文章编号：1009－3486(2002)04－0001－06

软件质量保证技术研究综述①

贲可荣[1,2]

(1.海军工程大学 管理工程系，湖北 武汉 430033；2.软件工程国家重点实验室，湖北 武汉 430072)

摘　要：概述了软件可靠性研究的必要性，从软件开发过程管理、质量保证标准、测试技术、可靠性评估与预测以及形式化规范和验证等5个方面给出了软件质量问题的解决办法，并对所研究的成果做了介绍.最后指出了在统计测试、形式化方法方面有待进一步研究的问题.

关键词：软件质量；能力成熟度模型(CMM)；软件标准；软件测试；软件可靠性；形式化方法

中图分类号：TP311.5；TP31.2　　文献标识码：A

1　研究背景

计算机系统的可靠性涉及硬件和软件两个方面.今天的计算机系统，其硬件质量越来越高，而软件质量问题却变得更加突出[1].

作战管理软件系统有自身的特点，比如：要求系统能够认别、跟踪定向目标，而这些目标的弹道特性在战前是未知的；通过联系于传感器和武器的计算机网络来进行计算，但是，因攻击者的反测量导致通信行为不可能预测，在实际使用之前，不可能在真实条件下测试系统；系统服务期很短暂，几乎不能容许人的交互，在服务期间，不能进行测试和程序修改；要求在一确定时限内，实时完成；无法预测用于计算的资源的有效性，武器系统包含大量传感器和武器，这些传感器和武器的绝大多数本身就需要大而复杂的软件系统，武器和传感器的类型随着开发而增加，武器和传感器的特征处于可变状态[2].

以上特性，注定了作战管理软件的可靠性的本质困难.但是，国外许多高技术条件下作战系统可靠地运行，说明了软件可靠性达到了设计要求.

资料表明，净室软件工程法能够保证交付的软件具有已知的MTTF，能开发出几乎没有缺陷的软件，可以以低代价生产高质量的软件[3].使用任务剖面、统计测试、测试用例生成工具、组合模型、评估及预测工具等可靠工程方面的成熟技术和积累的经验，可以缓解我国目前软件可靠性存在的严重困难，使开发的软件达到预期的要求.

因此，为我国我军高可靠性系统推荐适当的开发方法，制定保证可靠性的策略和标准，开发可靠性测评工具，非常必要.

2　解决问题的途径

2.1　软件开发的过程管理

由美国国防部资助研究的、用于评估软件开发机构的工程和管理水平的“软件能力成熟度模型”(CMM)，于1991年公布.CMM是由美国Carnegie Mellon大学软件工程研究所(CMU/SEI)的Humphrey等人在1987年前后提出的[4].

CMM是一个明确的框架，软件工程师可以使用、讨论和扩展，用一种开放的思想去看待CMM，而不

① 收稿日期：2002-01-25；修订日期：2002-04-06
基金项目：教育部骨干教师基金以及武汉大学软件工程国家重点实验室资助项目
作者简介：贲可荣(1963-)，男，副教授，博士.

万方数据

要被它所束缚.把注意力集中于目标,把详细具体的实践当作实例对待,什么合适使用什么.

能力成熟度模型的引进势必为我国的软件产业提供一个更有效的设计和认证思路,为进一步落实软件质量保证政策提供有效途径[5,6].

在小型软件机构中采用 CMM 和 ISO9001 已经成了软件工业的一种挑战.文献[7]总结了小机构面临的困难,提出了解决问题的建议和小型软件机构的过程改进模型.

文献[8]提出了 CMM 过程评估模型并给出了符号表达,用伪码程序描述了判断过程能力等级的算法.

1999 年底,在美国 SEI 召开了高成熟度研讨会.文献[9]概述了与会 26 个组织的状况,介绍了部分组织的投资回报数据、特有实践活动以及达到高等级遇到的问题.

文献[10]对基于软件能力成熟度模型(CMM)的传统软件开发方法与净室技术进行了比较,阐述了如何将净室技术引入到一个基于 CMM 的软件开发框架之中,介绍了从传统方法向净室技术转化的一些方法.

2.2 软件质量保证标准

在软件可靠性质量和软件可靠性管理等方面,国际电子委员会负责的 TC56 技术委员会在 1985 年成立了软件可靠性工作组,并制定了软件可靠性和维修性管理规范;美国国防部不惜重金制定了一整套软件可靠性检验标准,1994 年春天,美国向月球轨道上发射的名为 Clementine 的卫星,由于按这些标准对它的软件进行了严格的检验,从而确保了卫星正常工作;美国国防部还推荐了几种可靠性模型,例如,Schneidewind 模型,它已用于 NASA 航天飞机机载可靠性评估、三叉戟自控系统可靠性预计和分析方面的研究.

