



(12) 发明专利

(10) 授权公告号 CN 109903165 B

(45) 授权公告日 2020.10.16

(21) 申请号 201811538668.6

(22) 申请日 2018.12.14

(65) 同一申请的已公布的文献号  
申请公布号 CN 109903165 A

(43) 申请公布日 2019.06.18

(73) 专利权人 阿里巴巴集团控股有限公司  
地址 英属开曼群岛大开曼资本大厦一座四  
层847号邮箱

(72) 发明人 朱标 崔志伟 章鹏 陈诗奕

(74) 专利代理机构 北京晋德允升知识产权代理  
有限公司 11623

代理人 杨移

(51) Int. Cl.

G06Q 40/08 (2012.01)

(56) 对比文件

CN 108364106 A, 2018.08.03

CN 106779271 A, 2017.05.31

CN 108428188 A, 2018.08.21

CN 108960561 A, 2018.12.07

审查员 王莹

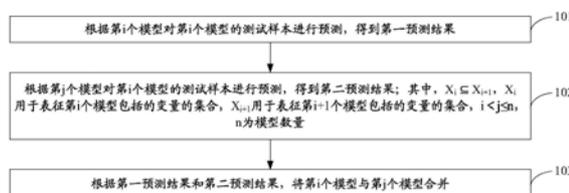
权利要求书2页 说明书9页 附图2页

(54) 发明名称

一种模型合并方法和装置

(57) 摘要

本发明提供了一种模型合并方法和装置,其中,该方法包括:根据第i个模型对所述第i个模型的测试样本进行预测,得到第一预测结果;根据第j个模型对所述第i个模型的测试样本进行预测,得到第二预测结果;其中, $X_i \subseteq X_{i+1}$ ,  $X_i$ 用于表征所述第i个模型包括的变量的集合, $X_{i+1}$ 用于表征第i+1个模型包括的变量的集合, $i < j \leq n$ , n为模型数量;根据所述第一预测结果和所述第二预测结果,将所述第i个模型与所述第j个模型合并。



1. 一种模型合并方法,包括:

根据第*i*个模型对所述第*i*个模型的测试样本进行预测,得到第一预测结果;

根据第*j*个模型对所述第*i*个模型的测试样本进行预测,得到第二预测结果;其中, $X_i \subseteq X_{i+1}$ , $X_i$ 用于表征所述第*i*个模型包括的变量的集合, $X_{i+1}$ 用于表征第*i*+1个模型包括的变量的集合, $i < j \leq n$ , $n$ 为模型数量,与所述第*i*个模型对应的业务阶段和与所述第*j*个模型对应的业务阶段不同,所述第*i*个模型为根据其对应的业务阶段的训练样本训练得到的,所述第*j*个模型为根据其对应的业务阶段的训练样本训练得到的;

根据所述第一预测结果和所述第二预测结果,将所述第*i*个模型与所述第*j*个模型合并。

2. 如权利要求1所述的模型合并方法,

所述根据所述第一预测结果和所述第二预测结果,将所述第*i*个模型与所述第*j*个模型合并,包括:

根据所述第一预测结果,确定所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率;

根据所述第二预测结果,确定所述第*j*个模型对所述第*i*个模型的测试样本的预测准确率;

当所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率和所述第*j*个模型对所述第*i*个模型的测试样本的预测准确率满足预设的合并条件时,将所述第*i*个模型与所述第*j*个模型合并。

3. 如权利要求2所述的模型合并方法,

所述第一预测结果,包括:所述第*i*个模型对所述第*i*个模型的测试样本的风险评分;

所述根据所述第一预测结果,确定所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率,包括:

根据所述第*i*个模型对所述第*i*个模型的测试样本的风险评分,确定所述第*i*个模型的测试样本是否存在风险;

根据存在风险的所述第*i*个模型的测试样本,确定所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率。

4. 如权利要求2所述的模型合并方法,

所述当所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率和所述第*j*个模型对所述第*i*个模型的测试样本的预测准确率满足预设的合并条件时,将所述第*i*个模型与所述第*j*个模型合并,包括:

