(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2014 年 9 月 4 日 (04.09.2014)

WIPO | PCT

(10) 国际公布号
WO 2014/131262 A1

(51) 国际专利分类号:
G06F 19/00 (2011.01)

(21) 国际申请号: PCT/CN2013/080279

(22) 国际申请日: 2013 年 7 月 29 日 (29.07.2013)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201310066324.0 2013 年 2 月 28 日 (28.02.2013) CN

(71) 申请人: **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: **陈焕华 (CHEN, Huanhua)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **潘璐伽 (PAN, Lujia)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: **北京中博世达专利商标代理有限公司 (BEIJING ZBSD PATENT&TRADEMARK AGENT LTD.)**; 中国北京市海淀区大柳树路 17 号富海大厦 B 座 501 室, Beijing 100081 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) **Title**: DEFECT PREDICTION METHOD AND DEVICE

(54) **发明名称** : 一种缺陷预测方法及装置

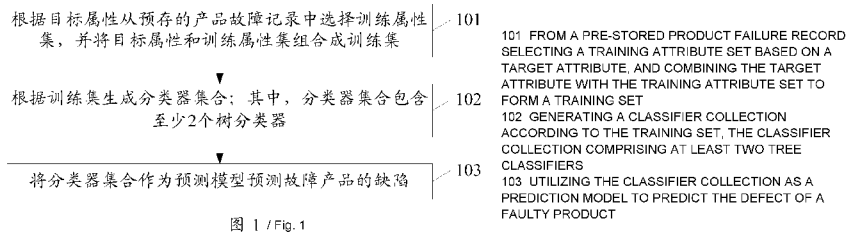

图 1 / Fig. 1

(57) **Abstract**: The present invention relates to the field of data processing. A defect prediction method and device, the method comprising: from a pre-stored product failure record selecting a training attribute set based on a target attribute, and combining the target attribute with the training attribute set to form a training set (101); the target attribute is the defect attribute of a historical faulty product; generating a classifier collection according to the training set, the classifier collection comprising at least two tree classifiers (102); and utilizing the classifier collection as a prediction model to predict the defects of a faulty product (103). The method is used in the defect prediction process of a faulty product to realize accurate and quick locating of a faulty product.

(57) **摘要**: 一种缺陷预测方法及装置，涉及数据处理领域。该方法包括：根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集（101）；所述目标属性为历史故障产品的缺陷属性；根据所述训练集生成分类器集合，所述分类器集合包含至少 2 个树分类器（102）；将所述分类器集合作为预测模型预测故障产品的缺陷（103）。上述方法用于故障产品的缺陷预测的过程中，实现对故障产品的准确及快速定位。

WO 2014/131262 A1

一种缺陷预测方法及装置

本申请要求于 2013 年 02 月 28 日提交中国专利局、申请号为 201310066324.0、发明名称为“一种缺陷预测方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及数据处理领域，尤其涉及一种缺陷预测方法及装置。

背景技术

随着时代的发展，能够满足人们需求的产品种类和数量逐渐增多，产品的质量是也已成为用户及企业关心的主要问题，特别是尤其对于企业来说，产品的质量就是企业的根本，因此降低产品的缺陷率对企业至关重要。而引起产品缺陷的原因主要是产品的生产工艺，包括产品的设计、所使用材料的质量、生产商能力等，因此对于企业来讲，若想降低产品的缺陷率，就需要分析并改进产品的生产工艺，从而提高产品质量。

每个产品都有关于该产品各方面的信息的记录，如原料来源、生产信息、测试信息、运输信息、使用信息等等，而当产品在使用或者生产过程中出现某一类型的缺陷或者故障时，引起这类缺陷或故障的因素和记录的该产品的信息具有一定的关联性。

现有技术提供一种故障产品缺陷预测方法，具体为利用记录的出现过故障的产品的信息，通过基于决策树的分类算法生成单一决策树，此时当产品出现故障时，便可以根据生成的决策树对故障产品的缺陷进行预测。而当记录的出现过故障的产品的信息的分类标签较多时，采用基于决策树的分类算法产生的单一决策树就容易引起过拟合或欠拟合，从而导致无法进行缺陷预测。因此当产品出现缺陷或者故障时，如何快速的定位故障点，并查找到故障原因已成为业界研究的重点。

发明内容

本发明的实施例提供一种缺陷预测方法及装置，实现了对故障产品的缺陷的准确及快速定位。

本发明的第一方面，提供一种缺陷预测方法，包括：

根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集；其中，所述目标属性为历史故障产品的缺陷属性；

根据所述训练集生成分类器集合；其中，所述分类器集合包含至少 2 个树分类器；

将所述分类器集合作为预测模型预测故障产品的缺陷。

结合第一方面，在一种可能的实现方式中，所述训练集包含 M 个训练单元，每个训练单元包含一个目标属性和一个训练属性集；

所述根据所述训练集生成分类器集合，包括：

从所述训练集中选取第一训练子集；

根据预设策略生成与所述第一训练子集相对应的第一树分类器；

从所述训练集中选取第二训练子集；

根据预设策略生成与所述第二训练子集相对应的第二树分类器；

从所述训练集中选取第 N 训练子集；其中，所述第 N 训练子集包含 M'个训练单元，所述 M'小于等于所述 M；

根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器；其中，所述 N 为大于等于 2 的整数；

将 N 个树分类器组合生成所述分类器集合。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，还包括：

当生成第 K-1 树分类器时，获取生成的 K-1 个树分类器的错误率；

当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率；以便当所述 K 个树分类器的错误率和所述 K-1 个树分类器的错误率

的差值小于预设的阈值时，将所述 K 个树分类器组合生成所述分类器集合；其中，所述 K 为小于等于 N 的整数。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，所述当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率，包括：

根据第一训练单元从所述分类器集合中选取第一类树分类器；

根据所述第一类树分类器生成所述第一训练单元的第一预测标签；

根据第二训练单元从所述分类器集合中选取第二类树分类器；

根据所述第二类树分类器生成所述第二训练单元的第二预测标签；

根据第 M 训练单元从所述分类器集合中选取第 M 类树分类器；其中，所述第 M 类树分类器为未使用第 M 训练单元生成树分类器的分类器集合,所述 M 为训练集中包含训练单元的个数；

根据所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签；

根据 M 个预测标签获取所述生成的 K 个树分类器的错误率。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，所述根据所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签，具体包括：

根据 $C^{OOB}(M,\ x_M)=\arg\max_{y}\sum_{C_j\in C_M^{OOB}}h(\varepsilon_j)I(C_j(x_M)=y)$ 生成所述第 M 预测标签；其中，$C^{OOB}(M,\ x_M)$ 为所述第 M 训练单元的第 M 预测标签，C_j 为第 j 树分类器，C_M^{OOB} 为所述第 M 类树分类器，$h(\varepsilon_j)$ 为第 j 树分类器的权重，$C_j(x_M)$ 为根据所述第 j 树分类器和所述第 M 训练单元中包含的训练属性集得到的目标属性，$y\in Y$，Y 为分类标签集合。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，所述根据 M 个预测标签获取所述生成的 K 个树分类器的错误率，具体包括：

根据 $E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$ 获取所述生成的 K 个树分类器的错误率；其中，$E(T)$ 为所述生成的 K 个树分类器的错误率，M 为所述训练集中训练单元的个数，$C^{OOB}(r,\ x_r)$ 为所述第 r 训练单元的第 r 预测标签，y_r 为第 r 训练单元的目标属性。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，在所述根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器之后，还包括：

从所述训练集中选取第 N'训练子集；其中，所述第 N'训练子集与所述第 N 训练子集的交集为空，所述第 N'训练子集包含至少一个训练单元；

根据所述第 N'训练子集获取所述第 N 树分类器的误预测率；

根据所述第 N 树分类器误预测率获取所述第 N 树分类器的权重。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，所述将所述分类器集合作为预测模型预测故障产品的缺陷，包括：

