



(12) 发明专利申请

(10) 申请公布号 CN 105045719 A

(43) 申请公布日 2015. 11. 11

(21) 申请号 201510524123. X

(22) 申请日 2015. 08. 24

(71) 申请人 中国科学院软件研究所

地址 100190 北京市海淀区中关村南四街 4 号

(72) 发明人 唐新烨 王青

(74) 专利代理机构 北京汇泽知识产权代理有限公司 11228

代理人 张瑾

(51) Int. Cl.

G06F 11/36(2006. 01)

权利要求书2页 说明书4页 附图3页

(54) 发明名称

基于修复缺陷的变更预测回归测试失效的方法及装置

(57) 摘要

本发明提供一种基于修复缺陷的变更预测回归测试失效的方法及装置,涉及源代码的变更分析技术领域,用于解决现有技术无法找到哪些相应的测试用例会引发回归测试的失效的问题。所述方法包括:通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型;使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型;使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测;通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。本发明适用于在修复软件源代码中存在的缺陷时提供导致回归测试失效的具体测试用例。



1. 一种基于修复缺陷变更的预测回归测试失效的方法,其特征在於,包括:

通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型;

使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型;

使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测;

通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。

2. 根据权利要求 1 所述的方法,其特征在於,所述通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型包括:

收集软件源代码变更的历史,对软件源代码变更的历史进行分析,得到三种特征类别共 18 个特征向量;其中,特征类别为“大小”的特征向量共 5 个,分别为:LA(lines of code added),LD(lines of code deleted),LF(number of files changed),NC(number of changed class),NM(number of changed method);

特征类别为“原子变更”的特征向量共 7 个,分别为:AC(has added classes),DC(has deleted classes),AM(has added methods),DM(has deleted methods),CM(has changed methods body),MR(has renamed methods),PC(has changed parameters of methods);

特征类别为“语义变更”的特征向量共 6 个,分别为:CC(number of changed dependencies),DD(number of deleted dependencies),AD(number of added dependencies),A/RF(has added/removed for blocks),A/RW(has added/removed while blocks),A/RI(has added/removed if blocks)。

3. 根据权利要求 1 所述的方法,其特征在於,所述 Logistic 回归模型如下:

$$P(y=1|x)=\pi(x)=\frac{1}{1+e^{-g(x)}}$$

其中, $g(x)=w_0+w_1x_1+\dots+w_nx_n$, w 为权重向量, x 为特征向量。

4. 根据权利要求 1 所述的方法,其特征在於,所述使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测包括:

将修复缺陷的变更按照时间顺序进行排序;

将修复缺陷的变更分为四份,每一份包含等量的引起回归测试失效的变更;

使用前三份数据作为训练集来训练模型,使用最后一份数据作为测试集。

5. 根据权利要求 1 所述的方法,其特征在於,所述对可能导致回归测试失效的测试用例进行预测和推荐包括:

推荐以下两种可能导致回归测试失效的测试用例:直接测试那些修复缺陷的变更所在的类的测试用例,以及测试调用修复缺陷所在的类的那些类的测试用例。

6. 一种基于修复缺陷变更的预测回归测试失效的装置,其特征在於,包括:

特征选取单元,用于通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型;

模型训练单元,用于使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型;

预测单元,用于使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预

测；

推荐单元,用于通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。

7. 根据权利要求 6 所述的装置,其特征在于,所述预测单元,用于将修复缺陷的变更按照时间顺序进行排序,将修复缺陷的变更分为四份,每一份包含等量的引起回归测试失效的变更,使用前三份数据作为训练集来训练模型,使用最后一份数据作为测试集。

8. 根据权利要求 6 所述的装置,其特征在于,所述推荐单元,用于推荐以下两种可能导致回归测试失效的测试用例:直接测试那些修复缺陷的变更所在的类的测试用例,以及测试调用修复缺陷所在的类的那些类的测试用例。

基于修复缺陷的变更预测回归测试失效的方法及装置

技术领域

[0001] 本发明涉及源代码的变更分析技术领域,尤其涉及一种基于修复缺陷的变更预测回归测试失效的方法及装置。

背景技术

[0002] 为了修复发现的缺陷,软件的源代码经常变更。这些修复缺陷的变更可能引起非预期的行为,使得与现在的回归测试用例的场景不一致,从而导致回归测试失效。研究表明,48.7%的修复缺陷的变更可能会使回归测试失效,也就意味着对于48.7%的变更,开发人员需要多次的运行回归测试。这是个非常耗时和枯燥的过程。因此,在运行回归测试之前,找到哪些相应的测试用例会引发回归测试的失效对于快速修复缺陷是非常有帮助的。