当*j*取不同的值、且所述不同的值均使得所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率和所述第*j*个模型对所述第*i*个模型的测试样本的预测准确率满足所述合并条件时,将所述第*i*个模型与第*a*个模型合并;其中, $a$ 用于表征所述不同的值中的最大值。

5. 如权利要求1-4中任一所述的模型合并方法,

所述将所述第*i*个模型与所述第*j*个模型合并,包括:

确定所述第*j*个模型为所述第*i*个模型与所述第*j*个模型合并的结果。

6. 一种模型合并装置,包括:

第一预测单元,用于根据第i个模型对所述第i个模型的测试样本进行预测,得到第一预测结果;

第二预测单元,用于根据第j个模型对所述第i个模型的测试样本进行预测,得到第二预测结果;其中,  $X_i \subseteq X_{i+1}$ ,  $X_i$ 用于表征所述第i个模型包括的变量的集合,  $X_{i+1}$ 用于表征第i+1个模型包括的变量的集合,  $i < j \leq n$ , n为模型数量,与所述第i个模型对应的业务阶段和与所述第j个模型对应的业务阶段不同,所述第i个模型为根据其对应的业务阶段的训练样本训练得到的,所述第j个模型为根据其对应的业务阶段的训练样本训练得到的;

合并单元,用于根据所述第一预测结果和所述第二预测结果,将所述第i个模型与所述第j个模型合并。

7.如权利要求6所述的模型合并装置,

所述合并单元,用于根据所述第一预测结果,确定所述第i个模型对所述第i个模型的测试样本的预测准确率;根据所述第二预测结果,确定所述第j个模型对所述第i个模型的测试样本的预测准确率;当所述第i个模型对所述第i个模型的测试样本的预测准确率和所述第j个模型对所述第i个模型的测试样本的预测准确率满足预设的合并条件时,将所述第i个模型与所述第j个模型合并。

8.如权利要求7所述的模型合并装置,

所述第一预测结果,包括:所述第i个模型对所述第i个模型的测试样本的风险评分;

所述合并单元,用于根据所述第i个模型对所述第i个模型的测试样本的风险评分,确定所述第i个模型的测试样本是否存在风险;根据存在风险的所述第i个模型的测试样本,确定所述第i个模型对所述第i个模型的测试样本的预测准确率。

9.如权利要求7所述的模型合并装置,

所述合并单元,用于当j取不同的值、且所述不同的值均使得所述第i个模型对所述第i个模型的测试样本的预测准确率和所述第j个模型对所述第i个模型的测试样本的预测准确率满足所述合并条件时,将所述第i个模型与第a个模型合并;其中,a用于表征所述不同的值中的最大值。

10.如权利要求6-9中任一所述的模型合并装置,

所述合并单元,用于确定所述第j个模型为所述第i个模型与所述第j个模型合并的结果。

## 一种模型合并方法和装置

### 技术领域

[0001] 本发明涉及计算机技术领域,特别涉及一种模型合并方法和装置。

### 背景技术

[0002] 在风险识别过程中,一般在业务的多个阶段设置模型,以识别对应阶段是否存在风险。

[0003] 以车险理赔业务为例进行说明,阶段包括:报案、查勘和核赔。每个阶段对应一个模型,例如,报案阶段对应报案模型,报案模型用于识别报案阶段是否存在风险。

[0004] 但是,在实际应用场景中,业务可能存在多个阶段,如果在每个节点都设置对应的模型,则风险识别的成本较高。因此,如何降低风险识别的成本,同时保证各个阶段模型的性能,是当前亟待解决的问题。

### 发明内容

[0005] 鉴于此,本发明实施例提供了一种模型合并方法和装置,能够降低风险识别的成本。

[0006] 第一方面,本发明实施例提供了一种模型合并方法,包括:

[0007] 根据第*i*个模型对所述第*i*个模型的测试样本进行预测,得到第一预测结果;

[0008] 根据第*j*个模型对所述第*i*个模型的测试样本进行预测,得到第二预测结果;其中,

$X_i \subseteq X_{i+1}$ ,  $X_i$ 用于表征所述第*i*个模型包括的变量的集合,  $X_{i+1}$ 用于表征第*i*+1个模型包括的变量的集合,  $i < j \leq n$ ,  $n$ 为模型数量;