统计所述故障产品的属性信息；

根据所述属性信息将所述分类器集合作为预测模型预测所述故障产品的缺陷得到分类标签集合；

根据所述分类器集合和所述分类器集合中每个树分类器的权重，获取所述分类标签集合中每个分类标签的信任值。

结合第一方面和上述可能的实现方式，在另一种可能的实现方式中，所述预设策略包括决策树算法。

本发明的第二方面，提供一种缺陷预测装置，包括：

处理单元，用于根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集；其中，所述目标属性为历史故障产品的缺陷属性；

生成单元，用于根据所述处理单元得到的训练集生成分类器集

合；其中，所述分类器集合包含至少 2 个树分类器；

预测单元，用于将所述生成单元生成的分类器集合作为预测模型预测故障产品的缺陷。

结合第二方面，在一种可能的实现方式中，所述训练集包含 M 个训练单元，每个训练单元包含一个目标属性和一个训练属性集；

所述生成单元，包括：

选取模块，用于从所述处理单元得到的所述训练集中选取第一训练子集；

生成模块，用于根据预设策略生成与所述选取模块选取的所述第一训练子集相对应的第一树分类器；

所述选取模块，还用于从所述处理单元得到的所述训练集中选取第二训练子集；

所述生成模块，还用于根据预设策略生成与所述选取模块选取的所述第二训练子集相对应的第二树分类器；

所述选取模块，还用于从所述处理单元得到的所述训练集中选取第 N 训练子集；其中，所述第 N 训练子集包含 M'个训练单元，所述 M'小于等于所述 M；

所述生成模块，还用于根据预设策略生成与所述选取模块选取的所述第 N 训练子集相对应的第 N 树分类器；其中，所述 N 为大于等于 2 的整数；

组合模块，用于将所述生成模块生成的 N 个树分类器组合生成所述分类器集合。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，所述生成单元还包括：

第一获取模块，用于当生成第 K-1 树分类器时，获取生成的 K-1 个树分类器的错误率；

第二获取模块，用于当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率；以便当所述 K 个树分类器的错误率和所述 K-1 个树分类器的错误率的差值小于预设的阈值时，将所述 K 个树分类

器组合生成所述分类器集合；其中，所述 K 为小于等于 N 的整数。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，所述第二获取模块，包括：

选取子模块，用于根据第一训练单元从所述分类器集合中选取第一类树分类器；

生成子模块，用于根据所述选取子模块选取的所述第一类树分类器生成所述第一训练单元的第一预测标签；

所述选取子模块，还用于根据第二训练单元从所述分类器集合中选取第二类树分类器；

所述生成子模块，还用于根据所述选取子模块选取的所述第二类树分类器生成所述第二训练单元的第二预测标签；

所述选取子模块，还用于根据第 M 训练单元从所述分类器集合中选取第 M 类树分类器；其中，所述第 M 类树分类器为未使用第 M 训练单元生成树分类器的分类器集合,所述 M 为训练集中包含训练单元的个数；

所述生成子模块，还用于根据所述选取子模块选取的所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签；

获取子模块，用于根据所述生成子模块生成的 M 个预测标签获取所述生成的 K 个树分类器的错误率。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，所述生成子模块，具体用于：

根据 $C^{OOB}(M,\ x_M)=\arg\max_{y}\sum_{C_j\in C_M^{OOB}}h(\varepsilon_j)I(C_j(x_M)=y)$ 生成所述第 M 预测标签；其中，$C^{OOB}(M,\ x_M)$ 为所述第 M 训练单元的第 M 预测标签，C_j 为第 j 树分类器，C_M^{OOB} 为所述第 M 类树分类器，$h(\varepsilon_j)$ 为第 j 树分类器的权重，$C_j(x_M)$ 为根据所述第 j 树分类器和所述第 M 训练单元中包含的训练属性集得到的目标属性，$y\in Y$，Y 为分类标签集合。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，所述获取子模块，具体用于：

根据 $E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$ 获取所述生成子模块生成的 K 个树分类器的错误率；其中，$E(T)$ 为所述生成的 K 个树分类器的错误率，M 为所述训练集中训练单元的个数，$C^{OOB}(r,\ x_r)$ 为所述第 r 训练单元的第 r 预测标签，y_r 为第 r 训练单元的目标属性。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，还包括：

选取单元，用于在所述生成模块根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器之后，从所述训练集中选取第 N' 训练子集；其中，所述第 N'训练子集与所述第 N 训练子集的交集为空，所述第 N'训练子集包含至少一个训练单元；

第一获取单元，用于根据所述选取单元选取的所述第 N'训练子集获取所述第 N 树分类器的误预测率；

第一获取单元，用于根据所述第一获取单元获取到的所述第 N 树分类器误预测率获取所述第 N 树分类器的权重。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，所述预测单元包括：

统计模块，用于统计所述故障产品的属性信息；

预测模块，用于根据所述统计模块统计的所述属性信息将所述分类器集合作为预测模型预测所述故障产品的缺陷得到分类标签集合；

第三获取模块，用于根据所述分类器集合和所述分类器集合中每个树分类器的权重，获取所述分类标签集合中每个分类标签的信任值。

结合第二方面和上述可能的实现方式，在另一种可能的实现方式中，所述预设策略包括决策树算法。

本发明实施例提供的一种缺陷预测方法及装置，根据目标属性从预存的产品故障记录中选择训练属性集，并根据目标属性和训练属性集组合成训练集生成包含至少 2 个树分类器的分类器集合，此

时当产品出现故障时，便可以将该分类器集合作为预测模型来预测故障产品的缺陷，利用该分类器集合作为预测模型，解决了采用单一决策树容易引起过拟合或欠拟合而导致无法对故障产品进行缺陷预测的问题，并且在实现了对故障产品的缺陷快速定位的同时也提高了对故障产品缺陷预测的准确率。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术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性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本发明实施例 1 提供的一种缺陷预测方法流程图；

图 2 为本发明实施例 2 提供的一种缺陷预测方法流程图；

图 3 为本发明实施例 3 提供的一种缺陷预测装置组成示意图；

图 4 为本发明实施例 3 提供的另一种缺陷预测装置组成示意图；

图 5 为本发明实施例 4 提供的一种缺陷预测装置组成示意图。

具体实施方式

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作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

实施例 1

本发明实施例提供一种缺陷预测方法，如图 1 所示，该方法可以包括：

101、根据目标属性从预存的产品故障记录中选择训练属性集，并将目标属性和训练属性集组合成训练集。

其中，当一个产品出现故障时，故障检测人员一般情况下都希望能够快速的定位出故障产品的缺陷类型或者导致产品出现故障

的器件，以便来节省维修人员的维修的时间，而要实现对故障产品的缺陷类型或者是导致产品出现故障的器件进行快速的定位，可以通过提前训练预测模型来实现，首先故障检测人员可以将生产环节或者使用过程中出现过故障的产品的信息进行收集并将这些信息记录到产品故障记录中，这样当训练预测模型的时候，便可以根据历史故障产品的缺陷属性从提前记录的出现过故障的产品的产品故障记录中选择建立预测模型所必须的属性作为训练属性集，其中，将历史故障产品的缺陷属性定义为目标属性，当根据目标属性选择好训练属性集之后，将目标属性和训练属性集组合生成训练集，具体的，训练集中可以包含多个训练单元，其中每个训练单元中包含一个目标属性和一个训练属性集。

102、根据训练集生成分类器集合；其中，分类器集合包含至少2个树分类器。

其中，当根据目标属性选择好需要的训练属性集，并将目标属性和训练属性集组合成训练集之后，便可以根据训练集生成分类器集合，具体的，该分类器集合中包含至少2个树分类器，每个树分类器根据预设的策略生成，并将生成的所有的树分类器共同组成分类器集合。该预设的策略可以是决策树算法等。

