类器、决策树分类器、随机森林分类器和支持向量机分类器。

[0030] 进一步，度量元信息包括：面向对象特征、代码复杂度、Solidity智能合约的函数，方法，变量类型，属性以及Solidity语言的限制。

[0031] 为了进一步优化上述技术方案，提取度量元信息的具体步骤：将针对面向对象特征和代码复杂度的CKBD度量元信息与针对Solidity智能合约的函数、方法、变量类型、属性以及Solidity语言的限制的SC-Sol度量元信息进行组合得到CKBD-SC-Sol度量元集。

[0032] 进一步，由于尚不存在Solidity智能合约的缺陷预测数据集，因此为了构建缺陷预测模型，首先从Xblock和Etherscan获得Solidity智能合约的源码，使用AST分析工具solidity-parser-antlr提取Solidity智能合约中的CKBD-SC-Sol度量元集，将提取出的CKBD-SC-Sol度量元集信息与对应的缺陷检测信息组合，构建Solidity缺陷预测数据集。

[0033] 为了进一步优化上述技术方案，获得缺陷信息的具体步骤：区块链智能合约检测平台通过输入Solidity源代码或以太坊合约地址，经过区块链智能合约检测平台分析检测后会输出缺陷类型以及对应的代码行号，将区块链智能合约检测平台输出的缺陷报告按照缺陷类型整理为每个contract/library含有的不同类型缺陷的缺陷数量；对于缺陷数量数据集，每个contract/library的缺陷数量即为各个类型缺陷数量之和；对于缺陷倾向性数据集，将各个类型的缺陷数量二值化，即缺陷数量大于1的contract/library的标签标记为1，否则标记为0。

[0034] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

[0035] 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下，在其它实施例中实现。因此，本发明

将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

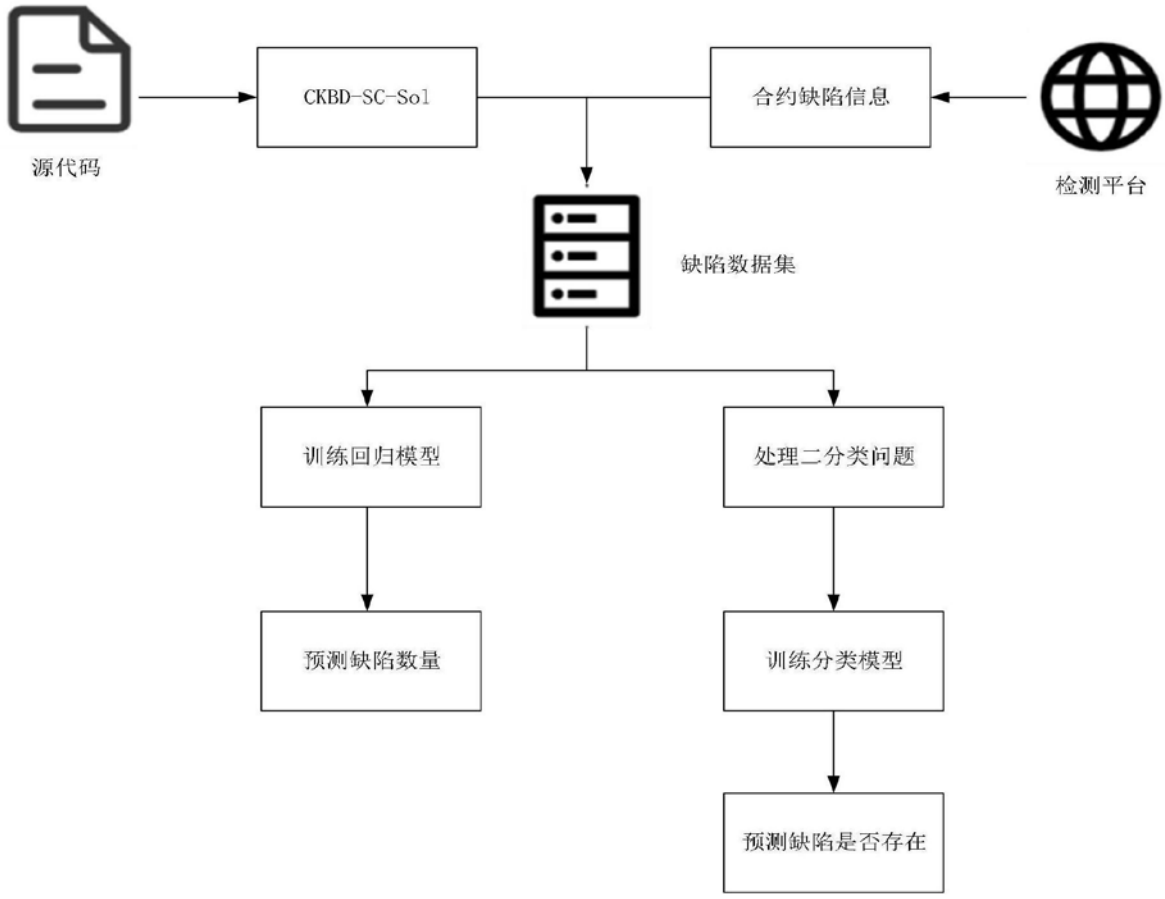

源代码
CKBD-SC-Sol
合约缺陷信息
检测平台
缺陷数据集
训练回归模型
处理二分类问题
预测缺陷数量
训练分类模型
预测缺陷是否存在

图1