

# 面向不同行业软件的质量评测模型及实践

聂南<sup>1</sup> 邓璐娟<sup>1</sup> 夏启明<sup>2</sup> 李健勇<sup>1</sup>

(郑州轻工业学院计算机与通信工程学院 郑州 450002)<sup>1</sup>

(武汉大学计算机学院湖北软件评测中心 武汉 430072)<sup>2</sup>

**摘要** 针对不同的行业类型软件的质量需求和应用,提出了建立对应的软件质量评测模型,论述了系统质量评价时涉及的相关质量特性选择根据和方法。还结合几个实际的工程测量类软件产品,给出了工程测量类软件测试模型的一个应用实例。测评过程展示了测试实践中遇到的问题和解决方法,如一些接口交互错误和格式化数据类型错误等。最后软件的质量得到了改进,并且统计出一个较全面和客观的评价结果。

**关键词** 质量评测模型,工程测量,接口,交互

**中图法分类号** TP312.56 **文献标识码** A

## Different Trade Software Quality Evaluation Model and Test Practice

NIE Nan<sup>1</sup> DENG Lu-juan<sup>1</sup> XIA Qi-ming<sup>2</sup> LI Jian-yong<sup>1</sup>

(School of Computer and Communication Engineering,Zhengzhou University of Light Industry,Zhengzhou 450002,China)<sup>1</sup>

(Hubei Software Testing Center,School of Computer,Wuhan University,Wuhan 430072,China)<sup>2</sup>

**Abstract** Aiming at different trade software quality requirement and application problems,it set up their corresponding quality evaluation and test model,and it described evaluation methods and selection basis related with system quality features. It also provided an application case of engineering measuring software quality evaluation model combined with several engineering measuring software. Test practice process was illustrated,which includes questions and resolutions,such as interface interaction errors and format data type errors. At last,the comprehensive and justice evaluation result was counted up,and software quality could be improved.

**Keywords** Quality evaluation and test model,Engineering measuring,Interface,Interaction

## 1 引言

GB/T 16260-2006《软件工程 产品质量》(ISO/IEC TR 9126-2:2003 IDT)系列标准及相关制定机构<sup>[1-3]</sup>,能够对软件产品质量进行综合评价。可是面对日新月异的软件产业及软件产品,该标准缺乏针对性和实效性,评测机构无法实现评测方法和技术的及时更新。结合被评价软件的质量需求与特点,选择适用的质量特性、子特性和度量属性,构建针对被评价软件的质量模型,是进行软件产品质量评价的关键步骤。面向质量评价与某类软件测试的研究已经涌现了许多成果。例如文献[4]结合和完善 Logiscope 和 McCabe 两种软件度量体系,探讨了航天型号软件的质量度量问题,建立其质量度量评估模型,但是未能进一步测试并对不同类型的软件进行度量。文献[5]提出一套适用于石油应用软件测试的方法,应用这套方法对地震数据理解释一体化大型软件系统进行综合测试,发现了许多影响软件质量的关键问题,并得到了及时的更改<sup>[6,7]</sup>。基于软件行为需求,使用 viewchart 形式化图,提出一种面向软件早期需求的可靠性评定方法。目前,行业化

正在企业生产中起着越来越重要的影响。面向不同的行业标准需求,已经开发出了许多对应的行业类型软件,如教育类软件、通讯类软件、医疗类软件、测量类软件等等。由于行业需求的不同特点及应用差别,致使对行业软件的评价与测试需要使用针对性的评估体系与测试方法。遗憾的是,目前还未有面向不同行业的软件质量评测模型出现。由于软件测试的方案与技术也不断被提出,需要针对行业软件的需求进行相应的功能、性能以及安全测试。本文建造了针对行业软件的质量评测模型,并且对某类型软件进行了大量实际的测试。结果发现该模型能够对不同类型软件进行较全面和客观的评测,从而达到促进和提高软件质量的目的。

