

# 软件测试面临的挑战与发展趋势

朱少民

(同济大学 软件学院, 上海 201804)

**摘要:** 结合当今软件研发模式的改变以及软件新技术的发展, 特别是人工智能、大数据、云计算、区块链、物联网等新技术的普遍应用, 阐述了软件测试所受到的各种挑战, 包括在系统功能测试、性能测试、安全性测试、可靠性测试等各个方面所面临的挑战。最后, 展望未来, 描绘出软件测试发展的五大趋势——敏捷化、高度自动化、云化、服务化和智能化。

**关键词:** 软件测试; 敏捷开发; 人工智能; 信息物理系统

**中图分类号:** TP606    **文献标识码:** A    **文章编号:** 1000-8829 (2020) 01-0001-04

**doi:** 10.19708/j.ckjs.2020.01.001

## 1 背景

当今, 已进入一个快节奏、碎片化的移动互联网时代, 面对新的软件应用环境, 其中最为显著的变化体现在人工智能、大数据、云计算、区块链、物联网等新技术的兴起, 这些新技术甚至改变了传统的软件形态, 以至于今天一切皆服务 (Everything as a Service), 例如:

- 基础设施即服务 (Infrastructure as a Service, IaaS);
- 平台即服务 (Platform as a Service, PaaS);
- 软件即服务 (Software as a Service, SaaS)。

这些新技术、新环境自然给软件测试带来新的挑战。

软件形态的变化自然催生了软件工程环境的变化, 也深深影响着软件研发模式、流程和实践, 其中最突出的是敏捷开发模式的兴起。软件开发的节奏越来越快, 力争做到持续交付, 但只有做到持续构建、持续集成、持续测试才能做到持续交付。其中最为困难的就是如何做到持续测试——高效、及时地完成测试工作, 这也给软件测试带来新的挑战。

## 2 挑战

### 2.1 AI 系统的测试

对 AI 软件的测试, 需要增加一个时间维度, 因为 AI 系统的能力是动态的, 系统具有自身学习能力, 随着时间增长, 其系统能力必须显著增强, 例

如谷歌公司的 AlphaGo。针对这样的黑盒测试过程, 其测试代价不小, 测试周期长, 不容易实现自动化测试。

像 AlphaGo 这样的系统还相对简单, 而智能语音应答系统的测试则更为复杂, 需要模拟各种上下文场景、用户的提问, 如图 1 所示, 很难判断其应答能力是否有明显提升。虽然可以通过大量的应答测试数据的统计结果确定语音识别、语义理解的正确率, 但其测试样本还是有较大的局限性。

深度神经网络学习算法, 通常被认为是黑箱模型, 即其具有不可解释性, 且鲁棒性差, 结果不稳定。因此, 一般需要借助大量的数据或实验来进行普适性验证, 这与传统的测试相差甚远。测试 (Test) 从本质上看就是“样本性的试验验证活动”, 传统测试可以借助“等价类划分”等类似方法来大幅降低样本量, 但对于 AI 软件, 这类方法已经失效。AI 软件的测试还依赖大数据, 基于大数据的自动产生、分析、呈现等技术, 能有效地验证 AI 软件的合理性, 但测试的工作量较大。例如, 为了测试 Numenta (工作模式更接近人的大脑的 AI 软件), IBM 资深研究员 Winfried Wilcke 带领 100 人的团队对其算法进行测试。面对俄罗斯的控制机器人集群的智能软件包 Unicum, 又如何测试呢? 它被称为“机器人之上的机器人”, 可以独立分配集群内部的机器人角色、确定集群中的核心、替换脱离的单位, 还会自动占据有利位置、搜寻目标, 并在自动模式下向操纵员申请作战与摧毁目标的许可。

引用格式: 朱少民. 软件测试面临的挑战与发展趋势[J]. 测控技术, 2020, 39 (1): 1-4.

