

基于机器学习的软件缺陷预测

黃柏曠

2020 年 10 月 5 日

1 摘要

随着虚拟仪器软件系统规模逐渐扩大, 其复杂程度也逐渐提高, 对于缺陷容忍度较低的高风险软件来说发生故障导致的后果很严重, 虚拟仪器软件的性能面临严峻的考验。[1] 软件缺陷预测是软件工程领域中与软件质量保证密切相关的重要的研究课题, 它对提高软件系统质量和优化测试资源分配都有重要意义。在软件工程数据挖掘领域中, 基于机器学习的静态软件缺陷预测根据软件历史仓库数据, 采用缺陷相关的度量对软件代码或开发过程进行分析, 利用机器学习方法来预测软件项目中待测试程序模块的缺陷倾向性或缺陷数量。[2]

2 關鍵字

软件缺陷、机器学习、预测

3 前言

IEEE729-1983 对软件缺陷进行了定义：一方面是指系统开发或维护过程中存在的错误、毛病等各种问题；另一方面是指系统没有达到客户需求的某种功能的失效。而软件缺陷的危害是巨大的，在软件开发早期的软件缺陷预测就很重要了。[4] 软件缺陷预测技术是指按照软件的基本属性规模、复杂性、开发方法和过程，以及已知缺陷来预测潜在但还未被发现的

缺陷 [5]。软件缺陷预测技术能帮助测试人员掌握软件失效模式、了解质量状态，并决定软件是否交予用户使用。软件缺陷预测技术可分成动态缺陷预测与静态缺陷预测 [6]。动态缺陷预测技术是以缺陷产生时间为基础，对系统缺陷随时间分布实施预测的技术；该技术利用时间分布统计、挖掘软件的缺陷，寻找缺陷基于软件开发周期的引入与移除规律。静态缺陷预测技术是指采用软件规模、复杂度、开发过程等可度量缺陷的元素及已有缺陷，预测软件潜在但还未暴露的缺陷；该技术以缺陷尽早检测为原则，既可减少缺陷修复成本，又能缩短缺陷修复时间。[7]

4 緒論

4.1 研究背景

作为计算机实现各项功能，辅助人们进行各项活动的载体，在当今世界各行各业中发挥着重要的作用。然而伴随各类软件的开发，软件缺陷的产生是不可避免的。关于软件缺陷 (defect) 的定义，学术界和产业界还有其他与之相关的术语，例如错误 (error)、故障 (fault)、失误 (mistake)、失效 (failure) 等。其中缺陷是软件中已经存在的部分，且可以通过修复进行消除。错误是一种人为的行为，如开发过程中无意识的技术错误，其后果必定导致软件某部分功能产生缺陷。失效和故障都是软件运行时产生的，失效是软件不能再完成规定的功能，而故障是指软件在运行时没有输出与设计相符合的结果。除人为操作失误以外，软件缺陷正逐渐成为导致计算机系统失效和停机的主要因素。在某些对可靠性有着严格要求的行业，譬如航空航天、医疗管理、金融系统等，如果潜在的软件缺陷得不到及时的清除，则可能造成很大的损失，例如：1999 年 NASA 在制造火星气候轨道探测器时在程序中使用了英制单位，而不是预定的公制单位，导致探测器的推进系统无法按照预设轨道运作，使其进入大气层的高度有误，最终瓦解碎裂并造成了 3.27 亿美元损失。软件缺陷预测技术通过对软件开发过程或者软件代码进行分析，以建立相应的预测模型，并对软件中潜在的缺陷进行预测，其结果可以为测试人员提前定位可能产生缺陷的模块或者提前预知模块中所含的缺陷数目，从而帮助和指导决策者分配有限的资源以进行更有针对性的软件测试。软件缺陷预测在软件工程中具有重要的意

义：(1) 使软件开发人员缩短开发周期，节省资源，更快的开发出高可靠性的软件；(2) 使测试过程和资源能更集中地针对易产生缺陷的模块，从而加快缺陷发现和修复的速度，从而降低软件开发的成本；(3) 通过提高测试的过程以提高软件的质量。

5 結論

6 參考文獻

- [1] 曾路, 汪浩. 基于机器学习的虚拟仪器软件缺陷预测模型研究 [J]. 自动化与仪器仪表, 2020(05):59-62.
- [2] 张志武. 基于机器学习的软件缺陷预测方法研究. 2018.
- [3] Metric-based software reliability prediction approach and its application[J] . Ying Shi, Ming Li, Steven Arndt, Carol Smidts. Empirical Software Engineering . 2017 (4)