## 2 行业软件评测模型及质量特性的选择

本节对文献[2]中的度量属性进行选择或调整,定义其权值参数的、有效性、一致性、无关性。给出以下一个面向不同行业类型软件的质量评测模型。

**定义 1** 度量项( $X_1 * T_1 + X_2 * T_2 + X_3 * T_3 + \dots + X_i * T_i$ )( $X_i$  为选择的度量, $T_i$  为度量的权重) $=V$ (软件质量评价

到稿日期:2010-03-05 返修日期:2010-06-16 本文受国家重点基础研究发展计划(2006CB708302),河南自然科学基金研究(基于 SOA 软件的单个服务测试研究),郑州轻工业学院博士基金(200903)资助。

聂南(1973-),男,博士,副教授,主要研究方向为软件工程与测试,E-mail:oraclenienan@gmail.com;邓璐娟(1963-),女,博士,教授,主要研究方向为控制工程与软件测试。

值)。

注:度量项包含若干子项( $X_1, X_2, X_3, \dots$ ),它们对应的权重  $T_1 + T_2 + T_3 + \dots + T_i = 1$ 。

**定义 2** 度量子项( $X_{i1} * T_{i1} + X_{i2} * T_{i2} + X_{i3} * T_{i3} + \dots + X_{in} * T_{in}$ )( $X_{ij}$ 为选择的度量子项, $T_{ij}$ 为度量的子项权重) $= X_i$ (度量子项)。

注:一个度量项  $X_i$ ,包含若干子项( $X_{i1}, X_{i2}, X_{i3}, \dots$ ),它们对应的权重  $T_{i1} + T_{i2} + T_{i3} + \dots + T_{in} = 1$ 。

**定义 3** 度量子项关联事件( $X_{i1} * R_{i1} + X_{i2} * R_{i2} + X_{i3} * R_{i3} + \dots + X_{in} * R_{in}$ )( $X_{ij}$ 为选择的某度量子项, $R_{ij}$ 为支持该度量子项关联值) $= X_i$ (关联支持度)(例如选择某精确度量项,与其关联的度量项依次可定义为准确度、安全等)。不同的关联主项决定着不同的关联度。计算关联度系数  $\beta_{k(i)}$ ,  $i=1, \dots, M$ (相关度量个数),  $\beta_{k(i)}$ 反映了度量值  $T_i$ 与真实目标的关联度。

行业软件度量模型的关联概率系数  $\beta_{k(i)}$ 的关联事件定义如下:

$$\begin{cases} \beta_{k(i)} \text{ 源于度量项,} & i=1, \dots, M \\ \text{没有量测源于度量项,} & i=0 \end{cases}$$

**定义 4** 不同的度量项其采用的测试方法也不同,给出以下度量项的测试方法关联矩阵:

度量项 \ 测试方法	T <sub>1</sub>	T <sub>2</sub>	...(第 i 项)	T <sub>N</sub>
X <sub>1</sub>	X <sub>1</sub> & T <sub>1</sub>	R <sub>12</sub>	...	R <sub>1n</sub>
X <sub>2</sub>	R <sub>21</sub>	X <sub>2</sub> & T <sub>2</sub>	...	R <sub>2n</sub>
...(第 i 项)	...	...	X <sub>N-1</sub> & T <sub>N-1</sub>	...
X <sub>N</sub>	R <sub>n1</sub>	R <sub>n2</sub>	...	X <sub>N</sub> & T <sub>N</sub>

度量项采用的首选测试方法为对角线关联度,定义为  $X_i \& T_i - X_N \& T_N$ 。其他的测试方法依照其关联度,再决定使用。

选择或调整不同行业类型软件度量属性时,依次选择以下 4 个度量特性。

1)首先功能需求是各类行业软件的首要需求,这也是行业软件评测的首要因素。行业软件所提供的内容和功能应该符合行业颁布的相关业务规范和客户要求。例如工程测量类软件功能的适合性、精确性、准确性是该行业业务规范最重要的度量属性,同时是满足使用方的必要条件之一。

2)可靠性也是不同行业软件的重要方面。各种软件,特别是军工、金融、医疗和测量等类型的软件对可靠性都有较高要求。评价时需要考虑软件可靠性中的成熟性、容错性、易恢复性等子特性,系统出现故障后,是否从备份系统恢复,是否能够提供紧急情况下的保障功能。易用性是软件比较重要的质量要求之一。通常,娱乐、教育类、通讯类等软件应该满足易用性中易理解、易学习、易操作、吸引力等子特性的要求。面向不同用户的软件,易用性子特性的重要程度略有差异。某些专业性要求较高的产品,可能要求软件界面风格、操作方式等符合行业的相关规定。

