

基于软件过程的软件可靠性预测运作模型^{*}

蒋永辉^{1,2}, 吴洪丽^{2,3}, 王政霞⁴

- (1. 海南大学 信息科学技术学院, 海南 海口 570228 ;
2. 海南师范大学 信息科学技术学院, 海南 海口 571158;
3. 中国科学院 成都计算机应用研究所, 四川 成都 610041;
4. 重庆交通大学 信息与计算科学系, 重庆 400074)

摘要: 提出了基于软件开发过程数据, 构造用于可靠性预测的过程数据库, 并分析了此数据库的内容与运作模型, 以期得到更加可信、准确的软件可靠性预测结果, 为后续的研究奠定数据基础。

关键词: 软件过程; 软件可靠性; 过程数据库

中图分类号: TP311.5

文献标识码: A

Software reliability prediction operation model based on software process

JIANG Yong Hui^{1,2}, WU Hong Li^{2,3}, WANG Zheng Xia⁴

- (1. College of Information Science & Technology, Hainan University, Haikou 570228, China;
2. College of Information Science and Technology, Hainan Normal University, Haikou 571158 China;
3. Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610041, China;
4. Department of Information and Computing Science, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: This paper presented process database used to predict the software reliability which was based on the data of software development process and analyzed the contents of this database and operating model in order to be more credible and accurate software reliability prediction and lay the data base for follow-up studying.

Key words: software process ; software reliability ; software process database

从 20 世纪 70 年代起, 软件可靠性预测技术大体上分为静态和动态两种。早期的静态预测主要研究可靠性和软件规模、程序复杂度等基本属性之间的关系, 以此预测软件的可靠性。20 世纪 90 年代初, 人们发现缺陷在软件中是随机分布的, 既而出现了针对缺陷分布的预测技术, 也有一些取得了比较好的成果和应用。例如基于软件规模的 Halstead 模型以及 Lipow 在此模型机上改进所得到的软件缺陷与可执行代码行之间的关系模型; Takahashi 将代码与具体的文档数量对应起来, 给出了缺陷密度的估计; Malaiya 等在假设模块规模符合指数分布的情况下, 提出了缺陷估算的公式: 缺陷密度 $D(s) = a/s + b + cs$, 指出软件缺陷取决于程序员能力、过程成熟度以及测试程度等。20 世纪 90 年代, 随着软件规模和复杂度的不断增长, 同时由于面向对象技术的出现, 许多基于面向对象度量元的缺陷预测技术涌现出来。而随着软

件过程技术的发展, 人们逐渐认识到了软件过程对于软件可靠性的影响^[1-6]。不难发现, 静态预测主要是基于缺陷相关的度量数据, 对缺陷的数量或者分布进行预测的技术; 而动态预测则是基于缺陷或者失效产生的时间, 对系统缺陷随时间的分布进行预测的技术。这些现有的软件可靠性预测模型一般只是对与缺陷有关的度量数据和软件可靠性测试后的失效数据进行分析, 而忽略了软件开发过程中大量有用信息。因此, 笔者提出了基于软件过程数据, 构造用于可靠性预测的过程数据库, 并分析了此数据库的内容与运作模型, 以期得到更加可信、准确的软件可靠性预测结果, 为后续的研究奠定数据基础。

1 软件过程与软件可靠性

软件过程可以定义为人们用来开发和维护软件以及相关产品(如工程计划、设计文档、规章、检测事例及用户手册)的一组活动、方法、实践及转换^[7]。软件可靠

^{*} 基金项目: 重庆自然科学基金(CSTC 2009BB2375), 重庆交通大学青年教师基金 2005(35)

软件天地 Software Technology

性是软件系统在规定时间内及规定的环境条件下,完成规定功能的能力。在 ISO/IEC 9126 的软件质量特性表述中,软件可靠性(Software Reliability)是软件质量的一个重要特性。而软件产品是软件过程的产物,软件过程的好坏决定了软件产品的好坏,软件过程的稳定也决定了软件产品质量的稳定。软件产品质量的稳定直接反映了软件可靠性的稳定。SEI 提出的 CMM 以及 CMMI 里也指出,软件过程能力成熟度反映了一个软件企业的过程能力以及开发能力,过程稳定的企业或者团队,其软件产品的质量也是相对稳定的。据此,不难得到这样的结论:稳定的软件过程,其软件产品的可靠性也是稳定的^[8]。