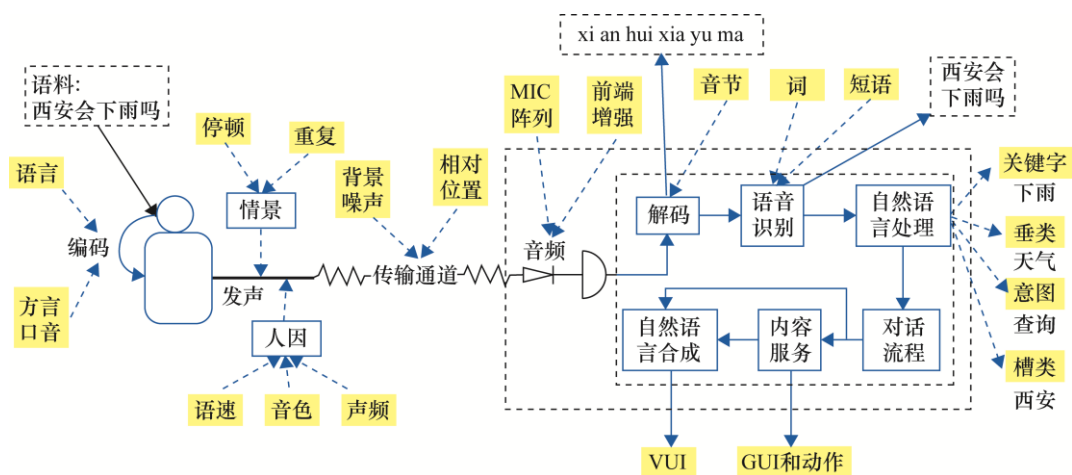


图1 智能语音应答系统的测试流程示意图

## 2.2 大数据的测试

大数据测试，并不比 AI 软件测试简单，更准确地说，大数据离不开 AI、AI 也离不开大数据。大数据的体量大、多样化和高速处理所涉及的数据生成、存储、检索和分析使得其测试也极具挑战。针对大数据的测试，主要面临的挑战有：

① 难以确定其测试预言（Test Oracle）——测试结果的判断准则，因为大数据的处理结果是否正确，有时很难确定。

② 大数据的多样性给测试数据准备、测试数据的组合带来极大的工作量，如何高效或有效地进行测试，也极具挑战。

③ 数据经过采集、存储、检索和分析后，在体量大、多样性和变化快、实时性的情况下，如何验证数据的一致性和完整性，需要开发相应的工具。

④ 在大数据的性能测试、安全性测试等方面，也会遇到业务场景过多的困难。

## 2.3 云服务测试

今天的云服务包含私有云、公有云、混合云等不同的形态，而其测试所面临复杂的系统架构、复杂的配置、大量的计算或存储节点等方面所带来的挑战，例如：配置参数的组合数能够达到几千万、几个亿甚至几十亿，如何完成这巨量的组合测试？虽然可以采用两两组合、三三组合等方法缩减组合数，但覆盖率相对较低，存在较大的质量风险。

其次，云服务的性能测试也很有挑战，主要体现在：

- 海量云端压力的模拟将面临高昂代价、甚至无法模拟；
- 需要网络运营商、CDN 服务提供商的配合与支持；
- 传统测试工具难以适应云计算平台的性能测试。

一般采用替代项——系统的可伸缩性（弹性）验证，以避免超大规模的性能测试。而且，云服务平台的容错性或可靠性测试也会面临挑战，例如：

- 错误类型过多；
- 错误信息表达不够充分；
- 业务错误码与 HTTP 错误码含义不匹配；
- 相同错误不同云产品表达不一致；
- 错误码超越可枚举集合。

因此诞生了一门新的工程方法（领域）——混沌工程。

## 2.4 区块链测试

区块链一般分为公有链、私有链、联盟链等不同类型，它们在管理、用户身份认证、节点数、共识机制、智能合约等许多方面均有不同，其测试所面临的挑战有所不同，但也面对一些共同的挑战。例如，安全性测试就很具挑战性，因为安全性始终是“矛与盾”的问题，没有绝对安全的技术、系统或平台，因此很难把握其安全性测试的度，更何况区块链技术平台还涉及基础设施（网络、存储设备等）安全性、加密算法安全性、网络协议安全性、共识机制安全性（包括共识算法漏洞）、智能合约安全性（如代码实现中的安全漏洞）、应用安全性（包括 Web/移动客户端软件、数字钱包等）、身份认证与鉴别等。