3)维护性反映了软件交付后是否可维护的能力,通常由客户方与开发方共同协作解决。由于系统复杂,同时软件更新比较快,常常在易于分析、易于改变、易于测试、稳定性等方面具有一定的要求。

4)可移植性反映了软件可应用于不同环境的能力,通常由客户方提需求。不同行业类软件皆有各自的共存性要求,

也具有硬件环境、软件环境的各自要求。特定情况下,对软件易替换性等也有一定要求。例如,包含硬件和分布式数据库的软件安装比较复杂,需要对易安装性做测试评价。

本节结合工程测量类软件的特点,对在文献[1]中给出的 112 个度量项进行选择或调整。可将度量项分为 A,B,C,D 4 类,建立其关联度值以便取舍。

A类:基本度量项,反映各类软件的基本质量要求的度量项,是必须选择的度量项,无特殊情况皆应采用。例如,软件功能性中的功能充分性、功能完整性以及易用性中功能的易理解性等度量项。其被选中,则关联度定义为 1.0。

B类:可选度量项,反映各类软件的可选质量要求,但该度量所需数据将在满足一定情况下才可能获得。或某些类软件可能存在此方面的要求(如演示的获得性要求,测量类软件就无此要求,而教育类软件则有此要求,效率部分作为测量类可选的度量项。在可获得相关数据或信息时,建议选择此类度量项)。与选中 A 类的关联度定义为  $0.5 < \text{关联度} \leq 0.75$ 。与 C,D 类的关联暂不考虑。

C类:非常规可选度量项,即不常使用的度量。与常规各类软件的特点不直接相关,但可能与某个特定的行业软件的质量要求相关,或可能是评价委托单位特别要求进行评价的度量项。与选中 A 类的关联度定义为  $0.25 < \text{关联度} \leq 0.5$ 。与 B,D 类的关联暂不考虑。

D类: $0 \leq \text{关联度} \leq 0.25$ ,与 B,C 类的关联暂不考虑。其与所评测的软件无关,不予考虑。

其中,建议为 A 类的度量项应尽可能地选用;具备条件时,B类度量项也应选用。根据其行业软件的特点和需求、国家标准和测试大纲,选择度量项(A,B,C),根据所选择的度量项对各度量项目的权重进行调整。测试项权值与即将实施的测试用例集实现关联对照,然后实施测试,并且对测试结果进行统计分析,最终得出软件质量的综合评定值。

给出 xxx 类软件产品质量评测模型评测序列图如图 1 所示。

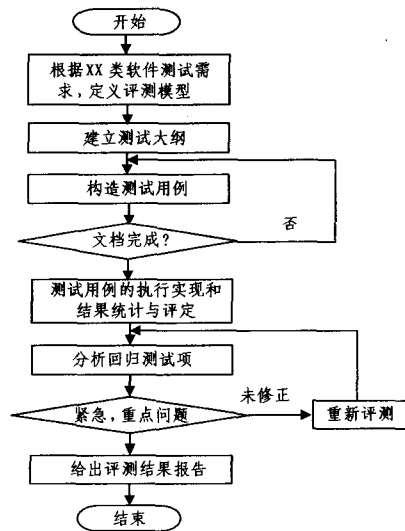


图 1 行业软件评测序列图

### 3 工程测量类产品质量评测实例

#### 3.1 工程测量软件及测试用例

工程测量类软件是指利用工程测量技术和网络数据库技

术向用户提供地形地理、设备或者行业网络系统的测量信息的软件系统,统称为工程测量类软件(Engineering Measurement Software)。工程测量类软件应用广泛,内容非常丰富,几乎涉及工程测量有关的所有服务。根据使用人群的不同,可以分为行业系统测量、设备测量和其他专业工程测量软件等。设备测量软件主要应用于各种设备的测试和度量中,如GPS测试设备主要用于运输监控管理系统;地理空间测量软件主要面向地质、地貌和空间环境的测量;行业系统测量主要面向行业系统,如电力、电信和能源等网络系统的测量。