在国家 863 项目“软件可靠性的关键技术与工具”中,借鉴美国 IEEE Std 1012 – 1987,IEEE Std 1028 – 1988,IEEE Std 1062 – 1993,提出了与我国我军软件可靠性标准相容的“软件验证和有效性标准”、“软件评审与审查标准”和“软件验证和有效性计划指南”.

文献[11]指出了如何测试文档,概述了最新的文档过程成熟度模型,探讨了成熟度等级、关键实践、评分方案和评估报告及发生变化的原因,指出正是关键实践的满意度推动了评估并为过程改进提供了所需关键信息.

2.3 软件测试技术

软件测试研究相当活跃,目前有 4 个主要方向,即验证技术、静态分析、动态测试和测试实例生成,提出了众多测试方法和测试工具,陆续推出了 Ada, C, FORTRAN, COBOL 等程序设计语言的程序测试环境,并且美国已开始了对软件测试可视化技术的研究[12].

软件测试的统计方法首先由 Harlan Mills 和 IBM 的同事开发出来.后来,John Musa 和 AT&T 的同事也开发了相似的方法.在工业界,典型地采用一组产品抽检协议进行合格性检验:首先,进行随机抽样,并应用投入运行使用中的一些测试特征进行测试,然后,进行分析和统计推断,最终,通过一定标准的产品才被认为是合格产品[3].

净室软件测试和认证方法——基于使用模型的统计测试——是这种协议在软件上的一种应用.统计测试时,需要开发出软件投入运行时的使用模型,测试用例由该使用模型随机产生,然后,按照数学和统计学模型对结果进行分析,获取软件的质量度量,并判断测试的充分性.传统的结构化测试方法是净室统计使用测试方法的一种补充,因此,不必放弃该方法.不过,大量实践表明,基于使用模型的测试更经济有效,并且能获得实用软件的高可靠性.

软件系统的基于使用模型的统计测试提供了软件产品和过程质量的度量标准,它将用于软件的整个生命期的管理和决策.由于使用模型是基于规范而不是基于代码的,因此源于模型构筑的洞察可用于产生在工程的早期阶段避免出现问题的有价值的管理决策.使用建模和统计测试具有如下主要优点:需求确认,资源和进度评估,人工挑选非随机测试用例,自动生成测试用例,有效地、高效地测试,聚焦测试

万方数据

(有偏抽样),量化测试管理,可靠性评估.

文献[13]对软件项目使用的各种测试方法进行了统计和分类,讨论了测试方法与应用规模之间的关系,功能点与测试实例数、测试人员数之间的关系,以及测试方法与测试人员组成之间的关系,分析了测试纠错效率.

文献[14]综述了基于逻辑谓词的软件测试策略,比较了这些策略的侦错效率.得到结论:布尔算子(BOR)测试选择策略在侦测谓词错误方面是实用和高效的.

文献[15]介绍了判断图形用户界面(GUI)行为和所给测试用例的预期行为是否一致的喻示方法.喻示法应用了 GUI 形式化模型,该模型用对象、对象属性和动作的集合来描述.若给定形式化模型和一个测试用例,喻示法自动为测试用例中的每个动作导出预期状态.我们从 GUI 的执行中获得对象和它们的属性,从而得到执行中的 GUI 的实际状态.使用从执行监视器获得的实际状态,喻示法自动比较每个动作产生的实际状态和预期状态,从而验证 GUI 对测试用例的正确性.

针对净室软件工程方法中的统计测试用例生成,国防科技大学于 2001 年开发了一软件支撑环境,支持软件的使用模型建立、分析、可靠性测试用例的生成以及可靠性测试的充分性度量.

2.4 可靠性评估与预测

软件可靠性的理论研究开始于 20 世纪 70 年代.现有的理论工作主要集中于软件可靠性建模及软件容错两个方面,但至今未能形成自己的理论体系.在已有的数十种软件可靠性模型中,还没有证明其中哪一个是广泛适用的.软件容错则基本沿用硬件容错的思路,没有形成自己的有效方法.相对而言,软件可靠性的工程实践却取得了长足的进步.目前,应用软件可靠性工程的机构包括 AT&T、Bellcore、NASA JPL、NASA 航天飞机计划、Lockheed-Martin、Lucent Technologies、Saab Military Aircraft、Microsoft、Motorola 等.