103、将分类器集合作为预测模型预测故障产品的缺陷。

其中，在生产或者使用过程中，若某个产品出现了故障，便可以根据生成的包含至少一个树分类器的分类器集合快速并准确的定位出该故障产品的缺陷。

本发明实施例提供的一种缺陷预测方法，根据目标属性从预存的产品故障记录中选择训练属性集，并根据目标属性和训练属性集组合成训练集生成包含至少2个树分类器的分类器集合，此时当产品出现故障时，便可以将该分类器集合作为预测模型来预测故障产品的缺陷，利用该分类器集合作为预测模型，解决了采用单一决策树容易引起过拟合或欠拟合而导致无法对故障产品进行缺陷预测的问题，并且在实现了对故障产品的缺陷快速定位的同时也提高了

对故障产品缺陷预测的准确率。

实施例 2

本发明实施例提供一种缺陷预测方法，如图 2 所示，该方法可以包括：

201、根据目标属性从预存的产品故障记录中选择训练属性集，并将目标属性和训练属性集组合成训练集。

具体的，当一个产品在生产过程中或者使用过程中出现故障时，一般情况下故障检测人员都希望可以快速的定位故障产品的缺陷类型或者出现故障的器件，而对于任何一种产品来说，故障或缺陷的出现都与该产品的客观信息有一定的关联性，例如，产品的型号、使用环境、原料来源等等。为了实现在产品出现故障或者缺陷时，能够快速的定位故障产品的缺陷类型或者出现故障的器件，可以从生产环节或者使用过程中出现过故障的产品的产品故障记录中选择出建立预测模型需要的属性，并将选择出来的属性组成训练集，利用该训练集来建立预测模型。

其中，首先要做的就是收集生产环节或者使用过程中的出现过故障的产品的属性信息，并将每个故障产品的属性信息记录下来。属性信息具体的可以分为以下几类：描述产品特征的属性、描述使用环境的属性、描述生产环节的属性以及缺陷属性。其中，描述产品特性的属性可以是产品名称、产品型号、组成部件等；描述使用环境的属性可以是使用周期、使用地点、使用气候等；描述生产环节的属性可以是生产日期、加工部门、检测记录等；缺陷属性则可以是缺陷类型、缺陷现象、缺陷根因、缺陷器件等。

需要说明的是，本发明实施例对记录的故障产品的属性信息的分类以及每种分类下记录的属性信息的种类不作限制，对记录故障产品的属性信息的形式也不作限制。

其次，由于对于故障产品来说，记录的属性信息有很多，而有些属性不是建立预测模型所必须要使用的属性，也就是说某些属性对判断故障产品的缺陷的作用不大，因此接下来要做的就是对故障

产品的属性信息进行筛选。可以理解的是，历史故障记录中记录的故障产品的属性信息中的缺陷属性也极有可能是将来出现故障的产品故障，即是将来出现故障的产品需要进行预测的属性，因此为了方便本领域技术人员的理解，我们将历史故障产品的缺陷属性称为目标属性，将根据历史故障产品的缺陷属性挑选出与其关联性较大的属性称为训练属性集，我们可以将目标属性和训练属性集组成训练集，这样便可以利用训练集来建立预测模型。筛选过程具体的可以是：针对目标属性，对记录的属性信息进行筛选，可以选出 X 个属性形成训练属性集，其中 X 可以是记录的属性信息中的全部属性，也可以是 1 个属性。例如，历史故障产品的缺陷属性为缺陷类型，即可以定义目标属性 Y={缺陷类型}，记录的故障产品的属性信息包括：产品名称、产品型号、组成部件、使用周期、使用地点、使用气候、生产日期、加工部门、检测记录、缺陷类型、缺陷现象、缺陷根因、缺陷器件，那么我们可以利用预设的规则在记录的故障产品的历史故障记录中的属性信息中选择建立预测模型所需要的属性来组成训练属性集，假设我们选出来的属性为：产品名称、生产日期、加工部门、使用周期，即可以定义训练属性集 X={产品名称、生产日期、加工部门、使用周期}，这样即可以定义训练集 T={产品名称、生产日期、加工部门、使用周期、缺陷类型}，当选出目标属性和训练属性集之后，便可以根据目标属性和训练属性集从历史故障记录中选取多个故障产品相对应的属性来生成训练集，该训练属性集中包含 M 个训练单元，每个训练单元包含一个历史故障产品的目标属性和训练属性集。其中，对于训练属性集中属性的选择有 2 个要求：一是利用训练属性集建立的预测目标属性的预测模型的准确率要高，这点要求可以通过重复的针对该目标属性选择不同的训练属性集组成训练集，并验证由不同生成的训练集建立的预测模型的准确性，从中选择准确性最高的作为建立预测模型所需的训练集，并可以将已知的缺陷的故障产品的目标属性去掉，将该故障产品在生产和制造过程中的属性信息作为测试数据，来检测生成的

树分类器准确性；二是训练属性集里的属性在故障产品被检测前是可获得的，例如，在上述记录的故障产品的属性信息中缺陷器件不能作为训练属性集中的属性，因为在故障检测前，并不能获知该故障产品是那个器件出现了故障。

需要说明的是，训练属性集的具体选择规则可以是遍历的方法，也可以是通过计算和目标属性的相关性来选出相关性最大的前 X 个属性作为训练属性集。计算和目标属性的相关性的选择方法是较为常用的方法，其中计算相关性的算法也有很多，一种最简单的相关性的计算方法是计算各属性和目标属性同时出现的频率，同时出现的频率越高，相关性便越大。在本发明实施例中，对训练属性集的选择方法及选择某些方法时需要运用的算法不作限制。

202、根据训练集生成分类器集合；其中，分类器集合包含至少 2 个树分类器。

其中，在根据目标属性从预存的产品故障记录中选择训练属性集，并组合成训练集之后，便可以根据训练集生成分类器集合。可以理解的是，目标属性和训练属性集组成的训练集可以包含 M 个训练单元，其中每个训练单元包含一个目标属性和一个训练属性集，即训练集 $T=\{(X_r,Y_r)，\ r=1,2,\cdots M\}$，其中 (X_1,Y_1) 即为第一训练单元。

根据训练集 $T=\{(X_r,Y_r)，\ r=1,2,\cdots M\}$ 生成一个分类器集合 $C=\{C_j，\ j=1,2,\cdots N\}$ 具体的可以是分为以下步骤，202a、202b 及 202c:

202a、从训练集中选取第 N 训练子集；其中 N 为大于等于 2 的整数。

其中，从训练集 $T=\{(X_r,Y_r)，\ r=1,2,\cdots M\}$ 中选取第 N 训练子集，该第 N 训练子集包含 M'个训练单元，M'小于等于 M，选取方法可以为可放回的随机抽样，本发明实施例在此不作限制。例如，可以从训练集中选取第一训练子集，第二训练子集...第 N 训练子集。

202b、根据预设策略生成与该第 N 训练子集相对应的第 N 树分类器。

其中，在从训练集中选取到第 N 训练子集之后，可以根据预设

的策略生成与该第 N 训练子集相对应的第 N 树分类器。该预设策略可以是生成树算法，具体的可以理解的是：将从训练集中选择的第 N 训练子集作为根节点，并按照分离算法选择分离属性和分离谓词，将根节点按照分离属性和分离谓词进行分裂，得到两个分支，对于每一个分支中的属性可以利用属性选择策略进行选择，然后对分支继续进行按照分离算法进行分裂，重复上述步骤直到得到最终生成的分支可以确定目标属性，最后再根据树裁剪策略对生成的树分类器进行检测。例如训练集 T={产品名称、生产日期、加工部门、使用周期、缺陷类型}，其中包含 M 个训练单元，第 N 训练子集为包含 M'个训练单元的集合并将该第 N 训练子集作为根节点，假设根据分离算法选择分离属性为使用周期、分离谓词为使用周期大于 50 天和使用周期小于等于 50 天，这样便可以根据分离属性和分离谓词将根节点分为 2 个分支，可以再继续选择分离属性和分离谓词进行分裂，直到可以确定目标属性。

其中，上述树分类器生成过程中使用的分离算法包括但不限于信息熵检验、基尼索引检验、开方检验、增益率检验；属性选择可以包括随机单个属性选择和随机多个属性选择，属性选择策略本发明实施例不作限制；树裁剪策略包括但不限于预裁剪策略、后裁剪策略。