随着我国大量行业对测量软件的需求,使人们越来越重视测量的效率和作用,因此对测量软件的质量进行测评也变得越来越严格。工程测量类软件提供形式多样的有关测量专用服务,如地理定位、查询、误差分析、设备度量等。这些服务都需要进行复杂的误差分析运算和数据转换。对工程测量类软件产品的测试,应重点关注其精确性和准确特性。此外,还应结合软件不同的类型,关注不同的方面。如设备测量软件,主要的使用对象是专业测量人员,需要专业的测量知识;面向广大普通用户的测量软件,还应着重关注其准确性、互操作性、成熟性、吸引力等。在一些非常重要场所使用的测量软件,还应关注其安全特性,比如操作系统安全、接口安全、数据安全等。

本节以若干工程测量软件为例,论述工程测量类软件产品质量评测模型的建立和应用。经过功能性、可靠性、易用性、效率、可移植性等方面的测试,并且经过测试记录分析、统计,获得相关统计数据如表1所列。

表1 若干工程测量软件测试统计表

工程测量软件	地理信息类	行业类	设备类
测试功能项	16	138	35
用例数目	235	2029	126
严重问题	9	77	5
一般问题	14	100	2
建议问题	5	23	2

严重问题指1)不能完全满足系统要求,基本功能未完全实现;2)影响系统要求或基本功能的实现;3)使操作者不方便或遇到麻烦,它不影响执行工作功能或重要功能。一般问题指其他不影响使用功能的表面问题。建议问题指为提高系统的质量提出的建议。开发方对发现的问题进行了修改,通过中心的回归测试后,项目产品功能基本完备。

### 3.2 “xxx软件 V1.0”评价过程及结果

我们对被评价软件《xxx软件 V1.0》建立对应的评测模型,并且进行了全面而严格的测试。通过对该系统的功能测试,系统运行正常,功能齐全,实用性强。《xxx软件 V1.0》是由C++编写的,具有地面常规控制测量和GPS控制测量数据处理、构件测量数据处理及图形显绘、变形监测数据分析等功能,可应用于核电工程的多种工程测量的数据处理和分析中。

#### 3.2.1 “xxx软件 V1.0”质量评测模型

根据被评价软件的质量需求模型,112个度量项选择A类12项,B类11项,C类17项,可确定的“xxx软件 V1.0”质量评价,如表2所列。软件质量评测模型宜在测试前完成,以利于确定测试内容。在评测模型中给出了特性、子特性、度量项的权重。

表2 “xxx软件 V1.0”质量评价价值与用例对照表

特性	权重比例	子特性	权重比例	度量项目类别与用例数目	权重比例
功能性	0.3	适合性	0.4	功能的充分性 A/130	0.4
				功能实现的完整性 A/130	0.3
		准确性	0.2	功能实现的覆盖率 A/130	0.3
				精度 A/42	0.7
				计算的准确性 A/42	0.2

权重、度量值依据度量类别与关联度和实际测试用例来确定。在进行评价之前,先从测试记录中获得原始数据,然后统计分析,计算出各质量特性的度量值,最后计算出软件的评价结果。

#### 3.2.2 使用的测试方法与技术

根据评测模型建立测试方案,该软件进行了功能性、可靠性、易用性、可移植性等方面的测试。我们采用了边界值分析(测试数值精度)、错误假定(测试计算准确性、数据移植)、逻辑覆盖(功能模块全面测试)、接口错误注入等测试方法。发现的问题列在测试用例统计数据表4中。软件安全性测试采用表3中的测试方法。