国内的软件可靠性研究始于 20 世纪 80 年代初,是在 70 年代国际软件可靠性研究高潮之后.在软件可靠性理论方面取得了一定成绩,但仅集中于软件可靠性建模方面,与国外相比有一定差距.国内在软件可靠性工程实践方面做了有益的工作,开发了一些特定的软件测试和可靠性评估环境."九五"期间我国大力开展了软件可靠性的研究和实践,取得了一批成果.

基于统计和 Markovian 链理论,文献[17]给出了计算机系统可靠性分配方法.该方法能够将整个系统的可靠性指标分配给子系统,进而分配给计算机系统和软件模块.本文阐述了基于该方法的工具.

模块化的软件系统的质量不仅与模块的可靠性有关,而且与软件的操作剖面有关.文献[18\]提出了软件系统模块的重要度概念,给出了依据模块的重要性对软件系统进行可靠性分配的方法.

许多主要的非齐次泊松模型的故障强度都满足顺序约束条件.在顺序约束条件下,文献[19\]提出了故障强度的约束极大似然估计(RMLE),并讨论了其性质.利用故障强度的 RMLE,获得了软件可靠性模型参数的加权最小二乘估计.

2001 年,国防科技大学开发了一可靠性工程环境.该环境抽象了软件可靠性模型的公共 API,能够定性和定量地分析软件可靠性模型对软件特征和失效数据的适用性,提供了基于模型组合方式的定制的软件可靠性模型的综合和智能优化.

2.5 形式化规范和验证

硬件和软件的功能不断增强,规模不断扩大,使得系统复杂性大大提高,由此也使得出错的可能性增大了,而且这些错误可能引起金钱、时间乃至生命的重大损失.软件工程的一个主要目的就是使开发者建构复杂但可靠的系统.达到这个目标的方法之一就是使用形式化方法.形式化方法是一种基于语言、技术及工具的数学方法,用于规范软件系统.形式化方法不能先验性地保证正确性,然而它能通过揭示系统的不一致性、模糊性及不完整性,极大地提高我们对一个系统的理解程度[20].

对软件规范而言,工业部门通过试验发现一些概念(如 Z)可以更精确地表达系统的某些性质.对硬件验证而言,工业界可采取诸如模型验证和定理证明来完善传统的方法.在这两个领域,研究者和实践者正在进行更大规模的研究,并从形式化方法中获益.

万方数据

(1) 规范

规范是描述一个系统及其特性的过程.形式化规范使用一种带有数学定义的语法和语义的语言.这类系统特性包括功能行为、计时行为、性能特征及内部结构.迄今为止,用规范来刻划行为特征最为成功.发展趋势之一是集成不同的规范语言,每一种语言处理系统的一个方面,另一趋势则是处理系统的非行为方面,如性能、实时限制、安全策略及结构设计.

一些形式化方法(如 Z、VDM 及 Larch)注重对顺序系统进行行为规范,状态用丰富的数学结构(如集合、关系、函数)来描述,状态转化由前置及后置条件给出.另一些方法(如 CSP、CCS、状态图、时态逻辑、I/O 自动机)针对并发系统的行为进行规范描述.状态一般遍及简单域(如整数),行为用事件的序列、树、偏序来定义.其他的如 RAISE、LOTOS 则结合了不同的两种方法,一方面处理丰富的状态空间,另一方面,处理由并发引起的复杂性.这些系统的共同之处就是都使用了抽象及合成的数学概念.

规范的过程就是精确记载事件的行为,这样做的好处是无形的——获得对需要规范的系统深刻理解.通过这个规范过程,开发者可以揭示设计缺陷、不一致性、模糊性及不完整性.这种行为的有形的副产品就是一个本身可以被形式化分析的人工产品,例如,检查内部一致性,或用来导出所规范系统的其他特性.规范是用户及设计者,设计者及实现者,实现者及测试者之间的一个有效沟通文档.它可以充当系统源代码的文档,不过是处在描述的更高层.

(2) 验证

发展较为完善的验证方法是模型检查和定理证明,比规范要先行一步.这些形式化方法可用于分析系统的特性.

模型检查技术依赖于构建一种有限系统模型并检查该模型的特性.由于系统状态有限,因此这种穷举状态空间检查是保证可以结束的.模型检查的技术难点在于设计能够处理大型搜索空间的算法和数据结构.模型检查主要用于硬件及原型验证.当前趋势是使用这种技术来分析软件系统的规范.模型检查的两种通用方法均用于实践.第一种,时态模型检查,是 20 世纪 80 年代由 Clarke 和 Emerson 分别独立开发的.这种方法中,规范要用时态逻辑来表达,系统以有限状态转换系统为模型.同时可以使用一种有效过程来检查某种有限状态转换系统是否符合这种规范.第二种方法中,规范是作为一种自动机给出的,系统也作为一种自动机来建构模型,并可与规范比较来决定它是否符合规范.