202c、重复以上步骤 202a、202b，生成 N 个树分类器，并将 N 个树分类器组合生成分类器集合。

其中，本发明实施例中的生成的树分类器的个数 N 可以是预先设置的门限值，即当生成的树分类器的个数达到预定的门限值时，便可以将生成的 N 个树分类器组成生成分类器集合，例如当预设的门限值 N 为 5 时，分类器集合 $C=\{C_1,C_2,C_3,C_4,C_5\}$。何时生成分类器集合也可以是通过计算生成的 K 个树分类器的错误率和生成的 K-1 个树分类器的错误率的差值来决定，具体的，当生成第 K-1 树分类器时，可以计算生成的 K-1 个树分类器的错误率，并且当生成第 K 树分类器时，计算生成的 K 个树分类器的错误率，这样当计算得到 K

个树分类器的错误率和K-1个树分类器的错误率的差值小于预设的阈值时，便将生成的K个树分类器组合生成分类器集合，其中，K为小于等于N的整数。

当生成第K树分类器时，生成的K个树分类器的错误率的计算方法为：对于训练集中的每一个训练单元，计算其预测标签，并根据该预测标签得到生成的K个树分类器的错误率。具体的，根据第一训练单元从分类器集合中选取第一类树分类器，并根据第一类树分类器生成第一训练单元的第一预测标签；根据第二训练单元从分类器集合中选取第二类树分类器，并根据第二类树分类器生成第二训练单元的第二预测标签，...根据第M训练单元从分类器集合中选取第M类树分类器，并根据第M类树分类器生成第M训练单元的第M预测标签；重复上述步骤，直到针对训练集中的每一个训练单元都对应计算出来该训练单元对应的预测标签再结束，最后根据计算出来的M个预测标签得到生成的K个树分类器的错误率。其中，第M类树分类器为未使用第M训练单元生成树分类器的分类器集合。

预测标签具体计算过程为，假设对于训练集中的第r训练单元（其中r为大于0，并小于等于M的正整数）来说，分类器集合中的树分类器可以分为两类，一类为使用第r训练单元生成的树分类器，另一类为未使用第r训练单元生成的树分类器，我们将未使用第r训练单元生成的树分类器组成一个集合，并称为第r类树分类器，记作C_r^{OOB}，那么第r训练单元的第r预测标签的具体计算公式为：

$$C^{OOB}(r,\ x_r)=\arg\max_{y}\sum_{C_j\in C_r^{OOB}}h(\varepsilon_j)I(C_j(x_r)=y)$$

其中，$C^{OOB}(r,\ x_r)$为第r训练单元的第r预测标签，C_j为第j树分类器，C_r^{OOB}为第r类树分类器，$h(\varepsilon_j)$为第j树分类器的权重，$C_j(x_r)$为根据第j树分类器和第r训练单元中包含的训练属性集得到的目标属性，y为分类标签，$y\in Y$，Y为根据第r训练单元和分类器集合得到的分类标签集合，$I(x)$是指标函数：$I(true)=1$，$I(false)=0$。

生成的K个树分类器的错误率的具体计算公式为：

$$E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$$

其中，$E(T)$为生成的K个树分类器的错误率，M为训练集中训练单元的个数，$C^{OOB}(r,\ x_r)$为所述第r训练单元的第r预测标签，y_r为第r训练单元的目标属性，$I(x)$是指标函数：$I(true)=1$，$I(false)=0$。

第j树分类器的权重的具体计算过程为：从训练集中选取第j'训练子集，然后根据第j'训练子集获取第j树分类器的误预测率，最后根据第j树分类器误预测率获取第j树分类器的权重。其中，所述第j'训练子集与所述第j训练子集的交集为空，所述第j'训练子集包含至少一个训练单元。具体的：将第j'训练子集记录为$T''=\{(x_r^*,\ y_r^*),\ r=1,2,\cdots N\}$，其中$T''\cap T'=\phi$，$T'$为生成第j树分类器的第j训练子集，第j树分类器的误预测率的具体计算公式为：

$$\varepsilon_j=\frac{1}{N}\sum_{r=1}^{N}I(C_j(x_r^*)\neq y_r^*)$$

其中，ε_j为第j树分类器的误预测率，N为第N'训练子集中训练单元的个数，$I(x)$是指标函数：$I(true)=1$，$I(false)=0$，$C_j(x_r^*)$为根据第j树分类器和第r训练单元中包含的训练属性集得到的目标属性，y_r^*为第r训练单元包含的目标属性。

第j树分类器的权重由公式$h(\varepsilon_j)$得到，其中，$h(x)=1-x$或

$h(x)=\log(\frac{1}{x})$。

203、统计故障产品的属性信息。

其中，当需要预测故障产品的缺陷时，可以先统计故障产品的属性信息，该属性信息是故障产品的在生产及使用过程中获得的数据，可以包括：产品名称、产品型号、组成部件、使用周期、使用地点、生产日期、加工部门等。

204、根据属性信息将分类器集合作为预测模型预测故障产品的缺陷得到分类标签集合。

其中，当将故障产品的属性信息统计出来之后，可以利用统计出来的该故障产品的属性信息，将提前训练好的分类器集合作为预测模型，预测故障产品的缺陷，由于生成的分类器集合中包含N个树分类器，因为采用该分类器集合预测出来的故障产品的缺陷将会出现多个预测结果，将预测出来的多个结果作为分类标签集合。采用本发明实施例提供的缺陷预测方法，不仅可以预测出故障产品的缺陷，还可以得到多个预测结果供维修人员参考，当维修人员根据预测出来的第一个预测结果检测故障产品时，发现第一个预测结果不是故障产品的缺陷时，便可以从分类标签集合中选择其他的预测结果来对故障产品进行检测，直到找到故障产品真正的缺陷，这样便可以节约维修人员的时间。

205、根据分类器集合和分类器集合中树分类器的权重，获取分类标签集合中每个分类标签的信任值。

其中，当根据统计出的故障产品的属性信息得到分类标签集合之后，为了让维修人员能够更快的定位出故障产品的缺陷，还可以根据分类器集合和分类器集合中树分类器的权重，计算分类标签集合中每个分类标签的信任值。分类标签的信任值的具体计算方法为：

$$U_{\mathrm{r}}(y)=\frac{1}{Z}\sum_{j=1}^{|C|}h(\varepsilon_j)I(C_j(x_r)=y)$$

其中，Y为分类标签集合，$y\in Y$；$U_{\mathrm{r}}(y)$为分类标签y的信任值；Z为归一化因子，$Z=\sum_{j=1}^{|C|}h(\varepsilon_j)$；$h(\varepsilon_j)$为第j树分类器的权重；$I(x)$是指标函数：$I(true)=1$，$I(false)=0$；$C_j(x_r)$为根据第j树分类器预测的故障产品的目标属性。

若通过公式计算出$U_{\mathrm{r}}(y)=0$，则表明该属性信息没有用于y的分类，此外，r可能的缺陷分类标签定义为$\{y\in Y|U_r(y)>\sigma\}$。

本发明实施例提供一种缺陷预测方法，根据目标属性从预存的产品故障记录中选择训练属性集，并根据目标属性和训练属性集组合成训练集生成包含至少2个树分类器的分类器集合，此时当产品

出现故障时，便可以将该分类器集合作为预测模型来预测故障产品的缺陷，利用该分类器集合作为预测模型，解决了采用单一决策树容易引起过拟合或欠拟合而导致无法对故障产品进行缺陷预测的问题，并且在实现了对故障产品的缺陷快速定位的同时也提高了对故障产品缺陷预测的准确率。

并且，当将分类器集合作为预测模型预测故障产品的缺陷时，还可以得到多个预测结果，并可以计算出每个预测结果的信任值，节约了维修人员定位缺陷的时间。

实施例 3

本发明实施例提供一种缺陷预测装置，如图 3 所示，包括：处理单元 31、生成单元 32、预测单元 33。

处理单元 31，用于根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集；其中，所述目标属性为历史故障产品的缺陷属性。