[0009] 根据所述第一预测结果和所述第二预测结果,将所述第*i*个模型与所述第*j*个模型合并。

[0010] 优选地,

[0011] 所述根据所述第一预测结果和所述第二预测结果,将所述第*i*个模型与所述第*j*个模型合并,包括:

[0012] 根据所述第一预测结果,确定所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率;

[0013] 根据所述第二预测结果,确定所述第*j*个模型对所述第*i*个模型的测试样本的预测准确率;

[0014] 当所述第*i*个模型对所述第*i*个模型的测试样本的预测准确率和所述第*j*个模型对所述第*i*个模型的测试样本的预测准确率满足预设的合并条件时,将所述第*i*个模型与所述第*j*个模型合并。

[0015] 优选地,

[0016] 所述第一预测结果,包括:所述第*i*个模型对所述第*i*个模型的测试样本的风险评分;

[0017] 所述根据所述第一预测结果,确定所述第*i*个模型对所述第*i*个模型的测试样本的

预测准确率,包括:

[0018] 根据所述第i个模型对所述第i个模型的测试样本的风险评分,确定所述第i个模型的测试样本是否存在风险;

[0019] 根据存在风险的所述第i个模型的测试样本,确定所述第i个模型对所述第i个模型的测试样本的预测准确率。

[0020] 优选地,

[0021] 所述当所述第i个模型对所述第i个模型的测试样本的预测准确率和所述第j个模型对所述第i个模型的测试样本的预测准确率满足预设的合并条件时,将所述第i个模型与所述第j个模型合并,包括:

[0022] 当j取不同的值、且所述不同的值均使得所述第i个模型对所述第i个模型的测试样本的预测准确率和所述第j个模型对所述第i个模型的测试样本的预测准确率满足所述合并条件时,将所述第i个模型与第a个模型合并;其中,a用于表征所述不同的值中的最大值。

[0023] 优选地,

[0024] 所述将所述第i个模型与所述第j个模型合并,包括:

[0025] 确定所述第j个模型为所述第i个模型与所述第j个模型合并的结果。

[0026] 第二方面,本发明实施例提供了一种模型合并装置,包括:

[0027] 第一预测单元,用于根据第i个模型对所述第i个模型的测试样本进行预测,得到第一预测结果;

[0028] 第二预测单元,用于根据第j个模型对所述第i个模型的测试样本进行预测,得到第二预测结果;其中, $X_i \subseteq X_{i+1}$ , $X_i$ 用于表征所述第i个模型包括的变量的集合, $X_{i+1}$ 用于表征第i+1个模型包括的变量的集合, $i < j \leq n$ ,n为模型数量;

[0029] 合并单元,用于根据所述第一预测结果和所述第二预测结果,将所述第i个模型与所述第j个模型合并。

[0030] 优选地,

[0031] 所述合并单元,用于根据所述第一预测结果,确定所述第i个模型对所述第i个模型的测试样本的预测准确率;根据所述第二预测结果,确定所述第j个模型对所述第i个模型的测试样本的预测准确率;当所述第i个模型对所述第i个模型的测试样本的预测准确率和所述第j个模型对所述第i个模型的测试样本的预测准确率满足预设的合并条件时,将所述第i个模型与所述第j个模型合并。

[0032] 优选地,

[0033] 所述第一预测结果,包括:所述第i个模型对所述第i个模型的测试样本的风险评分;

[0034] 所述合并单元,用于根据所述第i个模型对所述第i个模型的测试样本的风险评分,确定所述第i个模型的测试样本是否存在风险;根据存在风险的所述第i个模型的测试样本,确定所述第i个模型对所述第i个模型的测试样本的预测准确率。

[0035] 优选地,

[0036] 所述合并单元,用于当j取不同的值、且所述不同的值均使得所述第i个模型对所述第i个模型的测试样本的预测准确率和所述第j个模型对所述第i个模型的测试样本的预

测准确率满足所述合并条件时,将所述第i个模型与第a个模型合并;其中,a用于表征所述不同的值中的最大值。

[0037] 优选地,

[0038] 所述合并单元,用于确定所述第j个模型为所述第i个模型与所述第j个模型合并的结果。

[0039] 本发明实施例采用的上述至少一个技术方案能够达到以下有益效果:该方法根据模型对测试样本的预测结果,合并不同的模型,能够减少模型的数量,降低风险识别的成本。同时,相比于通过缺失值填充的方式直接复用模型,该方法能够保证各个阶段的模型性能,提高风险识别的准确率。