表3 安全保密性测试

类别		测试重点
验证	无效的用户验证	用户的授权和验证机制的方法(例如加密物防止交换数据,硬盘 license 验证等)。
参数保护	关键数据编辑操作	用户很容易就可以操作关键数据,如增、删、改。如果该值代表的是诸如设备精度的敏感数据,那么该应用程序就很容易受到攻击。
	IN/OUT 操作	数据导入,导出是否有信息丢失、损坏,包括格式、类型和精度等。
命令执行	缓冲区溢出	是否可以通过输入特殊信息。
	操作系统命令	是否可以通过输入特殊命令实现系统接口的访问控制。
接口	接口	包括组件(如 COM, DLL)接口错误,是否可以通过接口缓冲区溢出扰动函数错误注入导致系统崩溃[8,9],格式化(超长)字符串攻击使系统接口出现访问错误,例如文件名等;程序状态及程序环境的配置参数。
	信息泄露	是否可以通过文件内容获取 xx 设备参数配置,精度等技术信息。

#### 3.2.3 测试结果与数据统计

经过测试记录分析、统计,获得如下统计数据。该软件共有功能35个、设计功能测试用例126个、安全保密性测试用例58个、功能依从性测试用例39个、可靠性测试用例74个、易用性测试用例69个、可移植性测试用例5个。功能性测试中,发现问题(缺陷)9个,共发现2个安全问题,软件支持所要求的国际和国家行业测量标准。测试用例统计数据如表4所列。

表4 测试用例统计数据

测试用例表	通过用例	发现错误用例	关联度低于规定值
功能精确度	42	精度溢出1例	84
功能准确度	42	无	84
安全	57	接口注入文件名错1例	69

在获得各个度量项目的统计值后,即可结合质量评测模型,依次计算质量子特性、质量特性的评估值。最后根据评价准则,可计算出整个软件的质量水平,并确定其质量等级。质量评价结果如表5所列。

表5 质量评价结果

特性名称	功能性	可靠性	易用性	可移植性
统计结果	0.86875	0.45 * 0.885	0.25 * 0.715	0.15 * 0.975
		0.15 * 0.975	0.15 * 0.970	

然后,软件精度和安全等问题经过修正(浮点型变为双精度;对 COM 接口设置缓冲区防御机制,如限制字符长、判断是否含有二进制代码和采用更安全的函数等)重新度量,统计后的质量评价结果为 0.98X。

**结束语** 如何对软件的质量进行科学的评测是一个系统工程。应根据软件的特点、应用环境、适用人群、技术架构等提炼出软件隐含的质量需求。本文给出了一个针对不同行业软件质量评测模型,研发人员和用户在定义待测软件的质量评价标准时,可参考本模型确定其行业软件的度量项以及各组成部分的权重。例如,工程测量类软件由于更多地应用于地理、交通、能源等行业工程建设,对软件的数据精确度、计算准确度和数据的便于传输处理的通用性具有较高的需求,所以在制定工程测量类软件的质量评测模型时,应当提升功能的权重,特别是精确性和准确性等方面的需求。此外,A,B,C,D 度量类之间的关联规则还需要深入研究,因地制宜,从使用人员和应用平台等多方面考虑,这样才能使软件得到更充分的改进和有效的应用。

### 参考文献

- [1] GB/T16260.1-ISO/IEC 9126-1. 第 1-4 部分软件工程-产品质量[S]. 2006
- [2] Al-Kilidar H, Cox K, Kitchenham B. The use and usefulness of

(上接第 155 页)

测试的关键是自动化。PTT 可以根据给定的 PFD2PTT 算法自动地从 PFD 得到。测试翻译器从 PTT 中获取路径表达式并把其翻译成测试规格说明的测试脚本框架。测试引擎读取并执行测试规格说明进而生成测试报告。总体上,除了手工构造 PFD 和对测试脚本框架的小小编辑修改之外,我们的 Web 测试方法是自动化的。

为了构造 PFD,我们结合了传统的需求分析方法和 Web 软件的特性。为了生成测试用例,一种方法是将随机输入应用到 PFD 中,直到每个链接至少经过一次。然而,路径可能包含循环,这使得它变得很长。此外,为了获得经过每个链接一次且仅一次的测试路径,PFD 必须是一个欧拉图。若 PFD 不是一个欧拉图,则构造最优长度的测试路径是一个 NP 完全问题。

在测试时,可能存在这样的情况,即只有提供某个特定的输入,特定的链接才会出现在某个页面中。下一步的工作是探讨这种新情况下 Web 测试的问题,以改进提出的测试路径生成方法,并开发一个原型验证本文提出的 Web 测试模型。