生成单元 32，用于根据所述处理单元 31 得到的训练集生成分类器集合；其中，所述分类器集合包含至少 2 个树分类器。

预测单元 33，用于将所述生成单元 32 生成的分类器集合作为预测模型预测故障产品的缺陷。

进一步的，所述训练集包含 M 个训练单元，每个训练单元包含一个目标属性和一个训练属性集。

进一步的，如图 4 所示，所述生成单元 32 可以包括：选取模块 321、生成模块 322、组合模块 323。

选取模块 321，用于从所述处理单元 31 得到的所述训练集中选取第一训练子集。

生成模块 322，用于根据预设策略生成与所述选取模块 321 选取的所述第一训练子集相对应的第一树分类器。

所述选取模块 321，还用于从所述处理单元 31 得到的所述训练集中选取第二训练子集。

所述生成模块 322，还用于根据预设策略生成与所述选取模块

321 选取的所述第二训练子集相对应的第二树分类器。

所述选取模块 321，还用于从所述处理单元 31 得到的所述训练集中选取第 N 训练子集；其中，所述第 N 训练子集包含 M'个训练单元，所述 M'小于等于所述 M。

所述生成模块 322，还用于根据预设策略生成与所述选取模块 321 选取的所述第 N 训练子集相对应的第 N 树分类器；其中，所述 N 为大于等于 2 的整数。

组合模块 323，用于将所述生成模块 322 生成的 N 个树分类器组合生成所述分类器集合。

进一步的，所述生成单元 32 还可以包括：第一获取模块 324、第二获取模块 325。

第一获取模块 324，用于当生成第 K-1 树分类器时，获取生成的 K-1 个树分类器的错误率。

第二获取模块 325，用于当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率；以便当所述 K 个树分类器的错误率和所述 K-1 个树分类器的错误率的差值小于预设的阈值时，将所述 K 个树分类器组合生成所述分类器集合；其中，所述 K 为小于等于 N 的整数。

进一步的，所述第二获取模块 325 可以包括：选取子模块 3251、生成子模块 3252、获取子模块 3253。

选取子模块 3251，用于根据第一训练单元从所述分类器集合中选取第一类树分类器。

生成子模块 3252，用于根据所述选取子模块 3251 选取的所述第一类树分类器生成所述第一训练单元的第一预测标签。

所述选取子模块 3251，还用于根据第二训练单元从所述分类器集合中选取第二类树分类器。

所述生成子模块 3252，还用于根据所述选取子模块 3251 选取的所述第二类树分类器生成所述第二训练单元的第二预测标签。

所述选取子模块 3251，还用于根据第 M 训练单元从所述分类

器集合中选取第 M 类树分类器；其中，所述第 M 类树分类器为未使用第 M 训练单元生成树分类器的分类器集合,所述 M 为训练集中包含训练单元的个数。

所述生成子模块 3252，还用于根据所述选取子模块 3251 选取的所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签。

获取子模块 3253，用于根据所述生成子模块 3252 生成的 M 个预测标签获取所述生成的 K 个树分类器的错误率。

进一步的，所述生成子模块 3252 具体用于：根据

$C^{OOB}(M,\ x_M)=\arg\max_{y}\sum_{C_j\in C_M^{OOB}}h(\varepsilon_j)I(C_j(x_M)=y)$ 生成所述第 M 预测标签；其中，

$C^{OOB}(M,\ x_M)$ 为所述第 M 训练单元的第 M 预测标签，C_j 为第 j 树分类器，C_M^{OOB} 为所述第 M 类树分类器，$h(\varepsilon_j)$ 为第 j 树分类器的权重，$C_j(x_M)$ 为根据所述第 j 树分类器和所述第 M 训练单元中包含的训练属性集得到的目标属性，$y\in Y$，Y 为分类标签集合。

进一步的，所述获取子模块 3253 具体用于：根据

$E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$ 获取所述生成子模块 3252 生成的 K 个树分类器的错误率；其中，$E(T)$ 为所述生成的 K 个树分类器的错误率，M 为所述训练集中训练单元的个数，$C^{OOB}(r,\ x_r)$ 为所述第 r 训练单元的第 r 预测标签，y_r 为第 r 训练单元的目标属性。

进一步的，该装置还可以包括：选取单元 34、第一获取单元 35、第二获取单元 36。

选取单元 34，用于在所述生成模块 322 根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器之后，从所述训练集中选取第 N'训练子集；其中，所述第 N'训练子集与所述第 N 训练子集的交集为空，所述第 N'训练子集包含至少一个训练单元。

第一获取单元 35，用于根据所述选取单元 34 选取的所述第 N'训练子集获取所述第 N 树分类器的误预测率。

第二获取单元 36，用于根据所述第一获取单元 35 获取到的所

述第 N 树分类器误预测率获取所述第 N 树分类器的权重。

进一步的，所述预测单元 33 可以包括：统计模块 331、预测模块 332、第三获取模块 333。

统计模块 331，用于统计所述故障产品的属性信息。

预测模块 332，用于根据所述统计模块 331 统计的所述属性信息将所述分类器集合作为预测模型预测所述故障产品的缺陷得到分类标签集合。

第三获取模块 333，用于根据所述分类器集合和所述分类器集合中每个树分类器的权重，获取所述分类标签集合中每个分类标签的信任值。

本发明实施例提供一种缺陷预测装置，根据目标属性从预存的产品故障记录中选择训练属性集，并根据目标属性和训练属性集组合成训练集生成包含至少 2 个树分类器的分类器集合，此时当产品出现故障时，便可以将该分类器集合作为预测模型来预测故障产品的缺陷，利用该分类器集合作为预测模型，解决了采用单一决策树容易引起过拟合或欠拟合而导致无法对故障产品进行缺陷预测的问题，并且在实现了对故障产品的缺陷快速定位的同时也提高了对故障产品缺陷预测的准确率。

并且，当将分类器集合作为预测模型预测故障产品的缺陷时，还可以得到多个预测结果，并可以计算出每个预测结果的信任值，节约了维修人员定位缺陷的时间。

实施例 4

本发明实施例提供一种缺陷预测装置，如图 5 所示，包括：至少一个处理器 41、存储器 42、通信接口 43 和总线 44，该至少一个处理器 41、存储器 42 和通信接口 43 通过总线 44 连接并完成相互间的通信，其中：

所述总线 44 可以是工业标准体系结构（Industry Standard Architecture，ISA）总线、外部设备互连（Peripheral Component Interconnect，PCI）总线或扩展工业标准体系结构（Extended Industry

Standard Architecture，EISA）总线等。所述总线 44 可以分为地址总线、数据总线、控制总线等。为便于表示，图 5 中仅用一条粗线表示，但并不表示仅有一根总线或一种类型的总线。

所述存储器 42 用于存储可执行程序代码，该程序代码包括计算机操作指令。存储器 42 可能包含高速 RAM 存储器，也可能还包括非易失性存储器（non-volatile memory），例如至少一个磁盘存储器。

所述处理器 41 可能是一个中央处理器（Central Processing Unit，CPU），或者是特定集成电路（Application Specific Integrated Circuit，ASIC），或者是被配置成实施本发明实施例的一个或多个集成电路。

所述通信接口 43，主要用于实现本实施例的设备之间的通信。

所述处理器 41 执行所述程序代码，用于根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集；其中，所述目标属性为历史故障产品的缺陷属性，根据所述训练集生成分类器集合；其中，所述分类器集合包含至少 2 个树分类器，并将生成的分类器集合作为预测模型预测故障产品的缺陷。

进一步的，所述训练集包含 M 个训练单元，每个训练单元包含一个目标属性和一个训练属性集。所述处理器 41，还用于从所述训练集中选取第一训练子集，根据预设策略生成与所述第一训练子集相对应的第一树分类器；从所述训练集中选取第二训练子集，根据预设策略生成与所述第二训练子集相对应的第二树分类器；从所述训练集中选取第 N 训练子集，根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器，最后将生成的 N 个树分类器组合生成所述分类器集合。其中，所述第 N 训练子集包含 M'个训练单元，所述 M'小于等于所述 M，所述 N 为大于等于 2 的整数。