发明内容

[0003] 本发明提供一种基于修复缺陷的变更预测回归测试失效的方法及装置,能够提供导致回归测试失效的具体测试用例,便于快速修复软件源代码中存在的缺陷。

[0004] 本发明提供的基于修复缺陷的变更预测回归测试失效的方法,包括:

[0005] 通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型;

[0006] 使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型;

[0007] 使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测;

[0008] 通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。

[0009] 本发明提供的基于修复缺陷的变更预测回归测试失效的装置,包括:

[0010] 特征选取单元,用于通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型;

[0011] 模型训练单元,用于使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型;

[0012] 预测单元,用于使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测;

[0013] 推荐单元,用于通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。

[0014] 本发明提供的基于修复缺陷的变更预测回归测试失效的方法及装置,通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型,使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型,使用该预测模型对修复缺陷的变更是否会导致回归测试失效进行预测,通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。与现有技术相比,本发明能够提供导致回归测试失效的具体测试用例,便于快速修复软件源代码中存在的缺陷。

附图说明

[0015] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其它的附图。

[0016] 图1为本发明实施例提供的基于修复缺陷的变更预测回归测试失效的方法的流程图;

[0017] 图2为本发明实施例中构建特征模型的18个特征向量;

[0018] 图3为本发明实施例中推荐可能导致回归测试失效的测试用例时使用的静态调用图的关系图;

[0019] 图4为本发明实施例提供的基于修复缺陷的变更预测回归测试失效的装置的结构示意图。

具体实施方式

[0020] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0021] 本发明实施例提供一种基于修复缺陷的变更预测回归测试失效的方法,如图1所示,所述基于修复缺陷的变更预测回归测试失效的方法包括:

[0022] S11、通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型。

[0023] S12、使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型。

[0024] 其中,所述 Logistic 回归模型如下:

[0025]
$$P(y = 1 | x) = \pi(x) = \frac{1}{1 + e^{-g(x)}}$$

[0026] 其中, $g(x) = w_0 + w_1x_1 + \dots + w_nx_n$, w 为权重向量, x 为特征向量。

[0027] S13、使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测。

[0028] S14、通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。

[0029] 进一步地,所述通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型包括:

[0030] 收集软件源代码变更的历史,对软件源代码变更的历史进行分析,得到如图2所示的三种特征类别共18个特征向量;其中,特征类别为“大小”的特征向量共5个,分别为:LA(lines of code added),LD(lines of code deleted),LF(number of files changed),NC(number of changed class),NM(number of changed method);

[0031] 特征类别为“原子变更”的特征向量共7个,分别为:AC(has added classes),DC(has deleted classes),AM(has added methods),DM(has deleted methods),CM(has

changed methods body), MR(has renamed methods), PC(has changed parameters of methods) ;

[0032] 特征类别为“语义变更”的特征向量共 6 个, 分别为 :CC(number of changed dependencies), DD(number of deleted dependencies), AD(number of added dependencies), A/RF(has added/removed for blocks), A/RW(has added/removed while blocks), A/RI(has added/removed if blocks)。

[0033] 进一步地, 所述使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测包括 :

[0034] 将修复缺陷的变更按照时间顺序进行排序 ;

[0035] 将修复缺陷的变更分为四份, 每一份包含等量的引起回归测试失效的变更 ;

[0036] 使用前三份数据作为训练集来训练模型, 使用最后一份数据作为测试集。

[0037] 如图 3a 所示, 有 A、B、C 三个类。B 是继承的 A, C 调用了 A 类中的 f1 的函数。在新版本中修复缺陷的变更是增加了 A 中的 IF block。图 3b 中是与 A、B、C 类相对应的 3 个 Junit 回归测试类 TestA、TestB、TestC。其中 TestA 中的 test1 测试的是 A 类中的 f1 函数, TestB 中的 test2 测试的是 B 类中的 f2 函数, TestC 中的 test3 测试的是 C 类中的 f3 函数。

[0038] 进一步地, 所述对可能导致回归测试失效的测试用例进行预测和推荐包括 :

[0039] 推荐以下两种可能导致回归测试失效的测试用例 :

[0040] 第一种是直接测试那些修复缺陷的变更所在的类的测试用例。

[0041] TestA 类中的 test1 方法就是推荐的第一种测试用例。

[0042] 第二种是测试调用修复缺陷所在的类的那些类的测试用例。

[0043] 如图 3c 所示, A 类中 f1 函数被 B 类的 f2 和 C 类的 f3 函数调用, 所以当 A 类中的 f1 函数改变了, B 类和 C 类的语义也会改变。所以 B 类的 f2 函数和 C 类的 f3 函数就是推荐的第二种测试用例。