2 影响软件可靠性的因素

在对 13 个公司的调查结果进行分析后,参考文献[9]提出了基于经验数据的影响软件可靠性的 32 个因素。这 32 个因素贯穿了软件开发的各个阶段,其中最重要的是软件复杂度、程序员的技能、测试工作量、测试覆盖率、测试环境和程序规格说明书的改动频率 6 个因素^[9]。参考文献[10]指出了影响软件可靠性的主要因素包括:需求的变更、开发团队的经验、设计的变更、编码和测试、新技术、开发语言以及工具的使用、管理经验、高层管理的支持力度、度量和模型的使用情况^[10]。参考文献[11]认为开发的复杂程度、代码重用的比率以及软件开发团队的经验和受教育程度是影响软件可靠性的主要因素。参考文献[12]指出影响软件可靠性的因素可分为两大类:缺陷引入和缺陷检测。缺陷引入又由 16 个因素决定,而缺陷检测由 17 个因素决定。这些因素中,与过程有关的因素包括开发人员的能力、领域知识、团队协作、团队结构、管理能力、程序复杂度、沟通能力、项目管理、过程管理、变更控制、文档质量、需求质量、开发环境、计划偏移度、开发过程成熟度、产品调研等^[12]。软件过程度量的三大类基本内容分别是质量、时间和资源,质量表示在不同的过程运作阶段过程产品和过程执行有关的质量信息(例如软件问题及其生命周期,过程执行的偏离度等),时间表示过程运作中各项活动的计划时间和实际时间,资源表示在过程运作中不同阶段活动消耗的物资资源和人力资源。综合分析这些文献提出的影响软件可靠性的因素,不难发现,其中,很多因素也是软件过程度量所关注的信息。

3 软件可靠性预测度量体系

在已有的调查研究基础之上^[13-15],笔者构建了基于软件过程的软件可靠性预测度量体系。该体系模型共划分为 6 个层次:影响软件可靠性因素、影响软件可靠性因子、特性、子特性、度量以及度量元。其层次结构如图 1 所示。其中,软件可靠性取决于影响软件可靠性的性能如何,而对软件可靠性因素的评价表现在多个可靠性因子,不同的可靠性因子的影响基于不同的性能特性,特性可能有子特性,通过度量来评价可靠性因子的

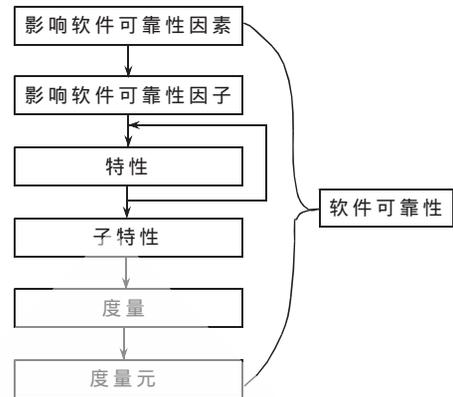


图 1 模型层次结构

特性和子特性,同时度量又有多种数据表现形式,即多个度量元。

对软件可靠性的预测,笔者确定了 6 个影响软件可靠性的主要因素,分别是:一般性因素、稳定性因素、过程依从性、开发文档、编码和人力资源因素。软件测试因其产生的数据是软件失效后的数据,更多的用于软件可靠性的动态预测,故此不予以考虑。对这 6 个可靠性因素的评价,笔者建立了如表 1 所示的度量和度量元,并据此来构建基于软件过程的软件可靠性预测度量体系。