进一步的，所述处理器 41，还用于当生成第 K-1 树分类器时，获取生成的 K-1 个树分类器的错误率，并且当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率，以便当所述 K 个树分类器的错

误率和所述 K-1 个树分类器的错误率的差值小于预设的阈值时，将所述 K 个树分类器组合生成所述分类器集合；其中，所述 K 为小于等于 N 的整数。

进一步的，所述处理器 41，还用于根据第一训练单元从所述分类器集合中选取第一类树分类器，根据所述第一类树分类器生成所述第一训练单元的第一预测标签；根据第二训练单元从所述分类器集合中选取第二类树分类器，根据所述第二类树分类器生成所述第二训练单元的第二预测标签；根据第 M 训练单元从所述分类器集合中选取第 M 类树分类器；根据所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签，最后根据生成的 M 个预测标签获取所述生成的 K 个树分类器的错误率。其中，所述第 M 类树分类器为未使用第 M 训练单元生成树分类器的分类器集合，所述 M 为训练集中包含训练单元的个数。

进一步的，所述处理器 41 还用于：根据

$C^{OOB}(M,\ x_M)=\arg\max_{y}\sum_{C_j\in C_M^{OOB}}h(\varepsilon_j)I(C_j(x_M)=y)$ 生成所述第 M 预测标签；其中，

$C^{OOB}(M,\ x_M)$ 为所述第 M 训练单元的第 M 预测标签，C_j 为第 j 树分类器，C_M^{OOB} 为所述第 M 类树分类器，$h(\varepsilon_j)$ 为第 j 树分类器的权重，$C_j(x_M)$ 为根据所述第 j 树分类器和所述第 M 训练单元中包含的训练属性集得到的目标属性，$y\in Y$，Y 为分类标签集合。并根据

$E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$ 获取生成的 K 个树分类器的错误率；其中，

$E(T)$ 为所述生成的 K 个树分类器的错误率，M 为所述训练集中训练单元的个数，$C^{OOB}(r,\ x_r)$ 为所述第 r 训练单元的第 r 预测标签，y_r 为第 r 训练单元的目标属性。

进一步的，所述处理器 41，还用于在所述根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器之后，从所述训练集中选取第 N'训练子集，根据所述第 N'训练子集获取所述第 N 树分类器的误预测率，根据所述第 N 树分类器误预测率获取所述第 N 树分类

器的权重。其中，所述第 N'训练子集与所述第 N 训练子集的交集为空，所述第 N'训练子集包含至少一个训练单元。

进一步的，所述处理器 41，还用于统计所述故障产品的属性信息，根据所述属性信息将所述分类器集合作为预测模型预测所述故障产品的缺陷得到分类标签集合，并根据所述分类器集合和所述分类器集合中树分类器的权重，获取所述分类标签集合中每个分类标签的信任值。

进一步的，所述预设策略包括决策树算法。

本发明实施例提供一种缺陷预测装置，根据目标属性从预存的产品故障记录中选择训练属性集，并根据目标属性和训练属性集组合成训练集生成包含至少 2 个树分类器的分类器集合，此时当产品出现故障时，便可以将该分类器集合作为预测模型来预测故障产品的缺陷，利用该分类器集合作为预测模型，解决了采用单一决策树容易引起过拟合或欠拟合而导致无法对故障产品进行缺陷预测的问题，并且在实现了对故障产品的缺陷快速定位的同时也提高了对故障产品缺陷预测的准确率。

并且，当将分类器集合作为预测模型预测故障产品的缺陷时，还可以得到多个预测结果，并可以计算出每个预测结果的信任值，节约了维修人员定位缺陷的时间。

通过以上的实施方式的描述，所属领域的技术人员可以清楚地了解到本发明可借助软件加必需的通用硬件的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在可读取的存储介质中，如计算机的软盘，硬盘或光盘等，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本发明各个实施例所述的方法。

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以所述权利要求的保护范围为准。

权 利 要 求 书

1、一种缺陷预测方法，其特征在于，包括：

根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集；其中，所述目标属性为历史故障产品的缺陷属性；

根据所述训练集生成分类器集合；其中，所述分类器集合包含至少 2 个树分类器；

将所述分类器集合作为预测模型预测故障产品的缺陷。

2、根据权利要求 1 所述的缺陷预测方法，其特征在于，所述训练集包含 M 个训练单元，每个训练单元包含一个目标属性和一个训练属性集；

所述根据所述训练集生成分类器集合，包括：

从所述训练集中选取第一训练子集；

根据预设策略生成与所述第一训练子集相对应的第一树分类器；

从所述训练集中选取第二训练子集；

根据预设策略生成与所述第二训练子集相对应的第二树分类器；

从所述训练集中选取第 N 训练子集；其中，所述第 N 训练子集包含 M'个训练单元，所述 M'小于等于所述 M；

根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器；其中，所述 N 为大于等于 2 的整数；

将 N 个树分类器组合生成所述分类器集合。

3、根据权利要求 1 所述的缺陷预测方法，其特征在于，还包括：

当生成第 K-1 树分类器时，获取生成的 K-1 个树分类器的错误率；

当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率；以便当所述 K 个树分类器的错误率和所述 K-1 个树分类器的错误率的差值小于预设的阈值时，将所述 K 个树分类器组合生成所述分类器集合；其中，所述 K 为小于等于 N 的整数。

4、根据权利要求 3 所述的缺陷预测方法，其特征在于，所述当

生成第 K 树分类器时，获取生成的 K 个树分类器的错误率，包括：

根据第一训练单元从所述分类器集合中选取第一类树分类器；

根据所述第一类树分类器生成所述第一训练单元的第一预测标签；

根据第二训练单元从所述分类器集合中选取第二类树分类器；

根据所述第二类树分类器生成所述第二训练单元的第二预测标签；

根据第 M 训练单元从所述分类器集合中选取第 M 类树分类器；其中，所述第 M 类树分类器为未使用第 M 训练单元生成树分类器的分类器集合,所述 M 为训练集中包含训练单元的个数；

根据所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签；

根据 M 个预测标签获取所述生成的 K 个树分类器的错误率。

5、根据权利要求 4 所述的缺陷预测方法，其特征在于，所述根据所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签，具体包括：

根据 $C^{OOB}(M,\ x_M)=\arg\max_{y}\sum_{C_j\in C_M^{OOB}}h(\varepsilon_j)I(C_j(x_M)=y)$ 生成所述第 M 预测标签；其中，$C^{OOB}(M,\ x_M)$ 为所述第 M 训练单元的第 M 预测标签，C_j 为第 j 树分类器，C_M^{OOB} 为所述第 M 类树分类器，$h(\varepsilon_j)$ 为第 j 树分类器的权重，$C_j(x_M)$ 为根据所述第 j 树分类器和所述第 M 训练单元中包含的训练属性集得到的目标属性，$y\in Y$，Y 为分类标签集合。

6、根据权利要求 5 所述的缺陷预测方法，其特征在于，所述根据 M 个预测标签获取所述生成的 K 个树分类器的错误率,具体包括：

根据 $E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$ 获取所述生成的 K 个树分类器的错误率；其中，$E(T)$ 为所述生成的 K 个树分类器的错误率,M 为所述训练集中训练单元的个数，$C^{OOB}(r,\ x_r)$ 为所述第 r 训练单元的第 r 预测标签，y_r 为第 r 训练单元的目标属性。