[0044] 本发明实施例提供的基于修复缺陷的变更预测回归测试失效的方法, 通过对修复缺陷的变更历史的分析, 获取有影响的因素, 从中选取特征向量, 构建特征模型, 使用 Logistic 回归模型, 对所构建的特征模型进行机器学习, 得到预测模型, 使用该预测模型对修复缺陷的变更是否会导致回归测试失效进行预测, 通过对静态调用图的分析, 对可能导致回归测试失效的测试用例进行推荐。与现有技术相比, 本发明能够提供导致回归测试失效的具体测试用例, 便于快速修复软件源代码中存在的缺陷。

[0045] 本发明实施例还提供一种基于修复缺陷的变更预测回归测试失效的装置, 如图 4 所示, 所述基于修复缺陷的变更预测回归测试失效的装置包括 :

[0046] 特征选取单元 11, 用于通过对修复缺陷的变更历史的分析, 获取有影响的因素, 从中选取特征向量, 构建特征模型 ;

[0047] 模型训练单元 12, 用于使用 Logistic 回归模型, 对所构建的特征模型进行机器学习, 得到预测模型 ;

[0048] 预测单元 13, 用于使用所述预测模型对修复缺陷的变更是否会导致回归测试失效进行预测 ;

[0049] 推荐单元 14, 用于通过对静态调用图的分析, 对可能导致回归测试失效的测试用

例进行推荐。

[0050] 进一步地,所述预测单元 13,可用于将修复缺陷的变更按照时间顺序进行排序,将修复缺陷的变更分为四份,每一份包含等量的引起回归测试失效的变更,使用前三份数据作为训练集来训练模型,使用最后一份数据作为测试集。

[0051] 进一步地,所述推荐单元 14,可用于推荐以下两种可能导致回归测试失效的测试用例:直接测试那些修复缺陷的变更所在的类的测试用例,以及测试调用修复缺陷所在的类的那些类的测试用例。

[0052] 本发明提供的基于修复缺陷的变更预测回归测试失效的装置,通过对修复缺陷的变更历史的分析,获取有影响的因素,从中选取特征向量,构建特征模型,使用 Logistic 回归模型,对所构建的特征模型进行机器学习,得到预测模型,使用该预测模型对修复缺陷的变更是否会导致回归测试失效进行预测,通过对静态调用图的分析,对可能导致回归测试失效的测试用例进行推荐。与现有技术相比,本发明能够提供导致回归测试失效的具体测试用例,便于快速修复软件源代码中存在的缺陷。

[0053] 本发明实施例提供的基于修复缺陷的变更预测回归测试失效的方法及装置,适用于在修复软件源代码中存在的缺陷时提供导致回归测试失效的具体测试用例。

[0054] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体 (Read-Only Memory, ROM) 或随机存储记忆体 (Random Access Memory, RAM) 等。