作为一个区块链的系统，则是一个完全去中心化的分布式网络——P2P 网络，而且这个网络很可能跨越多个子网、多个数据中心、多个运营商甚至多个国家，其边界是模糊的。但区块链测试，不仅涉及前端 API 与某个区块链节点之间的测试，还涉及大量区块链节点与节点之间的测试。因此，边界模糊、网络及其交易的复杂性都给系统的功能测试、兼容性测试、数据一致性验证、安全性测试和性能测试等带来挑战，特别是在共识算法的合法性、完整性、可终止性的验

证方面。另外一个挑战来自智能合约的验证,因为智能合约执行过程中,一切均需听命于事先设定好的代码,而且智能合约类似法律,事先设定好的代码一旦上线就不能轻易修改。如何确保智能合约百分之百正确?测试的一项重要原则就是测试是不能穷尽的,测试不能保证没有缺陷或百分之百正确。

## 2.5 物联网测试

在工业 4.0 和 5G 时代,万物互联变得司空见惯,Gartner 预计今年将有超过 200 亿台物联网设备投入使用。当这么多设备接入物联网中,需要实现异构互联互通,也意味着物联网会支持更为广泛的工业协议,如 MQTT、OPC UA、ModBus-TCP 等行业标准,并与现有的互联网系统连接起来,形成规模更大的系统,这必然给软件测试带来新的挑战,主要体现在如下方面。

① 测试环境复杂性,如异构互联、M2M 互联、软硬件结合紧密等,需要构建全新的仿真实验环境等。

② 高可靠性测试,因为系统涉及生命安全,需要确保系统的高可靠性,包括可靠性模拟、可靠性估算和可靠性监控等。

③ 实时性,如时间敏感网络(TSN)、实时数据采集、实时数据同步传输等,对测试环境、测试工具都有更高的要求。

④ 兼容性测试,如 EMI/EMC 共存测试,以及各种通信协议、数据采集方式等验证。

⑤ 网络安全策略的测试、配置组合测试,这方面测试工作量也很大,如何提高测试的有效性和高效性,也有很大挑战。

⑥ 另外还应对系统的可伸缩性、可扩展性、协同感知、大数据处理性能、智能特性等进行验证。

## 2.6 敏捷测试

快速迭代、快速交付,迫使人们能高效地完成软件测试,因此人们寄希望于自动化测试,毕竟工具的执行效率高。但从自动化测试的调查结果来看,大多数软件公司的自动化测试实际做得还不够好,只有 4% 的公司能做到 90% 以上的测试实现自动化,所以测试常常是敏捷开发的最大瓶颈,如图 2 所示。

自动化测试做得不够好有多方面的原因,不同的团队有不同的原因,但根据调查,主要的原因有:需求变更频繁、产品的可测试性差、单元测试不足、测试人员能力不足、测试工具不够强大等。

