



(12) 发明专利

(10) 授权公告号 CN 101487876 B

(45) 授权公告日 2011.08.03

(21) 申请号 200910078243.6

CN 1873625 A, 2006.12.06, 全文.

(22) 申请日 2009.02.23

CN 1916920 A, 2007.02.21, 全文.

(73) 专利权人 中国科学院计算技术研究所  
地址 100080 北京市海淀区中关村科学院南路6号

程文青. 一种基于 ATPG 的 ASIC 前端验证环境的构建. 《微电子学与计算机》. 2004, 第 21 卷 (第 6 期), 190-193.

审查员 肖靖

(72) 发明人 郭崎 沈海华 王朋宇

(74) 专利代理机构 北京律诚同业知识产权代理有限公司 11006

代理人 祁建国 梁挥

(51) Int. Cl.

G06F 11/36(2006.01)

G01R 31/3183(2006.01)

(56) 对比文件

JP 特开 2004-355130 A, 2004.12.16, 全文.

US 6141630 A, 2000.10.31, 全文.

CN 1688022 A, 2005.10.26, 全文.

CN 1845079 A, 2006.10.11, 全文.

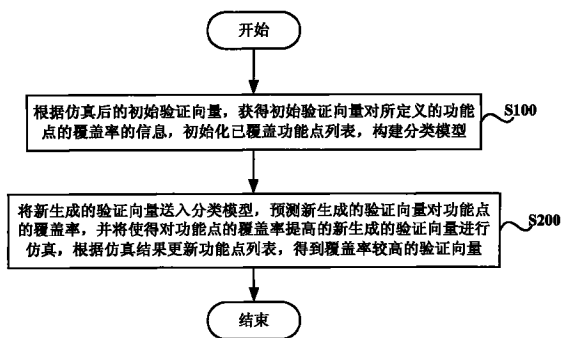
权利要求书 2 页 说明书 9 页 附图 4 页

(54) 发明名称

验证向量的优化方法及装置

(57) 摘要

本发明公开了验证向量的优化方法及装置。该方法,包括下列步骤:根据仿真后的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖功能点列表,构建分类模型;将新生成的验证向量送入所述分类模型,预测新生成的验证向量对所述功能点的覆盖率,并将使得对所述功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量。其能够在达到相同覆盖率的情况下,减少了需要进行仿真的验证向量的数目尽可能减少验证向量的使用,缩短仿真时间,从而提高验证效率,加速验证收敛。



1. 一种验证向量的优化方法,其特征在于,包括下列步骤:

步骤 100,根据仿真后的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖功能点列表,构建分类模型;

步骤 200,将新生成的验证向量的特征向量送入所述分类模型,预测新生成的验证向量对所述功能点的覆盖率,并将使得对所述功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量;

其中,所述步骤 100,包括下列步骤:

步骤 110,生成用户指定数目的初始验证向量集,将其中的初始验证向量在待验证设计上进行仿真,得到所述初始验证向量对所定义的功能点的覆盖率的信息,并将所述初始验证向量和与之对应的覆盖率的信息作为训练样本;

步骤 120,对所述训练样本进行预处理,提取验证向量特征,同时根据对每个所定义的功能点的覆盖情况对验证向量进行类别标记;

步骤 130,对步骤 120 中已进行类别标记的验证向量使用特征选择算法进行特征选择,得到特征向量集;

步骤 140,根据步骤 130 中选择的特征向量集,对验证向量集中的所有验证向量进行特征映射,得到特征矩阵;

步骤 150,将所述特征矩阵送入特定分类器中进行学习,得到分类模型。

2. 根据权利要求 1 所述的验证向量的优化方法,其特征在于,所述步骤 120,包括下列步骤:

步骤 121,对于由验证向量组成的训练样本,获取验证向量的特征表示,提取验证向量特征;

步骤 122,针对所定义的每一个功能点,结合所述训练样本中验证向量和覆盖率的信息,标记出该验证向量对每一个功能点的覆盖情况。

3. 根据权利要求 2 所述的验证向量的优化方法,其特征在于,步骤 122 中,所述标记,是指能覆盖到所述功能点的验证向量被标记为正样例,没有覆盖到所述功能点的验证向量被标记为负样例。

4. 根据权利要求 2 所述的验证向量的优化方法,其特征在于,所述步骤 200,包括下列步骤:

步骤 210,根据步骤 130 中所选择的特征向量所代表的特征,将新生成的验证向量映射为相应的特征向量;

步骤 220,将所述由新生成的验证向量映射的特征向量送入所述分类模型,判断所述新生成的验证向量是否能覆盖到所述功能点,从而预测得到新生成的验证向量对所述功能点的覆盖率,并与所述已覆盖功能点列表中相应表项一一对比;

步骤 230,判断所述新生成的验证向量是否能够覆盖新的功能点,若是,则进入步骤 230;否则,将其丢弃,不进行仿真;

步骤 240,将所述能够覆盖新的功能点的新生成的验证向量送入待验证设计中进行仿真,并用进行了仿真的验证向量的覆盖率结果更新已覆盖功能点列表,得到覆盖率较高的验证向量。

5. 一种验证向量优化装置,所述装置,包括:验证向量生成模块和覆盖率获取模块,其

特征在于,还包括:分类模型构建模块和验证向量优化模块,其中:

所述分类模型构建模块,用于根据经过仿真后的由所述验证向量生成模块生成的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖功能点列表,构建分类模型;

所述验证向量优化模块,用于将由所述验证向量生成模块生成的新生成的验证向量送入所述分类模型构建模块构建的分类模型,预测所述新生成的验证向量对所述功能点的覆盖率,并将使得对所述功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量;

其中,所述分类模型构建模块,包括:

特征提取模块,用于提取所述验证向量的特征;

特征向量选择模块,用于根据所述验证向量及其对所定义的功能点的覆盖率的信息,选择出对分类有效的特征向量;

特征向量映射模块,用于根据所述特征向量选择模块选择出来的特征向量,将验证向量集映射为特征矩阵;

训练模块,用于将所述特征矩阵送入特定分类器中进行训练,得到分类模型。

6. 根据权利要求 5 所述的验证向量优化装置,其特征在于,所述验证向量优化模块,包括:

特征向量获取模块,用于根据所述验证向量生成模块新生成的验证向量及所述提取验证向量的特征向量,获取新生成验证向量的特征向量,送入所述分类模型;

覆盖率预测模块,用于根据所述分类模型预测新生成的验证向量的覆盖率,判断是否能够覆盖到从未覆盖过的功能点;

高覆盖率验证向量获取模块,用于根据所述能够覆盖到从未覆盖过的功能点的新生成的验证向量送入所述覆盖率获取模块后得到的覆盖率信息,更新覆盖率列表,进而获得高覆盖率验证向量。

## 验证向量的优化方法及装置

### 技术领域

[0001] 本发明涉及超大规模集成电路 (Very Large Scale Integrated circuits, VLSI), 设计领域, 主要是关于模拟验证中的覆盖评估技术, 特别是涉及验证向量的优化方法及装置。

### 背景技术

[0002] 随着集成电路的广泛应用, 对功能正确性及速度、功耗、可靠性等都有严格要求。其中, 功能正确性是最基本的要求。2003 年度的国际半导体技术发展报告指出, 验证已经成为集成电路设计流程中开销最大的工作。验证不仅是研究领域中的热点, 它与企业的经济效益也是直接相关的。因此, 必须要有高效的验证的设计方法来将集成电路的设计错误发生的可能性降低到最小, 同时又要缩短产品面市时间, 从而实现经济目标。

[0003] 一般来讲, 验证的设计方法包括两大类: 模拟验证和形式验证。由于状态爆炸等一系列问题的存在, 形式验证难以应用到大型集成电路的设计验证当中去, 因此模拟验证成为当前 VLSI 设计验证中的主要技术; 而由于模拟验证无法证明设计的正确性, 只能说明到某个阶段没有发现错误, 在实际应用中又常常需要结合覆盖评估技术。在模拟验证中, 覆盖评估通过定义各种各样的覆盖率信息可以在一定程度上衡量验证进行的程度。其是通过定义一系列的功能点, 然后检测其是否被覆盖过, 来评估整个验证进度。在整个验证流程中需要始终保持对覆盖率数据的监测和分析, 并以此来指导测试程序的生成和仿真过程。

[0004] 目前, 在处理器的验证中, 随机测试向量是最主要的验证向量, 整个验证过程中超过四分之三的设计错误都是通过运行随机测试向量发现的。理论上来说, 如果能明确测试向量和需要覆盖的功能点之间的关系, 除了指定初始指令模板外, 验证过程完全可以不需要人工干预而自动完成, 但是由于测试向量和功能点之间的复杂网状联系, 在大规模的实际工业应用中还有待深入研究。由于约束随机测试向量的方法可以指定验证的范围, 在仿真初期验证效率可能会比较高, 覆盖率增长较快。但是当仿真进行到一定程度时, 由于所定义覆盖空间的复杂性和其中存在某些难以到达的状态, 覆盖率的提高开始变得十分缓慢, 此时所运行的大量测试程序都具有冗余性, 而在运行仿真前无法知道该测试程序对覆盖率提高的贡献, 大量的仿真时间都用在了已覆盖到的功能点的验证上, 这是影响当前验证效率的重要因素。

[0005] 综上所述, 现有模拟验证技术的仿真速度难以让人满意, 同时所使用的验证向量存在大量的冗余性。

### 发明内容

[0006] 本发明的目的在于过滤对提高当前覆盖率没有帮助的验证向量, 只将可能提高覆盖率的验证向量送入仿真执行的过程, 在达到相同覆盖率的情况下, 减少了需要进行仿真的验证向量的数目, 尽可能减少验证向量的使用, 缩短仿真时间, 从而提高验证效率, 加速验证收敛。

[0007] 为实现本发明的目的而提供一种验证向量的优化方法,包括下列步骤:

[0008] 步骤 100,根据仿真后的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖功能点列表,构建分类模型;

[0009] 步骤 200,将新生成的验证向量送入所述分类模型,预测新的验证向量对所定义的功能点的覆盖率,并将使得对所定义的功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量。

[0010] 所述步骤 100,包括下列步骤:

[0011] 步骤 110,生成用户指定数目的初始验证向量集,将其中的初始验证向量在待验证设计上进行仿真,得到所述初始验证向量对所定义的功能点的覆盖率的信息,并将所述初始验证向量和与之对应的覆盖率的信息作为训练样本;

[0012] 步骤 120,对所述训练样本进行预处理,提取验证向量特征,同时根据对每个所定义的功能点的覆盖情况对验证向量进行类别标记;

[0013] 步骤 130,对步骤 120 中已进行类别标记的验证向量使用特征选择算法进行特征选择,得到特征向量集;

[0014] 步骤 140,根据步骤 130 中选择的特征向量集,对验证向量集中的验证向量进行特征映射,得到特征矩阵;

[0015] 步骤 150,将所述特征矩阵送入特定分类器中进行学习,得到分类模型。

[0016] 所述步骤 120,包括下列步骤:

[0017] 步骤 121,对于由验证向量组成的训练样本,获取验证向量的特征表示,提取验证向量特征;