4 基于软件过程的软件可靠性预测

一个过程稳定的软件企业,其软件产品的可靠性也是稳定的。通过对影响软件可靠性的过程因素的分析,笔者确定了上述的度量和度量元。而软件企业若考虑从历史数据中得到关于以前项目的过程等信息,就必须拥有过程数据库。过程数据库的建设需要人员和软件工具的支撑,如图 2 所示。该过程数据库建设模型结合前述分析,并综合考虑了软件度量技术和 CMM/CMMI 中的度量技术后抽象而成。

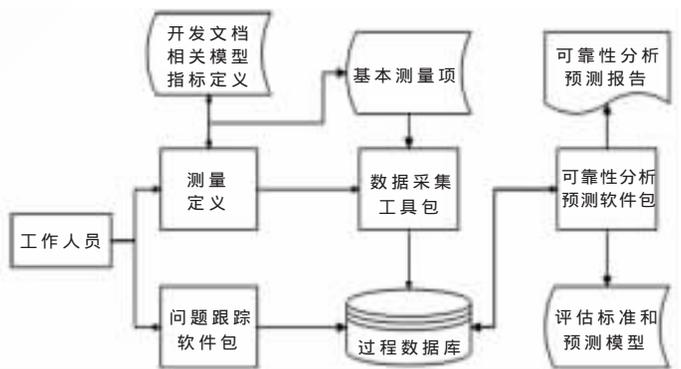


图 2 用于可靠性预测的软件数据库建设模型

根据企业的测量定义,工作人员可以知道应该采集什么数据,并通过数据采集工具包使测量数据进入过程数据库系统;另外工作人员也可以通过问题跟踪软件包(例如需求跟踪系统)直接使测量数据进入过程数据库。

表 1 基于软件过程的软件可靠性度量体系

软件可靠性因素	度量	度量元
一般性因素	开发环境	处理器,存储设备,输入/输出设备,通信设备,系统软件
	软件复杂度	Halstead、Macbe,节点数,路径数,产品的创新度,代码重用比例
	项目环境	软件类别,开发难度,编程语言,开发工作量,时间限制
	创新	新技术的使用、开发语言以及工具的使用
稳定性	需求变更请求及控制	变更数目、类型、状态、来源等,需求项数目、型、文档页数等
	设计变更以及变更控制	设计的规模大小,设计的变更数量、状态、频度,受影响的设计变更数量
	代码变更	代码规模,代码类型,代码变更数量、状态、版本、频度,增加软件单元,修改软件单元,删除软件单元
	文档变更	文档规模,文档产生时间,文档状态,文档负责人,变更频度
	数据库变更	数据库结构,数据库规模,数据库数据规模,数据库复杂度
	过程变更请求	变更数目,变更类型,变更时间,变更频度,变更来源,变更状态
过程依从性	度量和模型的使用情况	阶段评价标准
	过程偏差	偏差数目、状态,过程规范和组织方针数目,过程规范和组织方针类型,评审时间,评审方式,评审频度
	过程管理成熟度	开发活动的范围和职责的定义,开发活动的权利和义务的定义,项目进展报告的准确程度,是否坚持计划,项目计划和管理中的度量精度,项目计划和管理过程中历史数据的有效性,项目风险的估计,开发计划的执行情况,项目管理任务的执行,测试是否有计划
开发文档	文档质量	支持文档的创建贯穿在整个生命周期,系统文档的品质和完整性,需求文档的质量,设计文档的质量,项目文档的详细程度
编码	程序员技能	开发人员关于开发过程的相关知识,项目平台/环境的经验,相似开发环境的经验,对开发工具和开发语言的经验,开发人员关于支持工具的使用
	程序员的组成结构	编程经验,开发人员的技术水平,设计人员的技术水平,开发团队的经验,开发团队的能力,对关键成员的依赖,开发成员时间的有效性,技术人员使用的有效性,专家的使用,团队大小,组成开发团队的复杂程度,开发所涉及的团队数目,开发队伍的规模
	人员投入	人员投入数量,工作时间,人员工作对象
	人员变更情况	人员角色变更,人员数量变更,人员技术能力变更
人力资源因素	团队协作	开发人员的工作压力,开发人员和团队对项目环境的适应性,开发团队的协作工作能力,开发团队人员的稳定性,开发人员的态度,相关部门对开发团队的支持度,客户对开发团队的支持度,最终用户对开发工作的支持度
	开发团队的经验和教育程度	项目平台/环境的经验,对开发工具和开发语言的经验,开发人员关于支持工具的使用,开发相似应用的经验,从事相关领域的工作时间,从事相关领域的成功项目数,相关领域的角色