与定理证明相比较,模型检查是完全自动化的且速度较快,有时几分钟内就能得到结果.模型检查可以检查部分规范,即使未对系统的正确性作完整说明.模型检查的最大好处是提供反例,从而说明设计中的细微错误,并指导排错.

定理证明是一种推理技术,其中系统及其特性均用某种逻辑公式表达.这种逻辑由一组公理和推导规则给出.定理证明就是一个从系统的公理推导某一特征的证明过程.证明过程中引用了公理和规则,并尽可能采用导出定义和中间引理,尽管可以手工证明,但我们强调的是机器辅助证明.在硬、软件中对安全性要求较高的特性的机器验证上,定理证明器的使用不断增多[23].

定理证明器可以大致按高度自动化的通用程序到特定的交互系统来分类.自动化了的系统作为通用搜索过程十分有用,而且在解决不同的组合问题方面十分成功.而交互式系统则更适用于系统化的数学形式化开发及机械化的形式化方法.

与模型检查相比,定理证明可以直接处理无穷状态空间,它依赖于诸如结构化归纳技术在无限域上证明.交互式定理证明器需要与人的交互,因此定理证明慢且易出错.在寻找证明的过程中,用户也能经常获得对系统或要证明的特征的进一步的认识.

计算机系统的发展要求使用数学方法对其可靠性进行定义和验证.形式化方法作为一种实用的数学方法,在计算机系统的可靠性研究中得到了广泛的应用.文献[21]论述了形式化方法在当前计算机软硬件可靠性工程中的应用,介绍了一些典型实例,并讨论了未来的发展方向.文献[22]提出了一种用于描述程序断言的时态逻辑演算系统,并证明了该系统具有足够的表达能力和推理能力.文献[23]介绍了形式化方法的基本理论,考察了不同的形式化方法及其选择和应用问题,最后通过具体实例说明它们是如何在关键系统中应用的.

万方数据

3 下一步工作

军方拟对软件单独计价,拟将软件的可靠性从装备可靠性中独立出来,提出独立的度量指标和度量方法,有许多工作亟待开展.根据软件工程研究现状[25,26],结合我们的研究基础,我们将在两个方面开展研究:

(1) 通过统计测试进行认证

统计使用测试是净室方法的组成部分,其目的并不是传统开发方法中所说的发现尽可能多的错误,而是为了确定软件可靠性.软件可靠性不仅依赖于软件正确性,也依赖于软件的使用方法.如果在一定状态或输入下软件发生失效,那么整个软件的可靠性还需看这种状态或输入发生的概率大小.这就是统计使用测试的理论依据.

净室方法原来的使用模型是简单 Markov 模型.该模型在大型多用户系统中很容易发生状态的组合爆炸.这个问题可以用分层 Markov 模型解决.

确认是对质量完成情况的控制.通过使用测试进行验证可以预测软件在实际运行中的可靠性.由于使用软件模型进行验证需要大量的失效数据,这一点在净室开发法中难以得到满足,为此,需要对统计测试法进行改进并采用层次 Markov 使用模型.

(2) 形式化方法在软件质量保证技术中的应用

过去 10 年来技术的进步使人们有能力克服更难更大的问题.该领域的科研进步依赖于基础研究、新的方法、新的工具以及集成不同的方法,并使研究人员合作以促进技术转化.

方法集成最有希望的一个方向是把模型检查和定理证明结合起来,从而可以从这两个部分获得好处.一种方法是把模型检查作为演绎框架中的一种决策过程,如同 PVS 及 STeP 中所做的那样.另一种结合演绎和模型检查的方法是使用演绎来获得有限状态抽象,而实现本身可用模型检查验证.

这些抽象常用于适合模型检查的问题,而很少用于严格验证的问题.演绎亦可用于验证基于假设的证明约束,这些证明约束包括了已经由模型检查独立检验的部件实现.归纳可以与模型检查相结合来验证由有限状态过程网络组成的系统.

在整个系统开发过程中,形式化方法与非形式化方法相互补充.正如 Praxis 在 CDIS 实例中所做的,并非取代而是协助完善非形式化方法.形式化方法已在规范和验证中显示了其威力,在需求分析、精化和测试中值得进一步开发.

致谢:本文所述工作还得到了国家自然科学基金资助(项目号:69673009);国家 863 计划资助(项目号:863-306-ZT02-04-2);高等学校重点实验室访问学者基金资助.