[0018] 步骤 122,针对所定义的每一个功能点,结合所述训练样本中验证向量和覆盖率的信息,标记出该验证向量对每一个功能点的覆盖情况;

[0019] 步骤 122 中,所述标记,是指能覆盖到所述功能点的验证向量被标记为正样例,没有覆盖到所述功能点的验证向量被标记为负样例。

[0020] 所述步骤 200,包括下列步骤:

[0021] 步骤 210,根据步骤 130 中所选择的特征向量所代表的特征,将新生成的验证向量映射为相应的特征向量;

[0022] 步骤 220,将所述由新生成的验证向量映射的特征向量送入所述分类模型,判断所述新生成的验证向量是否能覆盖到所述功能点,从而预测得到新生成的验证向量对所定义的功能点的覆盖率,并与所述已覆盖功能点列表中相应表项一一对比;

[0023] 步骤 230,判断所述新生成的验证向量是否能够覆盖新的功能点,若是,则进入步骤 240;否则,将其丢弃,不进行仿真;

[0024] 步骤 240,将所述能够覆盖新的功能点的新生成的验证向量送入待验证设计中进行仿真,并用进行了仿真的验证向量的覆盖率结果更新已覆盖功能点列表,得到覆盖率较高的验证向量。

[0025] 为实现本发明的目的还提供一种验证向量优化装置,所述装置,包括:验证向量生成模块和覆盖率获取模块,还包括:分类模型构建模块和验证向量优化模块,其中:

[0026] 所述分类模型构建模块,用于根据经过仿真后的由所述验证向量生成模块生成的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖

功能点列表,构建分类模型;

[0027] 所述验证向量优化模块,用于将由所述验证向量生成模块生成的新生成的验证向量送入所述分类模型构建模块构建的分类模型,预测所述新生成的验证向量对所述功能点的覆盖率,并将使得对所述功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量。

[0028] 所述分类模型构建模块,包括:

[0029] 特征提取模块,用于提取所述验证向量的特征;

[0030] 特征向量选择模块,用于根据所述验证向量及其对所定义的功能点的覆盖率的信息,选择出对分类有效的特征向量;

[0031] 特征向量映射模块,用于根据所述特征向量选择模块选择出来的特征向量,将验证向量集映射为特征矩阵;

[0032] 训练模块,用于将所述特征矩阵送入特定分类器中进行训练,得到分类模型。

[0033] 所述验证向量优化模块,包括:

[0034] 特征向量获取模块,用于根据所述验证向量生成模块新生成的验证向量及所述提取验证向量的特征向量,获取新生成验证向量的特征向量,送入所述分类模型;

[0035] 覆盖率预测模块,用于根据所述分类模型预测新生成的验证向量的覆盖率,判断是否能够覆盖到从未覆盖过的功能点;

[0036] 高覆盖率验证向量获取模块,用于根据所述能够覆盖到从未覆盖过的功能点的新生成的验证向量送入所述覆盖率获取模块后得到的覆盖率信息,更新覆盖率列表,进而获得高覆盖率验证向量。

[0037] 本发明的有益效果在于:

[0038] 1. 本发明的验证向量的优化方法及装置,是一种基于机器学习的验证向量优化方法;

[0039] 2. 本发明的验证向量的优化方法及装置,能够在验证过程中达到相同的覆盖率要求的前提下,尽可能减少验证向量的使用,缩短仿真时间,从而提高验证效率,加速验证收敛。

## 附图说明

[0040] 图1是本发明验证向量的优化方法的流程图;

[0041] 图2是本发明中构建分类模型的流程图;

[0042] 图3是本发明中对新生成的验证向量进行优化的流程图;

[0043] 图4是本发明验证向量优化装置的结构示意图;

[0044] 图5是本发明中分类模型构建模块的结构示意图;

[0045] 图6是本发明中验证向量优化模块的结构示意图。

## 具体实施方式

[0046] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明的一种验证向量的优化方法及装置进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0047] 本发明的一种验证向量的优化方法及装置,是根据已进行的验证向量的仿真信息,对新产生的验证向量的覆盖率信息进行预测,提前判断其是否可能提高当前覆盖率,如果可以再送入到待验证设计执行仿真操作,否则认为是冗余的验证向量而重新产生。其能够在验证过程中达到相同的覆盖率要求的前提下,尽可能减少验证向量的使用,缩短仿真时间,从而提高验证效率,加速验证收敛。

[0048] 下面结合上述目标详细介绍本发明的一种验证向量的优化方法,其是一种基于机器学习的验证向量优化方法,其目的在于,解决现有验证流程中由于大量的仿真时间用在了冗余的验证向量的运行上而导致的验证效率不够高的问题。

[0049] 图 1 是本发明一种验证向量的优化方法的流程图,如图 1 所示,所述方法包括下列步骤:

[0050] 步骤 100,根据仿真后的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖功能点列表,构建分类模型;

[0051] 所述功能点,是用户根据待验证设计的需要定义的。

[0052] 步骤 200,对新生成的验证向量进行优化。具体包括:将新生成的验证向量的特征向量送入分类模型,预测新生成的验证向量对功能点的覆盖率,并将使得对所述功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量。

[0053] 图 2 是本发明中构建分类模型的流程图,如图 2 所示,所述步骤 100,包括下列步骤:

[0054] 步骤 110,生成用户指定数目的初始验证向量集,将其中的初始验证向量在待验证设计上进行仿真,得到相应初始验证向量对所定义的功能点的覆盖率信息,并将所述初始验证向量和与之对应的覆盖率信息作为训练样本;

[0055] 在实际验证工作中,可以通过人工手写验证向量,或者利用验证向量生成软件等方法生成用户指定数目的初始验证向量集。较佳地,作为一种可实施方式,本发明中是利用指令生成引擎来生成用户指定数目的初始验证向量集。