### 参考文献

- [1] Li Z, Alaeddine N, Tian J. Multi-Faceted Quality and Defect Measurement for Web Software and Source Contents[J]. Journal of Systems and Software, 2009, 83(1): 18-28
- [2] Hower R. Web Site Test Tools and Site Management Tools. Software QA and Testing Resource Center[OL]. <http://www.softwareqatest.com/qatweb1.html>. Accessed on Feb. 2010
- [3] Lucca G A D, Fasolino A R. Testing Web-based Applications: The State of the Art and Future Trends[J]. Information and Software Technology, 2006(48): 1172-1186

the ISO/IEC 9126 quality standard [C] // 2005 International Symposium on Empirical Software Engineering (ISESE 2005). Noosa Heads, Australia. IEEE, 2005; 126-132

- [3] Garcia F, Bertoa M, Calero C, et al. Towards a consistent terminology for software measurement[J]. Information and Software Technology, 2006, 48(8): 631-644
- [4] 孙梦璘,甘志强. 航天型号软件代码质量度量评估实现[J]. 系统工程与电子技术, 2009, 31(4): 956-133
- [5] 柳永坡,张志良,吴际忠,等. 石油应用软件的测试方法研究与实践[J]. 石油地球物理勘探, 2008, 43(6): 731-735
- [6] Alipour H, Isazadeh A. Software Reliability Assessment Based on a Formal Requirements Specification [C] // International Conference on Human System Interactions. Publisher: IEEE Computer Society, 2008; 311-316
- [7] Alipour H, Isazadeh A. Software Reliability Prediction Based on a Formal Requirements Specification[C] // Advances in Computer Science and Engineering 13th International CSI Computer Conference, Kish Island, Iran, 2008; 816-820
- [8] 聂南,谢晓东,甘勇,等. 基于 XML API 的组件扩展接口变异测试方法[J]. 计算机科学, 2008, 35(6): 283-287
- [9] 陈锦富,卢炎生,谢晓东. 软件错误注入测试技术研究[J]. 软件学报, 2009, 20(6): 1425-1444

- [4] Dobolyi K. An Exploration of User-Visible Errors to Improve Fault Detection in Web-based Applications[D]. School of Engineering and Applied Science, University of Virginia, 2009
- [5] Halfond W G J, Anand S, Orso A. Precise Interface Identification to Improve Testing and Analysis of Web Applications[C] // The 18th International Symposium on Software Testing and Analysis. ACM, NY, USA, 2009
- [6] Subraya B M, Subrahmanya S V. Object Driven Performance Testing of Web Applications[C] // The First Asia-Pacific Conference on Quality Software. HongKong, China, Oct. 2000
- [7] Andrews A, Offutt J, Alexander R. Testing Web Applications by Modeling with FSMs [J]. Software and Systems Modeling, 2005, 4(3): 326-345
- [8] 钱忠胜,缪淮扣. 面向用户会话的 Web 应用测试用例生成及其优化[J]. 计算机科学与探索, 2008, 2(6): 627-640
- [9] Benedikt M, Freire J, Godefroid P. VeriWeb: Automatically Testing Dynamic Web Sites[C] // The 11th International World Wide Web Conference. Honolulu, HI, May 2002
- [10] Reza H, Ogaard K, Malge A. A Model based Testing Technique to Test Web Applications Using Statecharts[C] // ITNG' 08. Washington, DC, USA, 2008; 183-188
- [11] 邓小鹏,邢春晓,蔡莲红. Web 应用测试技术进展[J]. 计算机研究与发展, 2007, 44(8): 1273-1283
- [12] Beizer B. Software Testing Techniques(2nd edition)[M]. International Thomson Computer Press, 1990
- [13] Offutt J, Wu Y, Du X, et al. Bypass Testing of Web Applications [C] // The 15th IEEE International Symposium on Software Reliability Engineering. Saint-Malo, Bretagne, France, Nov. 2004
- [14] Jia X, Liu H, Qin L. Formal Structured Specification for Web Applications Testing[C] // 2003 Midwest Software Engineering Conference. Chicago, USA, June 2003