过程数据库中除存储了影响软件可靠性的因素的度量值外,还应包含理解和解释实际度量数据、评价其合理性和适用性所需要的信息和索引,以及可靠性分析的标准、模型和基线值、优先级等。

通过对过程数据库里的基本测量项数据进行分析计算,得到关于以往项目的软件可靠性相关的度量,此度量数据经过评估标准和预测模型分析,形成当前软件产品的可靠性分析预测报告。

基于知识和数据挖掘的软件可靠性预测是现在和未来几年研究的方向。本文首先分析讨论了软件过程与

软件可靠性的关系,然后分析了影响软件可靠性的因素,最后讨论了建立基于软件过程数据的过程数据库,并抽象出运作模型,为后续的基于知识和数据挖掘的软件可靠性研究奠定数据基础。软件可靠性预测中一些其他重要问题,如软件可靠性预测环境的构建、软件可靠性预测模型的选择以及定量评价数据的验证等问题,还有待进一步探讨。

参考文献

- [1] HALSTEAD M H. Elements of software science. New York: Elsevier, North-Holland, 1977.

《微型机与应用》2010年第6期

- [2] OTTENSTEIN L. Predicting numbers of errors using software science. ACM SIGMETRICS Performance Evaluation Review, 1981, 10(1):157-167.
- [3] OTTENSTEIN L. Quantitative estimates of debugging requirements. IEEE Trans. on Software Engineering, 1979, SE-5(5):504-514.
- [4] LIPOW M. Number of faults per line of code. IEEE Trans. on Software Engineering, 1982, 8(4):437-439.
- [5] TAKAHASHI M, KAMAYACHI Y. An empirical study of a model for program error prediction. IEEE Trans. on Software Engineering, 1989, 15(1):82-86.
- [6] MALAYIA Y, DENTON J. Module size distribution and defect density. In: Proc. of the 11th Int'l Symp. on Software Reliability Engineering. New York: IEEE Computer Society Press, 2000. 62-71.
- [7] ZAHRAAN S 著. 软件过程改进[M]. 陈新, 译. 北京: 中信出版社, 2002.
- [8] 吴洪丽. 支持软件过程改进的软件过程度量研究[D]. 重庆: 重庆大学计算机学院, 2004.
- [9] ZHANG X, PHAM H. An analysis of factors affecting software reliability. The Journal of Systems & Software, 2000, 50(1):43-56.
- [10] SCHNEBERGER S L. Distributed computing environments: effects on software maintenance difficulty. Journal of Systems and Software, 1997, 37(2):101-116.
- [11] EVANCO W M, LACOVARA R. A model-based framework for the integration of software metrics. Journal of Systems and Software, 1994, 26(1):77-86.
- [12] JACOBS J J, MOLL V. Identification of factors that influence defect injection and detection in development of software intensive products. Information and Software Technology, 2007, 49(7):774-789.
- [13] 王铁江, 郦萌. 软件可靠性的模糊综合评估模型[J]. 计算机工程与应用, 2002, 38(20):23-26.
- [14] 田涛, 王昕. 一种基于模糊综合评判的软件可靠性模型选择方法[J]. 空军工程大学学报: 自然科学版, 2002, 3(2):56-59.
- [15] 马飒飒, 陈自力. 软件可靠性定量评估过程研究[J]. 计算机测量与控制, 2005, 13(5):503-505.

(收稿日期: 2009-12-15)

作者简介:

蒋永辉, 男, 1979年生, 硕士研究生, 主要研究方向: 软件过程技术与方法, 文本挖掘。

吴洪丽, 女, 年生, 博士研究生, 主要研究方向: 软件过程技术与方法、软件可靠性。

王政霞, 女, 1977年生, 博士研究生, 主要研究方向: 非线性系统, 软件工程。