[0056] 所述初始验证向量的覆盖率信息,是指初始验证向量是否覆盖过用户根据待验证设计的需要定义的一系列功能点的情况。

[0057] 在实际验证工作中,所述覆盖率信息可以通过用户自己设计的覆盖率采集工具等方式在仿真过程中自动获得相应功能点的覆盖率报告。较佳地,作为一种可实施方式,本发明中是利用电子设计自动化 (Electronic Design Automation, EDA) 工具得到相应功能点的覆盖率报告,其中包括每一个验证向量对所定义的所有功能点的覆盖次数等信息。

[0058] 作为一种可实施方式,在指令级随机验证中,验证向量对应的是由指令生成引擎产生的指令序列,如 `mfhi $10 ;lb $31,24251($2)` 等。对于功能覆盖率而言,需要由验证工程师和设计者共同编写功能覆盖率模型。例如针对微处理的 Decoder 模块,可能重点在于验证是否对于所有的指令操作码都覆盖到了,得到的可能结果如: `{mfhi 3 ;lb 2}`,表示的是对 `mfhi` 指令覆盖了 3 次,而对 `lb` 指令覆盖了 2 次

[0059] 步骤 120,对所述训练样本进行预处理,提取验证向量特征,同时根据对每个所定义的功能点的覆盖情况对验证向量进行类别标记;

[0060] 对于由验证向量组成的训练样本,要对其进行训练,首先要获取验证向量的特征

表示,并提取验证向量特征。然后,针对所定义的每一个功能点,结合训练样本的相应信息,标记出该验证向量对该功能点的覆盖情况,其中能覆盖到该功能点的验证向量称为正样例,没有覆盖到该功能点的验证向量称为负样例。

[0061] 所述步骤 120,包括下列步骤:

[0062] 步骤 121,对于由验证向量组成的训练样本,获取验证向量的特征表示,提取验证向量特征;

[0063] 对于微处理器的指令级随机验证而言,验证向量意味着由约束引擎产生的一系列指令序列。指令序列最直观表示就是其汇编代码,由于指令序列对功能点的覆盖情况不仅仅表现在单条指令所能覆盖到的功能点,同时更多的功能点是由指令的不同组合序列才可能覆盖到的,因此采用指令的顺序组合作为验证向量的特征表示是比较自然的方式。

[0064] 在文本处理领域,一般采用向量空间模型来表示样本,即表示成为一个 M 维的无序向量,其中 N-Gram 是文本表示中使用较为广泛的概念,对于自然语言而言,连续的词语之间是有必要关联的,所以才用 N-Gram 来表示文本特征。

[0065] 除了使用 N-Gram 作为特征表示,还可能采用别的特征表示方式来进行,可能不同的设计者有不同的考虑,比如可能会使用验证向量的二进制序列,将整个 01 序列作为该验证向量的特征向量,只要能将其表示为向量空间模型就可以了。作为一种可实施方式,本发明中的验证向量的汇编表示使用 N-Gram 可以得到以下的特征表示,如表 1 所示:

[0066] 表 1 汇编码的 2-Gram 特征表示

[0067]

汇编码	2-Gram 特征
mfhi \$10	mfhi \$10 lb \$31, 24251(\$2)
lb \$31, 24251(\$2)	lb \$31, 24251(\$2) sub \$5, \$13,
sub \$5, \$13, \$11	\$11 sub \$5, \$13, \$11 addu \$19, \$20,
addu \$19, \$20, \$24	\$24 addu \$19, \$20, \$24 swl \$10,
swl \$10, -15728(\$2)	-15728(\$2)

[0068] 对于 RISC 处理器,除了 load 和 store 等访存类指令,其它指令一般都只能访问寄存器,因此指令的操作码就规定了该条指令能够访问内存单元还是寄存器。对于上述特征而言,由于考虑了操作数而使得特征分布过于稀疏,不利于分类器进行学习;同时对于某些情况,如:

[0069] add \$19, \$20, \$24      add \$18, \$19, \$23

[0070] sub \$11, \$10, \$12      sub \$12, \$11, \$13

[0071] 以上两种情况所能覆盖到的功能点并无很大区别,所以进行权衡之后,考虑对上述特征进行简化作为验证向量特征,只使用其操作码部分,因此得到的验证向量特征如表 2 所示:

[0072] 表 2 简化的汇编码的 2-Gram 特征

[0073]



汇编码	2-Gram 特征
mfhi \$10	mfhi lb
lb \$31, 24251(\$2)	lb sub
sub \$5, \$13, \$11	sub addu
addu \$19, \$20, \$24	addu swl
swl \$10, -15728(\$2)	

[0074] 步骤 122, 针对所定义的每一个功能点, 结合所述训练样本中验证向量和覆盖率的信息, 标记出该验证向量对每一个功能点的覆盖情况;

[0075] 能覆盖到功能点的验证向量被标记为正样例, 没有覆盖到功能点的验证向量被标记为负样例。

[0076] 步骤 130, 对步骤 120 中已进行类别标记的验证向量使用特征选择算法进行特征选择, 得到特征向量集;

[0077] 由于步骤 120 中所提取的验证向量特征的数目是十分巨大的, 直接将其映射为特征向量会使得特征空间维数过多, 可能造成“维数灾难”等问题, 影响学习速度, 同时也可能有噪声成分的存在, 将影响机器学习的准确度。因此, 需要通过特征选择算法从大量可选特征中提取出对区分不同类别更为重要的特征来表征原有的验证向量, 以减少数据噪声, 降低特征维数, 使得既能保证训练的速度, 同时又能得到最佳的分类准确度。