7、根据权利要求 2 所述的缺陷预测方法，其特征在于，在所述根据预设策略生成与所述第N训练子集相对应的第N树分类器之后，还包括：

从所述训练集中选取第 N'训练子集；其中，所述第 N'训练子集与所述第 N 训练子集的交集为空，所述第 N'训练子集包含至少一个训练单元；

根据所述第 N'训练子集获取所述第 N 树分类器的误预测率；

根据所述第 N 树分类器误预测率获取所述第 N 树分类器的权重。

8、根据权利要求 7 所述的缺陷预测方法，其特征在于，所述将所述分类器集合作为预测模型预测故障产品的缺陷，包括：

统计所述故障产品的属性信息；

根据所述属性信息将所述分类器集合作为预测模型预测所述故障产品的缺陷得到分类标签集合；

根据所述分类器集合和所述分类器集合中每个树分类器的权重，获取所述分类标签集合中每个分类标签的信任值。

9、根据权利要求 2-8 中任一权利要求所述的缺陷预测方法，其特征在于，所述预设策略包括决策树算法。

10、一种缺陷预测装置，其特征在于，包括：

处理单元，用于根据目标属性从预存的产品故障记录中选择训练属性集，并将所述目标属性和所述训练属性集组合成训练集；其中，所述目标属性为历史故障产品的缺陷属性；

生成单元，用于根据所述处理单元得到的训练集生成分类器集合；其中，所述分类器集合包含至少 2 个树分类器；

预测单元，用于将所述生成单元生成的分类器集合作为预测模型预测故障产品的缺陷。

11、根据权利要求 10 所述的缺陷预测装置，其特征在于，所述训练集包含 M 个训练单元，每个训练单元包含一个目标属性和一个训练属性集；

所述生成单元，包括：

选取模块，用于从所述处理单元得到的所述训练集中选取第一训练子集；

生成模块，用于根据预设策略生成与所述选取模块选取的所述第一训练子集相对应的第一树分类器；

所述选取模块，还用于从所述处理单元得到的所述训练集中选取第二训练子集；

所述生成模块，还用于根据预设策略生成与所述选取模块选取的所述第二训练子集相对应的第二树分类器；

所述选取模块，还用于从所述处理单元得到的所述训练集中选取第 N 训练子集；其中，所述第 N 训练子集包含 M'个训练单元，所述 M'小于等于所述 M；

所述生成模块，还用于根据预设策略生成与所述选取模块选取的所述第 N 训练子集相对应的第 N 树分类器；其中，所述 N 为大于等于 2 的整数；

组合模块，用于将所述生成模块生成的 N 个树分类器组合生成所述分类器集合。

12、根据权利要求 10 所述的缺陷预测装置，其特征在于，所述生成单元还包括：

第一获取模块，用于当生成第 K-1 树分类器时，获取生成的 K-1 个树分类器的错误率；

第二获取模块，用于当生成第 K 树分类器时，获取生成的 K 个树分类器的错误率；以便当所述 K 个树分类器的错误率和所述 K-1 个树分类器的错误率的差值小于预设的阈值时，将所述 K 个树分类器组合生成所述分类器集合；其中，所述 K 为小于等于 N 的整数。

13、根据权利要求 12 所述的缺陷预测装置，其特征在于，所述第二获取模块，包括：

选取子模块，用于根据第一训练单元从所述分类器集合中选取第一类树分类器；

生成子模块，用于根据所述选取子模块选取的所述第一类树分类

器生成所述第一训练单元的第一预测标签;

所述选取子模块,还用于根据第二训练单元从所述分类器集合中选取第二类树分类器;

所述生成子模块,还用于根据所述选取子模块选取的所述第二类树分类器生成所述第二训练单元的第二预测标签;

所述选取子模块,还用于根据第 M 训练单元从所述分类器集合中选取第 M 类树分类器;其中,所述第 M 类树分类器为未使用第 M 训练单元生成树分类器的分类器集合,所述 M 为训练集中包含训练单元的个数;

所述生成子模块,还用于根据所述选取子模块选取的所述第 M 类树分类器生成所述第 M 训练单元的第 M 预测标签;

获取子模块,用于根据所述生成子模块生成的 M 个预测标签获取所述生成的 K 个树分类器的错误率。

14、根据权利要求 13 所述的缺陷预测装置,其特征在于,所述生成子模块,具体用于:

根据 $C^{OOB}(M,\ x_M)=\arg\max_{y}\sum_{C_j\in C_M^{OOB}}h(\varepsilon_j)I(C_j(x_M)=y)$ 生成所述第 M 预测标签;其中,$C^{OOB}(M,\ x_M)$ 为所述第 M 训练单元的第 M 预测标签,C_j 为第 j 树分类器,C_M^{OOB} 为所述第 M 类树分类器,$h(\varepsilon_j)$ 为第 j 树分类器的权重,$C_j(x_M)$ 为根据所述第 j 树分类器和所述第 M 训练单元中包含的训练属性集得到的目标属性,$y\in Y$,Y 为分类标签集合。

15、根据权利要求 14 所述的缺陷预测装置,其特征在于,所述获取子模块,具体用于:

根据 $E(T)=\frac{1}{M}\sum_{r=1}^{M}I(C^{OOB}(r,\ x_r)=y_r)$ 获取所述生成子模块生成的 K 个树分类器的错误率;其中,$E(T)$ 为所述生成的 K 个树分类器的错误率,M 为所述训练集中训练单元的个数,$C^{OOB}(r,\ x_r)$ 为所述第 r 训练单元的第 r 预测标签,y_r 为第 r 训练单元的目标属性。

16、根据权利要求 11 所述的缺陷预测装置,其特征在于,还包

括：

选取单元，用于在所述生成模块根据预设策略生成与所述第 N 训练子集相对应的第 N 树分类器之后，从所述训练集中选取第 N'训练子集；其中，所述第 N'训练子集与所述第 N 训练子集的交集为空，所述第 N'训练子集包含至少一个训练单元；

第一获取单元，用于根据所述选取单元选取的所述第 N'训练子集获取所述第 N 树分类器的误预测率；

第二获取单元，用于根据所述第一获取单元获取到的所述第 N 树分类器误预测率获取所述第 N 树分类器的权重。

17、根据权利要求 16 所述的缺陷预测装置，其特征在于，所述预测单元包括：

统计模块，用于统计所述故障产品的属性信息；

预测模块，用于根据所述统计模块统计的所述属性信息将所述分类器集合作为预测模型预测所述故障产品的缺陷得到分类标签集合；

第三获取模块，用于根据所述分类器集合和所述分类器集合中每个树分类器的权重，获取所述分类标签集合中每个分类标签的信任值。

18、根据权利要求 11-17 中任一权利要求所述的缺陷预测装置，其特征在于，所述预设策略包括决策树算法。

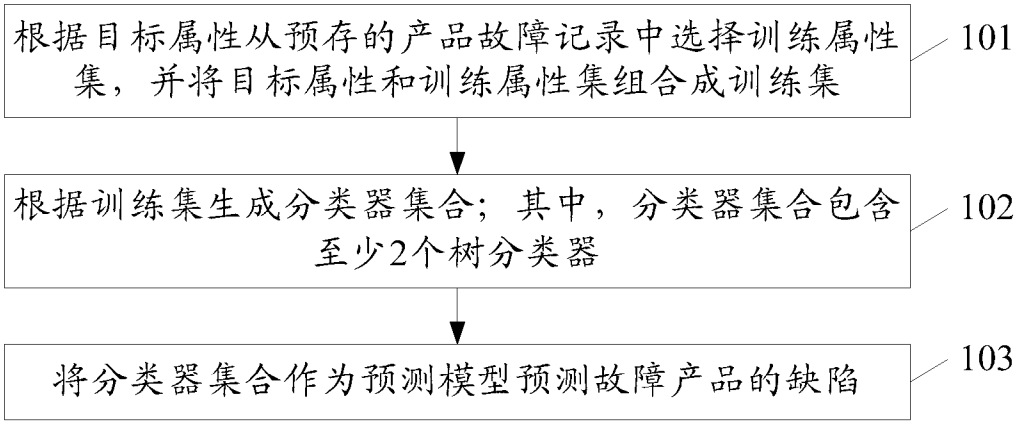
根据目标属性从预存的产品故障记录中选择训练属性集，并将目标属性和训练属性集组合成训练集
101
根据训练集生成分类器集合；其中，分类器集合包含至少2个树分类器
102
将分类器集合作为预测模型预测故障产品的缺陷
103