# 3 未来发展趋势

未来,软件测试发展有以下五大趋势。

## 3.1 敏捷化

无论是传统开发的测试,还是敏捷开发的测试,

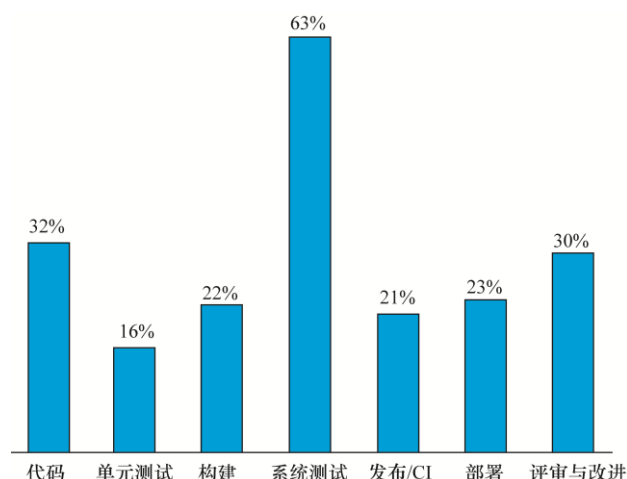


图 2 影响产品交付的主要软件开发环节

在这快节奏的、竞争激烈的时代,测试敏捷化都会成为一种趋势。为此,2019 年初中国电子工业标准化技术协会信息技术服务分会(ITSS)发布了《测试敏捷化白皮书》。测试敏捷化有助于企业持续改进各种测试实践,快速反馈软件质量,提升测试效率,实现自我驱动、灵活赋能、加速价值交付、高效稳定的目标,具体表现为:

➤ 测试左移,加强需求评审、设计评审,推行验收测试驱动开发(ATDD),以及测试驱动设计,让开发做更多的测试,至少做好单元测试和代码评审;

➤ 测试右移,开展在线测试(含性能、安全、易用性、可靠性)、日志/数据分析,反向改进产品;

➤ 激发软件交付和运维团队全员的测试主观能动性,与需求、开发和运维等工作相互促进,使测试成为驱动交付质量与效率持续提升的最主要力量之一。

## 3.2 高度自动化

不久前,Gartner 立足于“以人为本的智慧空间”,发布了 2020 年十大战略性技术趋势,排在第一的就是“超级自动化(Hyper-automation)”,即运用包括机器学习在内的先进技术,不仅包括一系列的自动化工具,还涉及自动化过程的成熟度。在软件研发中,追求高度的自动化测试也是顺理成章的事情。

通常所说的“自动化测试”,其实只能算“半自动化测试”——只有测试执行是自动化的,甚至执行都不能真正做到自动化,因为执行是不稳定的,需要人工干涉,执行的结果也需要人工分析,而且自动化测试的脚本需要人工开发,不能自动生成。

真正的软件测试自动化,不局限于使用工具执行测试,更应涵盖测试目标定位、测试分析、测试设计、质量度量、测试过程监控与评估和测试结果评估等全流程的自动化,让自动化无处不在,覆盖测试的各个



方面,这其中需要全面引入基于模型的测试(MBT)。基于测试模型生成测试数据、测试脚本,从而真正实现测试自动化。而且,应借助自动化测试流程的改进、自动化基础设施的夯实,持续提高自动化测试的成熟度。

### 3.3 云化

测试的基础设施采用当今的虚拟机、容器技术,这使测试环境更容易维护、系统更容易部署,从而更好地支持自动化测试,有利于整合测试资源、提高资源使用的效率,降低企业成本。而且,可以更好地融

合开发、测试和运维,收集更多的研发数据,更好地支持测试服务化、智能化。

### 3.4 服务化

让软件测试成为一种服务(Test as a Service, TaaS),简单地说,就是让所有的测试能力可以通过API来实现,构建测试中台。例如腾讯公司WeTest事业部已经建立较完整的测试中台服务,如图3所示,任何研发人员均可以按需自动获取测试的能力,这样也使研发人员愿意做更多的测试,实现测试左移。