[0055] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求的保护范围为准。

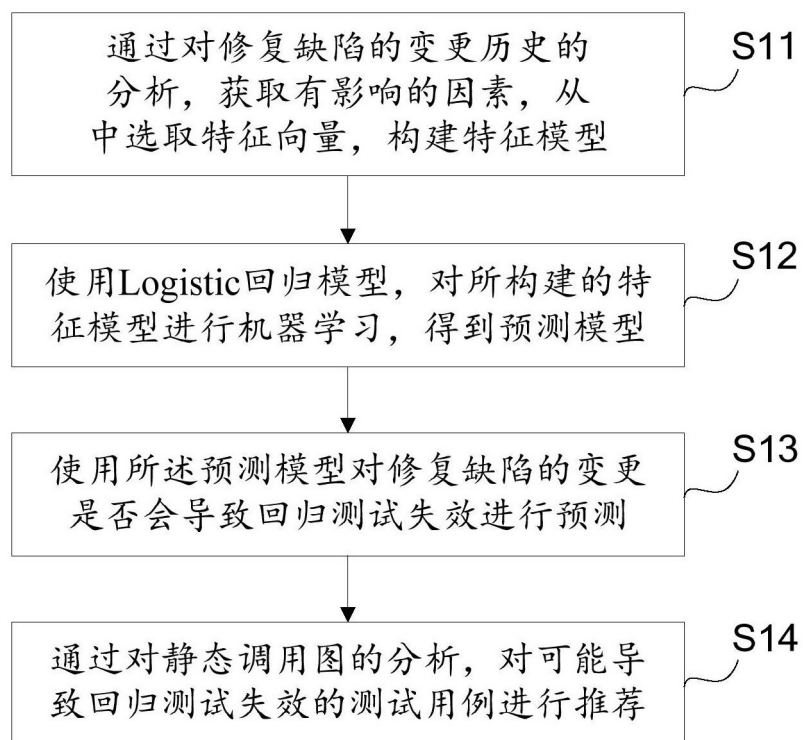


图 1

特征类别	特征名称	类型	定义
大小	LA	Numeric	lines of code added
	LD	Numeric	lines of code deleted
	LF	Numeric	number of files changed
	NC	Numeric	number of changed class
	NM	Numeric	number of changed method
原子变更	AC	Boolean	has added classes
	DC	Boolean	has deleted classes
	AM	Boolean	has added methods
	DM	Boolean	has deleted methods
	CM	Boolean	has changed methods body
	MR	Boolean	has renamed methods
	PC	Boolean	has changed parameters of methods
语义变更	CC	Numeric	number of changed dependencies
	DD	Numeric	number of deleted dependencies
	AD	Numeric	number of added dependencies
	A/RF	Boolean	has added/removed for blocks
	A/RW	Boolean	has added/removed while blocks
	A/RI	Boolean	has added/removed if blocks

图 2

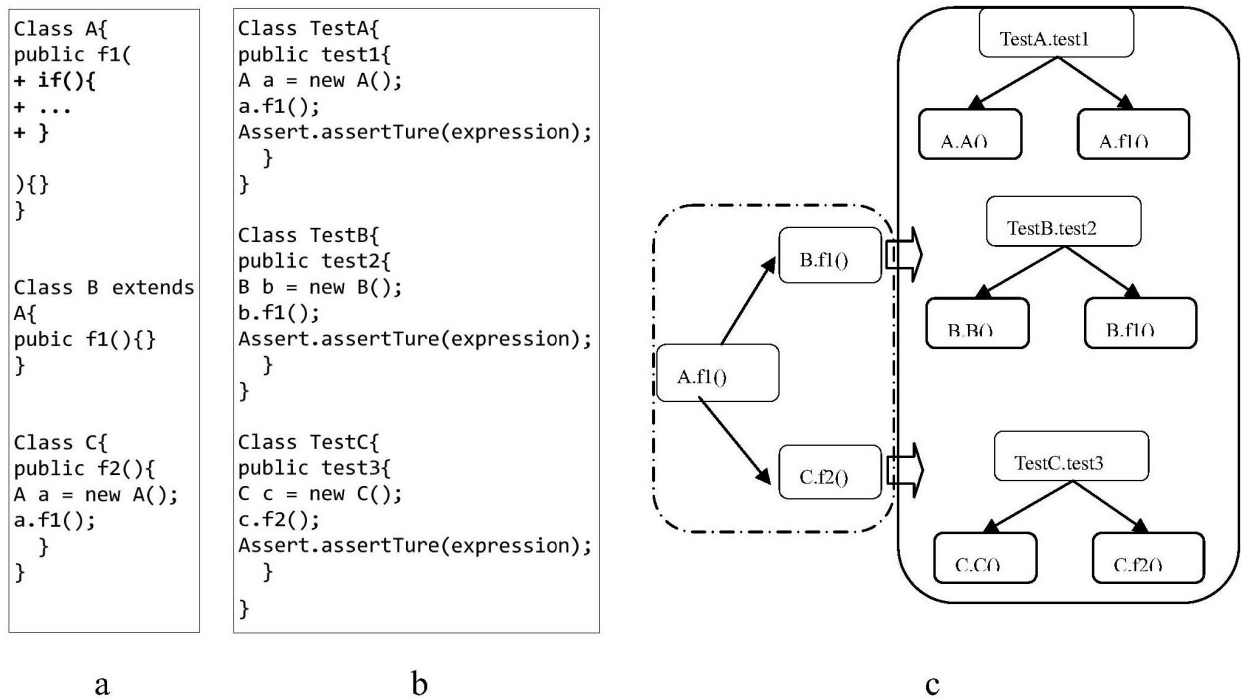


图 3

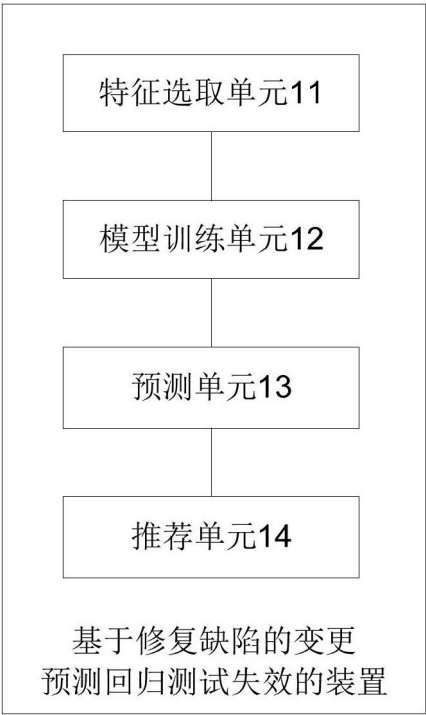


图 4