[0078] 所述特征选择算法是机器学习领域的现有技术, 文本处理中的特征选择方式都可以采用, 如文档频数 (DF)、信息增益 (Information Gain)、互信息 (Mutual Information), CHI 统计等等。

[0079] 作为一种可实施方式, 本发明使用信息增益来进行特征向量的挑选。信息增益源于信息论, 反映了一个特征在区别不同类别时的重要程度, 通过计算各个特征的信息增益值来剔除得分较低的特征, 是因为这些特征对分类的帮助十分有限。信息增益的概念建立在变量熵的基础上, 变量的熵表示该变量的不确定程度, 一个变量 X 的熵值按如下方式定义:

$$[0080] \quad H(X) = -\sum_{i=1}^n P(x_i) \times \log_2(P(x_i))$$

[0081] 其中  $x_i$  是变量 X 的所有可能取值。

[0082] 信息增益则是在增加了对变量的观测值后, 熵的变化情况:

$$[0083] \quad IG(X, Y) = H(X) - H(X|Y)$$

[0084] 对应到我们的实际问题, 假设对于验证向量集 S, 针对某个功能点, 覆盖到该功能点的样本数为 p, 未覆盖到该功能点的样本数为 n, 总的样本数  $t = p+n$ , 对于某个特征 f 在样本集 S 中的信息增益表示为:

$$[0085] \quad IG(S, f) = H(S) - \sum_{v \in \text{values}(f)} \frac{p_v + n_v}{p+n} H(I_v)$$

[0086] 由于这里是二分类问题, S 的取值只有正样例和负样例两种情况, 分别对应覆盖到

某个功能点的样例和未覆盖到某个功能点的样例。此外,特征  $f$  只有两种取值,  $\text{values}(f) = \{0, 1\}$ , 分别表示属性  $f$  出现在某个样例中和不在某个样例中出现的情况。由此得到各个符号具体含义如下:

[0087]  $p_0$ : 不包含特征  $f$  的正样例;

[0088]  $n_0$ : 不包含特征  $f$  的负样例;

[0089]  $I_0$ : 不包含特征  $f$  的所有样例, 即  $p_0+n_0$ ;

[0090]  $p_1$ : 包含特征  $f$  的正样例;

[0091]  $n_1$ : 包含特征  $f$  的负样例;

[0092]  $I_1$ : 包含特征  $f$  的所有样例, 即  $p_1+n_1$ ;

[0093] 在所有特征的信息增益计算出来后, 提取其中最高的 800 项作为最终的特征向量集。

[0094] 步骤 140, 根据步骤 130 中选择的特征向量集, 对验证向量集中所有验证向量进行特征映射, 得到特征矩阵;

[0095] 所述特征映射可以采取布尔映射、出现次数映射(根据该特征出现的次数)、出现频率映射(根据该特征出现的频率)等多种方式。较佳地, 本发明采用的映射方法是直接地进行布尔(Boolean)映射: 对于特征  $j$ , 如果在验证向量  $i$  中出现过, 则特征矩阵  $m[i][j] = 1$ , 否则  $m[i][j] = 0$ 。

[0096] 步骤 150, 将所述特征矩阵送入特定分类器中进行学习, 得到分类模型;

[0097] 现有的机器学习理论中得到广泛应用的分类器, 如  $k$  近邻(kNN), 朴素贝叶斯(naïve Bayes), 判定树(Decision tree)等等都可以对特征矩阵进行训练, 从而得到分类模型。

[0098] 较佳地, 作为一种可实施方式, 本发明使用支持向量机对步骤 140 中得到的特征矩阵进行训练, 得到相应的分类模型。该方法为现有技术, 在此不再一一赘述。

[0099] 图 3 是本发明中对新生成的验证向量进行优化的流程图, 如图 3 所示, 所述步骤 200, 包括下列步骤:

[0100] 步骤 210, 根据训练过程中所选择的特征, 将新生成的验证向量映射为相应的特征向量;

[0101] 所述特征, 是步骤 130 中所选择的特征向量所代表的特征。

[0102] 步骤 220, 将所述由新生成的验证向量映射的特征向量送入所述分类模型, 判断所述新生成的验证向量是否能覆盖到所述功能点, 从而预测得到新生成的验证向量对所述功能点的覆盖率, 并与所述已覆盖功能点列表中相应表项一一对比;

[0103] 本步骤 210、220 中, 将新生成的验证向量映射为特征向量的方法和得到覆盖率的方法同步骤 140 和 110, 只是将步骤 110 中, 将初始验证向量在待验证设计上进行仿真改为根据训练过程中所选择的特征。因此不再一一赘述。

[0104] 步骤 230, 判断所述新生成的验证向量是否能够覆盖新的功能点, 若是, 则进入步骤 240; 否则, 将其丢弃, 不进行仿真;

[0105] 所述新的功能点, 指的是以前所有已仿真过的验证向量从未覆盖到的功能点, 即不在全局的已覆盖到功能点列表中的功能点。

[0106] 步骤 240, 将所述能够覆盖新的功能点的新生成的验证向量送入待验证设计中进

行仿真,并用进行了仿真的验证向量的覆盖率结果更新已覆盖功能点列表,得到覆盖率较高的验证向量。

[0107] 对于所有新产生的验证向量,通过不断重复上述预测过程,可以得到其对定义的所有功能点的覆盖情况,如果预测该验证向量能够覆盖到的功能点都是已覆盖到功能点列表中的功能点,那么说明该验证向量是冗余的,没有必要进行仿真,将其丢弃。但是如果预测该验证向量能够覆盖到的功能点不在当前的已覆盖到功能点列表中,则说明该验证向量可能覆盖到新的功能点,将其进行仿真,并用仿真的结果更新已覆盖到功能点列表。上述过程完全由机器执行,实现了基于机器学习的验证向量优化方法。