## 附图说明

[0040] 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他的附图。

[0041] 图1是本发明一个实施例提供的一种模型合并方法的流程图;

[0042] 图2是本发明一个实施例提供的一种模型变量的关系的示意图;

[0043] 图3是本发明另一个实施例提供的一种模型合并方法的流程图;

[0044] 图4是本发明一个实施例提供的一种模型合并装置的结构示意图。

## 具体实施方式

[0045] 为使本发明实施例的目的、技术方案和优点更加清楚,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的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。

[0046] 在传统的车险反欺诈领域,一般在每个阶段(风险点)设置一个模型。无论是从数据角度还是模型角度,这种做法简单、直接,无需关注不同阶段之间在业务、数据方面的关联。但是,随着模型数量的增加,其研发、部署、客户接入的成本快速提高。

[0047] 另一种常见的做法是保留变量最全的阶段模型,例如,车险反欺诈场景中的核赔阶段模型,其余阶段模型通过缺失值填充的方式复用核赔阶段模型。这种做法直接把模型数量压缩到了极致,研发、部署、客户接入成本做到了最低,但是,其没办法保证各个阶段的模型性能。

[0048] 鉴于此,本发明实施例提供了一种模型合并方法,如图1所示,该方法可以包括以下步骤:

[0049] 步骤101:根据第i个模型对第i个模型的测试样本进行预测,得到第一预测结果。

[0050] 其中,i为变量,其取值范围为(0,n),其中,n为模型数量,即阶段数量。例如,第1个模型对第1个模型的测试样本进行预测,第2个模型对第2个模型的测试样本进行预测。模型可以是预先设置的,也可以是根据相应阶段的训练样本训练得到的。例如,第i个模型由其对应的训练样本训练得到。

[0051] 步骤102:根据第j个模型对第i个模型的测试样本进行预测,得到第二预测结果;其中,  $X_i \subseteq X_{i+1}$ ,  $X_i$ 用于表征第i个模型包括的变量的集合,  $X_{i+1}$ 用于表征第i+1个模型包括的变量的集合,  $i < j \leq n$ , n为模型数量。

[0052] 与i类似, j也为变量, 其取值范围为 (i, n]。对于i的每一个值, j的值可能存在多个。例如, n=4, 当i=1时, j的值可以为2、3和4中任意一个或多个。当i=2时, j的值可以为3和4中任意一个或多个。

[0053] 在实际应用场景中, 每个模型对应相应的业务阶段, 不同的业务阶段存在时间上的先后顺序。不同的业务阶段按照时间顺序, 对应的模型分别为第1个模型、第2个模型、第3个模型.....第n个模型。

[0054] 在车险理赔的业务场景中, 模型包括报案模型、查勘模型和核赔模型, 分别用于识别报案、查勘、核赔三个阶段存在的风险。这三个阶段存在时间上的先后顺序, 首先被保人在事故后进行报案(报案阶段), 保险公司在接到报案后会派出查勘员进行事故现场查勘(查勘阶段), 然后在调查结束后, 保险公司会对案件进行核赔(核赔阶段), 核赔一旦完成, 赔款就会打到被保人的指定账户中。

[0055] 从报案阶段到查勘阶段, 最后到核赔阶段, 保险公司获取到的案件信息会越来越丰富, 所以每个阶段的信息是后一个阶段信息的子集。对应到模型中, 前一阶段模型包含的变量的集合  $X_i$  和后一阶段模型包含的变量的集合  $X_{i+1}$  是包含的关系, 即  $X_i \subseteq X_{i+1}$ , 各个阶段的变量关系如图2所示。