参考文献:

[1] 贲可荣,马良荔,刘孟仁.软件可靠性分析及对策 [A].第四届全国计算机应用联合学术会议论文集 [C].北京:电子工业出版社,1997.

[2] 贲可荣,邹云松,刘孟仁.战略防御系统软件可靠性分析 [J].海军工程学院学报,1998,(2):5-10.

[3] Prowell S J, Trammell C J, Linger R C, *at el*.净室软件工程——技术与过程 [M]. 贲可荣,张志祥,张秀山,等译.北京:电子工业出版社,2001.

[4] CMU-SEI.能力成熟度模型(CMM) 软件过程改进指南 [M].刘孟仁,贲可荣,马良荔,等译.北京:电子工业出版社,2001.

[5] 马良荔,刘孟仁,贲可荣.软件工程能力成熟度模型研究 [J].计算机应用研究,1998,15(6):8-10.

[6] 曾 浩,贲可荣.通过提升 CMM 级别改进软件质量 [J].舰船电子工程, 2001,(2):35-41.

[7] 张 宇,贲可荣.小型软件开发单位软件过程改进模型探讨 [J].计算机应用研究,2001,18(11):1-4

[8] 贲可荣,俞立军.能力成熟度模型及其评估算法 [J].舰船电子工程,2001,(5):56-61.

[9] 贲可荣,俞立军.国际高成熟度研讨会综述 [J].舰船电子工程(已录用),2002.

[10] 熊 伟,贲可荣.净室技术与软件能力成熟度模型的融合 [J].武汉大学学报,1999,45(5B):691-694.

万方数据

[11] 孙 宁,贲可荣.文档测试与文档过程成熟度模型探讨 [J].计算机应用研究(已录用),2002.
[12] Binder R V.面向对象系统的测试 [M].华庆一译.北京:人民邮电出版社,2001.
[13] 石剑琛,贲可荣,汤志国.测试方法综述 [J].武汉大学学报,1999,45(5B):687-690.
[14] 汤志国,贲可荣.基于逻辑谓词的测试方法探讨 [J].计算机科学,2000,27(11.增刊):121-123.
[15] 曾 浩,贲可荣.图形用户界面软件的自动测试 [A].全国软件技术与软件产业研讨会论文集 [C].大连:大连出版社,2001.
[16] 颜 炯,贲可荣.可靠软件的开发及认证方法 [J].计算机工程与科学,1999,21(A1):98-101.
[17] Ben Ke-rong,Yan Jiong,Wang Ji. Computer System Reliability Allocation Method and Supporting Tool[J]. Wuhan University Journal of Natural Sciences,2001,6(1-2):505-510.
[18] 张志华,贲可荣.基于用户操作剖面的软件可靠性分配研究 [J].计算机工程与科学,1999,21(A1):94-97.
[19] 张志华,贲可荣.一类 NHPP 模型参数的估计方法 [J].计算机科学,2000,27(11.增刊):115-118.
[20] 贲可荣,马良荔,刘孟仁.正确看待形式化方法 [A]. New Technologies on Computer Software [C], International Academic Publisher,1997.
[21] 贲可荣,颜 炯.形式化方法在计算机可靠性工程中的应用 [J].计算机工程与科学,1999,21(A1):89-93.
[22] BEN Kerong, CHEN Huowang, WANG Bingshan. PTL sequent calculus system [J]. SCIENCE IN CHINA(Series A),1995,38(5):598-607.
[23] 颜 炯,贲可荣.逻辑与程序正确性 [J].计算机科学,1999,26(7.增刊):57-60.
[24] 熊 伟,贲可荣.形式化方法及其在关键系统中的应用 [J].计算机科学,2000,27(11.增刊):27-30.
[25] Stiller E, LeBlane C.基于项目的软件工程——面向对象研究方法 [M].贲可荣,张秀山译.北京:机械工业出版社,2002.
[26] 俞立军,贲可荣.软件工程问题综述 [J].计算机科学,2001,28(9.增刊):328-332.

Outlines of software quality assurance technologies research

BEN Ke-rong[1,2]

(1. Dept. of Management Sci., Naval Univ. of Engineering, Wuhan 430033, China;
2. State Key Laboratory of Software Engineering, Wuhan 430072, China.)

Abstract: This paper outlines the necessity of software reliability research, gives some software quality improvement techniques in regard to development process management, quality assurance standards, testing techniques, reliability estimation and predication, formal specification and verification, introduces some achievements that we have obtained recent years. Lastly, the paper points out further research problems on statistics testing and formal methods.

Key words: software quality; capability maturity model (CMM); software standards; software testing; software reliability; formal methods

万方数据