图 1

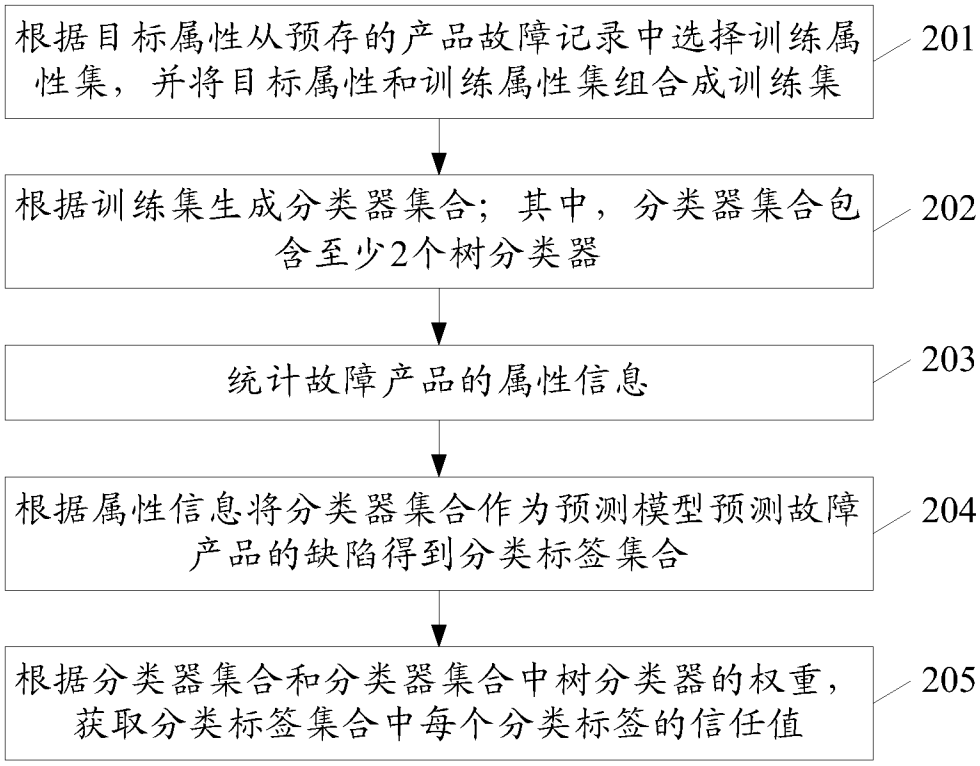
根据目标属性从预存的产品故障记录中选择训练属性集，并将目标属性和训练属性集组合成训练集
201
根据训练集生成分类器集合；其中，分类器集合包含至少2个树分类器
202
统计故障产品的属性信息
203
根据属性信息将分类器集合作为预测模型预测故障产品的缺陷得到分类标签集合
204
根据分类器集合和分类器集合中树分类器的权重，获取分类标签集合中每个分类标签的信任值
205

图 2

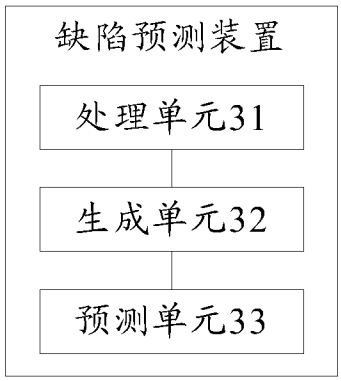
缺陷预测装置
处理单元31
生成单元32
预测单元33

图 3


缺陷预测装置
处理单元31
生成单元32
选取模块321
生成模块322
第一获取模块324
第二获取模块325
选取子模块3251
生成子模块3252
获取子模块3253
组合模块323
选取单元34
第一获取单元35
第二获取单元36
预测单元33
统计模块331
预测模块332
第三获取模块333

图 4

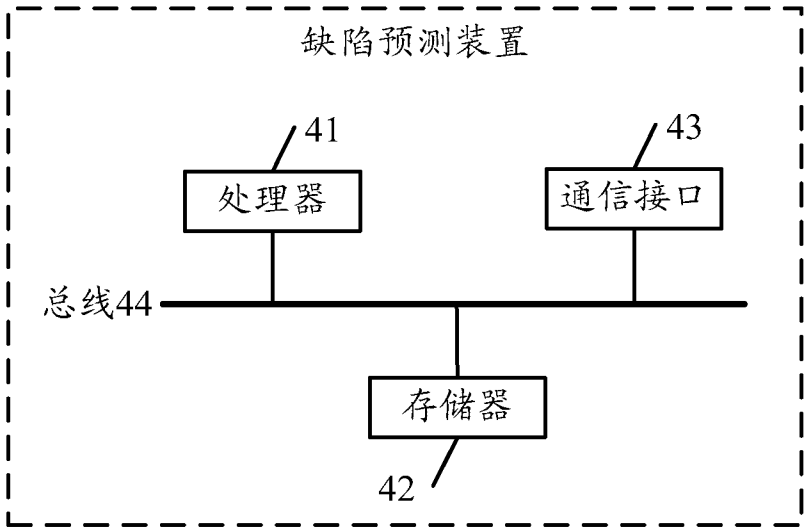

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2013/080279

A. CLASSIFICATION OF SUBJECT MATTER

G06F 19/00 (2011.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNABS, CNKI, WPI, EPODOC: forecast, bug, disfigurement, failure, fault, malfunction, trouble, attribute, data, sort, class, classify, equip, training, set

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101799320 A (BEIJING INFORMATION SCIENCE & TECHNOLOGY UNIVERSITY), 11 August 2010 (11.08.2010), see claims 1 and 3, and description, paragraphs 14-19	1, 10
A		2-9, 11-18
Y	CN 101556553 B (INSTITUTE OF SOFTWARE, CHINESE ACADEMY OF SCIENCES), 06 April 2011 (06.04.2011), see claims 1, 2 and 7, and description, paragraphs 6-17	1, 10
A		2-9, 11-18
A	CN 102928720 A (GUANGDONG POWER GRID COMPANY), 13 February 2013 (13.02.2013), see the whole document	1-18

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 November 2013 (26.11.2013)	**12 December 2013 (12.12.2013)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **ZHAO, Jing** Telephone No.: (86-10) **62411884**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2013/080279

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101799320 A	11.08.2010	CN 101799320 B	25.05.2011
CN 101556553 B	06.04.2011	CN 101556553 A	14.10.2009
CN 102928720 A	13.02.2013	None	

国际检索报告

国际申请号
PCT/CN2013/080279

A. 主题的分类

G06F 19/00 (2011.01) i

按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC: G06F /-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

CNPAT,CNABS,CNKI,WPI,EPODOC: 预测，缺陷，故障，属性，数据，分类，训练，集；forecast, bug, disfigurement, failure, fault, malfunction, trouble, attribute, data, sort, class, classify, equip, training, set

C. 相关文件

类　型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CN101799320A (北京信息科技大学) 11.8 月 2010(11.08.2010) 参见权利要求 1,3、说明书 14 至 19 段	1,10
A		2-9,11-18
Y	CN101556553B (中国科学院软件研究所) 06.4 月 2011(06.04.2011) 参见权利要求 1,2,7、说明书第 6 至 17 段	1,10
A		2-9,11-18
A	CN102928720A (广东电网公司) 13.2 月 2013(13.02.2013) 参见全文	1-18

☐ 其余文件在 C 栏的续页中列出。　　☒ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
26.11 月 2013(26.11.2013)	**12.12 月 2013 (12.12.2013)**
ISA/CN 的名称和邮寄地址： 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号：　(86-10)62019451	受权官员 赵婧 电话号码：(86-10) **62411884**

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2013/080279

检索报告中引用的专利文件	公布日期	同族专利	公布日期
CN101799320A	11.08.2010	CN101799320B	25.05.2011
CN101556553B	06.04.2011	CN101556553A	14.10.2009
CN102928720A	13.02.2013	无	