[0108] 本步骤中,根据进行仿真后得到的覆盖率报告,对全局的已覆盖到功能点列表进行更新,表示验证进行到当前对所有功能点的累积覆盖情况。用于判断新生成的验证向量是否对当前的覆盖率有所提高。因为通过预测过程得到的是可能提高覆盖率的验证向量,经过预测过程仅仅只能知道该验证向量对每个功能点的覆盖与否的情况,具体是否能够对当前覆盖率有所提高,是通过与全局的已覆盖到功能点列表进行比对来得到的。这样验证过程中在达到相同的覆盖率要求的前提下,能够尽可能减少验证向量的使用,缩短仿真时间,从而提高验证效率,加速验证收敛。

[0109] 相应于本发明的一种验证向量的优化方法,本发明还提供一种验证向量优化装置。

[0110] 图4是本发明一种验证向量优化装置的结构示意图,如图4所示,所述验证向量优化装置,包括:验证向量生成模块1,用于生成用户指定数目的验证向量集;覆盖率获取模块2,用于将验证向量在待验证设计上进行仿真,得到相应验证向量的覆盖率信息,还包括:分类模型构建模块3,验证向量优化模块4,其中:

[0111] 所述分类模型构建模块3,用于根据经过仿真后的由所述验证向量生成模块1生成的初始验证向量,获得所述初始验证向量对所定义的功能点的覆盖率的信息,初始化已覆盖功能点列表,构建分类模型;

[0112] 所述验证向量优化模块4,用于将由所述验证向量生成模块1生成的新生成的验证向量送入所述分类模型构建模块3构建的分类模型,预测所述新生成的验证向量对所述功能点的覆盖率,并将使得对所述功能点的覆盖率提高的新生成的验证向量进行仿真,根据仿真结果更新功能点列表,得到覆盖率较高的验证向量。

[0113] 图5是本发明中分类模型构建模块的结构示意图,如图5所示,所述分类模型构建模块3,包括:

[0114] 特征提取模块31,用于提取所述验证向量的特征;

[0115] 特征向量选择模块32,用于根据所述验证向量及其对所定义的功能点的覆盖率的信息,选择出对分类有效的特征向量;

[0116] 特征向量映射模块33,用于根据所述特征向量选择模块32选择出来的特征向量,将验证向量集映射为特征矩阵;

[0117] 训练模块34,用于将所述特征矩阵送入特定分类器中进行训练,得到分类模型。

[0118] 图6是本发明中验证向量优化模块的结构示意图,如图6所示,所述验证向量优化模块4,包括:

[0119] 特征向量获取模块41,用于根据所述验证向量生成模块新生成的验证向量及所述

提取验证向量的特征向量,获取新生成验证向量的特征向量,送入所述分类模型;

[0120] 覆盖率预测模块 42,用于根据所述分类模型预测新生成验证向量的覆盖率,判断是否能够覆盖到从未覆盖过的功能点;

[0121] 高覆盖率验证向量获取模块 43,用于根据所述能够覆盖到从未覆盖过的功能点的新生成验证向量送入所述覆盖率获取模块 2 后得到的覆盖率信息,更新覆盖率列表,进而获得高覆盖率验证向量。

[0122] 本发明的有益效果在于:

[0123] 1. 本发明的一种验证向量的优化方法及装置,是一种基于机器学习的验证向量优化方法;

[0124] 2. 本发明的一种验证向量的优化方法及装置,能够在验证过程中达到相同的覆盖率要求的前提下,尽可能减少验证向量的使用,缩短仿真时间,从而提高验证效率,加速验证收敛。