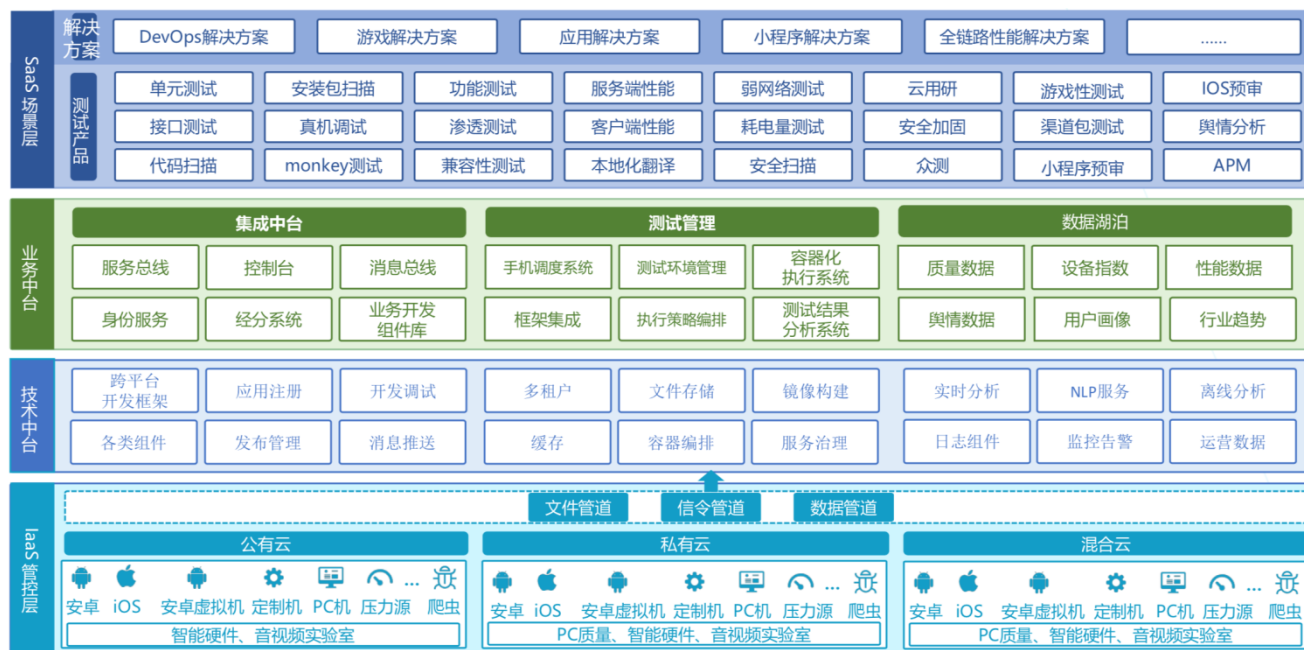


图3 腾讯WeTest事业部测试中台服务

### 3.5 智能化

今天,互联网、存储能力、技术能力和大数据将AI推向第三次浪潮,AI能够服务其他行业,自然能够服务于测试,而且在上述自动化、云化、服务化的基础上,容易收集更多的研发数据,并构建统一的代码库,把所有的代码放在一起更有利于机器学习,使AI能更好地发挥作用,包括测试数据的自动生成、自主操控软件、缺陷和日志的智能分析、优化测试分析与设计等。例如,Eggplant AI导入已有测试资产创建模型、使用智能算法选择最佳测试集运行测试,基于模型算法能最大程度减少构建与维护的成本,其测试覆盖率与之前相比也提高了80%~90%。

另外,人机交互智能更有价值,不能完全依靠机器,未来的测试机器人是需要人去训练的,即相当于把测试工程师的经验和知识、对业务流程和业务场景的理解赋能给机器人。现在人类可以给工具做按钮、菜单、文字、图标的识别训练,但这些工

具还不能真正认知业务流程。未来基于知识图谱和MBT的发展,在特定的业务领域可以帮助计算机提高认知,实现从感知智能向认知智能演化,让测试机器人对业务有良好的认知能力,这才算是实现真正意义上的智能测试。

□



#### 作者简介

朱少民,国内知名测试专家、同济大学特聘教授、中国科技大学教指委委员。近30年来一直从事软件工程、质量管理等工作,先后获得多项科技进步奖,出版了20多部著作,代表作主要有《软件测试方法和技术(第3版)》《全程软件测试(第3版)》等,近5年帮助国内近百家企业提升其软件研发能力水平,并经常在国内外学术会议或技术大会上发表演讲,曾担任IEEE QRS等多个国际学术会议程序委员、IEEE ICST 2019工业论坛主席、《软件学报》审稿人等。