[0125] 通过结合附图对本发明具体实施列的描述,本发明的其它方面及特征对本领域的技术人员而言是显而易见的。

[0126] 以上对本发明的具体实施列进行了描述和说明,这些实施列应被认为其只是示例性的,并不用于对本发明进行限制,本发明应根据所附的权利要求进行解释。

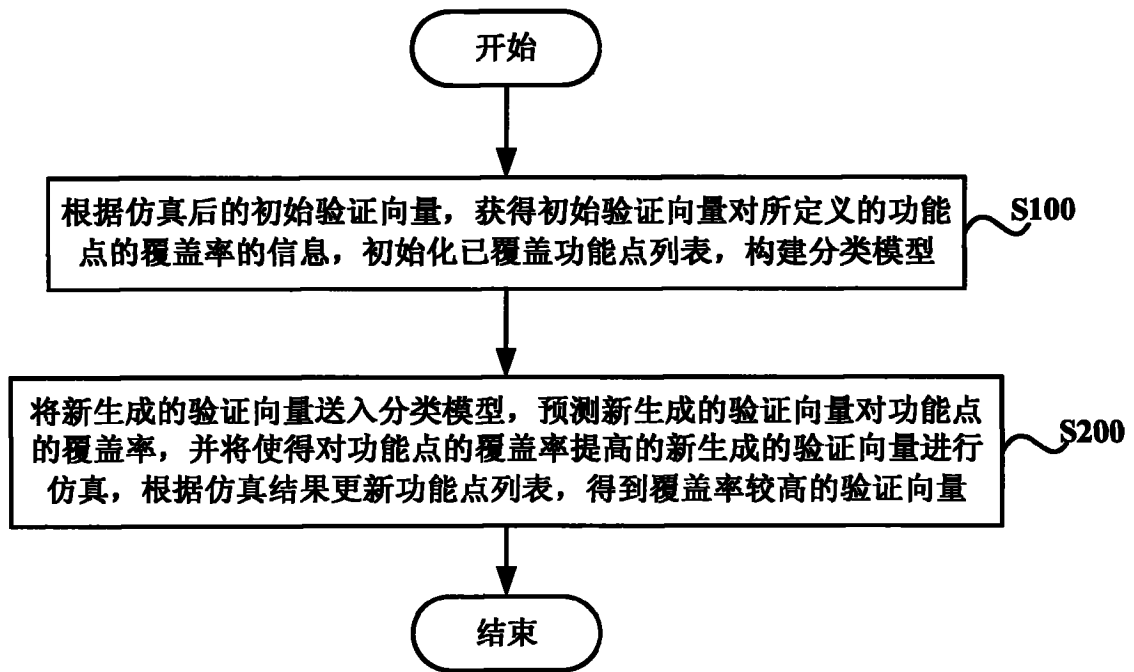


图 1

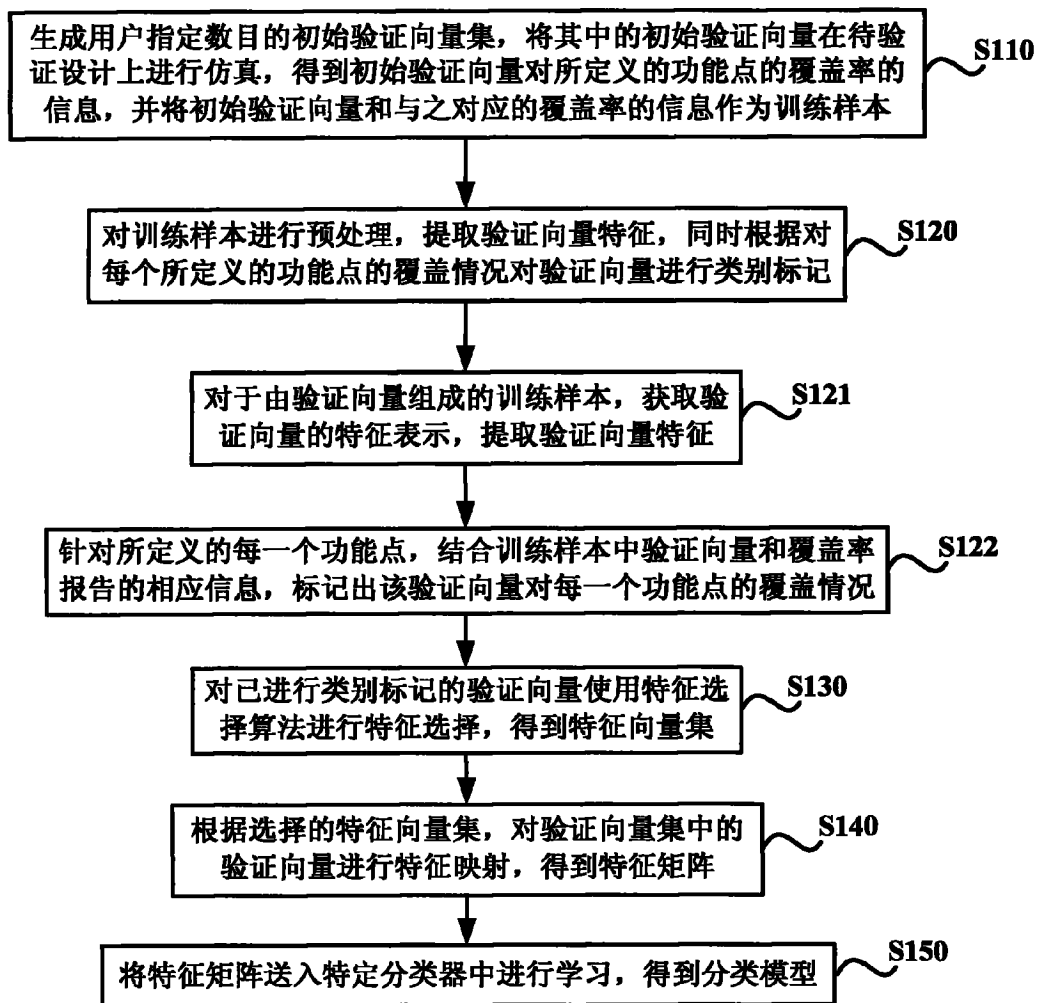


图 2

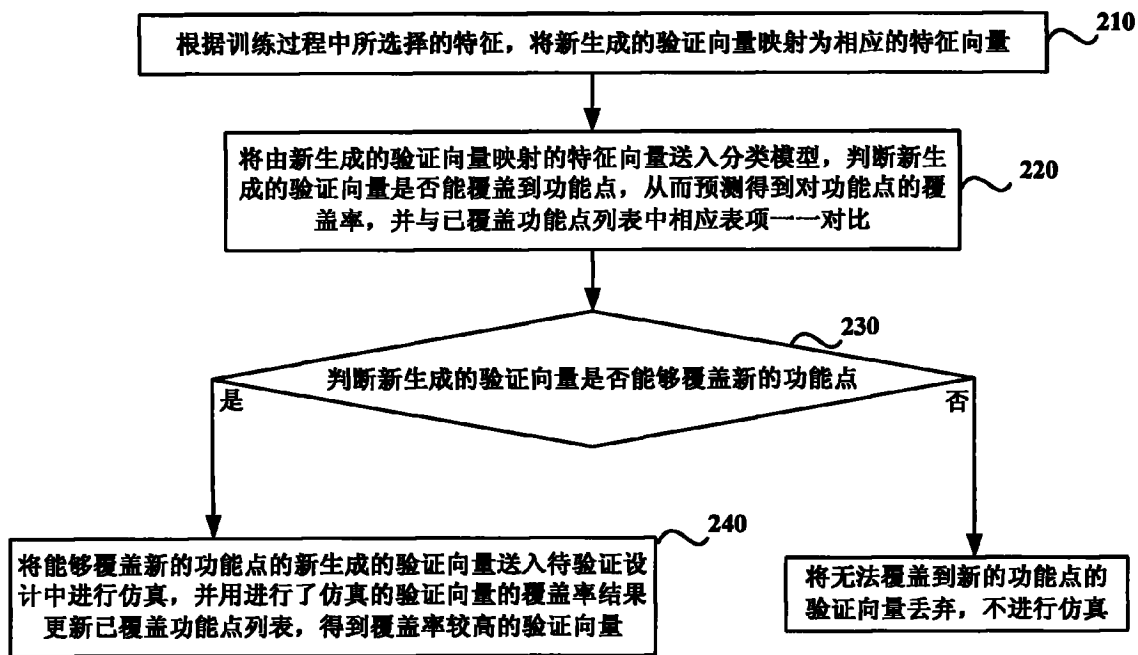


图 3

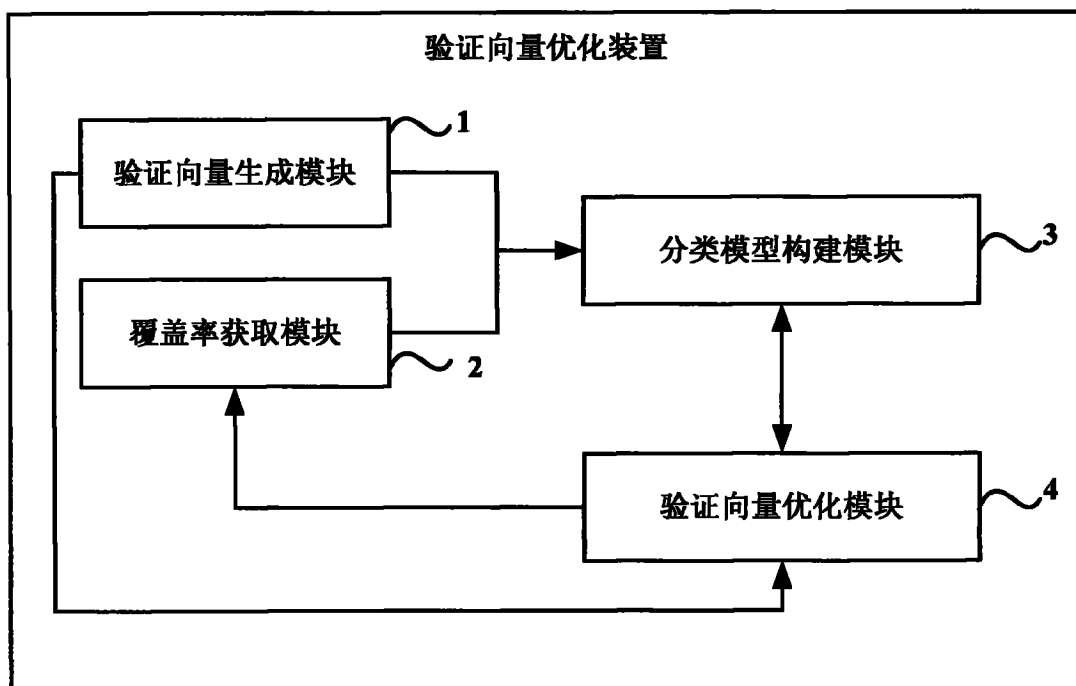


图 4

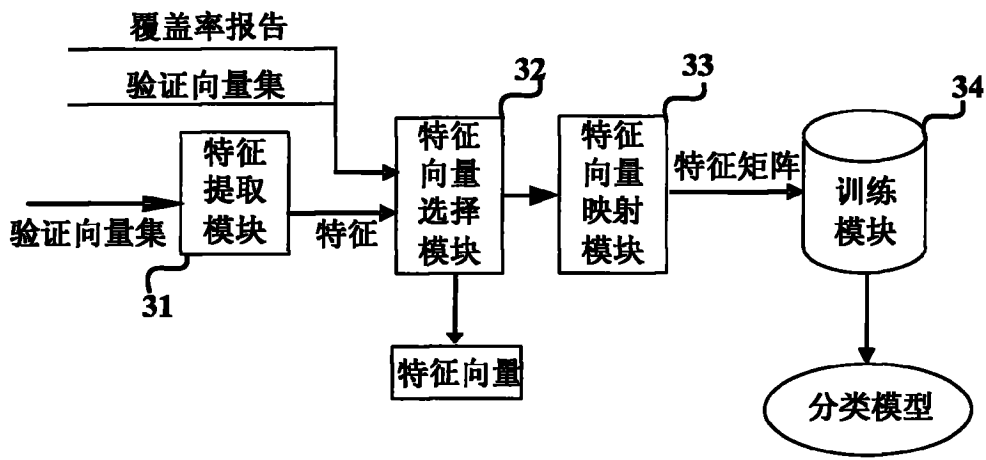


图 5

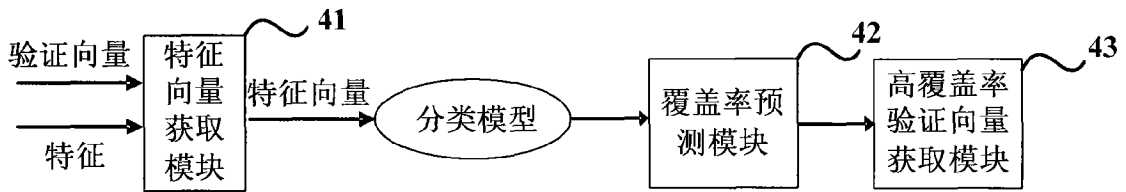


图 6