[0056] 步骤103:根据第一预测结果和第二预测结果, 将第i个模型与第j个模型合并。

[0057] 在本说明实施例中, 由于i和j可以取不同的值, 因此, 第一预测结果的数量和第二预测结果的数量可能为多个, 但模型的合并方法是一致的。

[0058] 例如, n=3, 存在与i=1对应的第一预测结果、与i=2对应的第一预测结果和与i=3对应的第一预测结果。

[0059] 当i=1时, 存在与j=2对应的第二预测结果和与j=3对应的第二预测结果。

[0060] 在实际应用场景中, 可以根据第一预测结果和第二预测结果, 确定第i个模型的性能与第j个模型的性能是否相近, 如果相近, 则将第i个模型与第j个模型合并。基于此, 上述报案模型、查勘模型和核赔模型可以根据表1进行合并。

[0061] 表1

	核赔模型和查勘模型性能是否相近	核赔模型和报案模型性能是否相近	查勘模型和报案模型性能是否相近	合并结果
[0062]	是	是	-	只保留核赔模型
	是	否	-	保留核赔模型、报案模型
	否	是	-	保留核赔模型、查勘模型
	否	否	是	保留核赔模型、查勘模型
	否	否	否	保留全部三个模型

[0063] 第i个模型的性能与第j个模型的性能是否相近, 可以用第i个模型对第i个模型的

测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率的差值衡量。

[0064] 在此情况下,步骤103具体包括:

[0065] A1:根据第一预测结果,确定第i个模型对第i个模型的测试样本的预测准确率。

[0066] 第一预测结果,包括:第i个模型对第i个模型的测试样本的风险评分。

[0067] A1具体包括:

[0068] A11:根据第i个模型对第i个模型的测试样本的风险评分,确定第i个模型的测试样本是否存在风险。

[0069] 在实际应用场景中,可以将第i个模型对第i个模型的测试样本的风险评分与预设的风险范围进行匹配,以确定第i个模型的测试样本是否存在风险

[0070] A12:根据存在风险的第i个模型的测试样本,确定第i个模型对第i个模型的测试样本的预测准确率。

[0071] 在本发明实施例中,第i个模型对第i个模型的测试样本的预测准确率=预测为存在风险的第i个模型的测试样本中真实存在风险的第i个模型的测试样本的数量/预测(确定)为存在风险的第i个模型的测试样本的数量。其中,真实存在风险指的是已确认存在风险,例如,具有风险标签的样本为真实存在风险的样本。

[0072] A2:根据第二预测结果,确定第j个模型对第i个模型的测试样本的预测准确率。

[0073] 第二预测结果与第一预测结果类似。第二预测结果,包括:第j个模型对第i个模型的测试样本的风险评分;

[0074] 确定第j个模型对第i个模型的测试样本的预测准确率的方法与上述A1中的过程类似,此处不再赘述。

[0075] A3:当第i个模型对第i个模型的测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率满足预设的合并条件时,将第i个模型与第j个模型合并。

[0076] 合并条件可以为第i个模型对第i个模型的测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率的差值小于预设的合并阈值。例如,合并阈值为5%,即两种预测准确率的差值小于5%时,才能够将第i个模型与第j个模型合并。

[0077] A3具体包括:

[0078] 当j取不同的值、且不同的值均使得第i个模型对第i个模型的测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率满足合并条件时,将第i个模型与第a个模型合并;其中,a用于表征不同的值中的最大值。

[0079] 沿用上例,预设的合并阈值为5%。第1个模型对第1个模型的测试样本的预测准确率和第2个模型对第1个模型的测试样本的预测准确率的差值为2%,同时,第1个模型对第1个模型的测试样本的预测准确率和第3个模型对第1个模型的测试样本的预测准确率的差值为3%,则将第1个模型与第3个模型合并。

[0080] 该方法根据模型对测试样本的预测结果,合并不同的模型,能够减少模型的数量,降低风险识别的成本。同时,相比于通过缺失值填充的方式直接复用模型,该方法能够保证各个阶段的模型性能,提高风险识别的准确率。

[0081] 在本发明的一个实施例中,将第i个模型与第j个模型合并,包括:确定第j个模型为第i个模型与第j个模型合并的结果。

[0082] 例如,第1个模型与第2个模型合并指的是保留第2个模型,合并前如果存在第1个

模型与第2个模型,则合并后仅存在第2个模型,仅使用第2个模型进行风险识别。

[0083] 需要说明的是,当j存在多个值时,可以每个值均执行步骤102,也可以仅对其部分值执行步骤102。例如, $i=1, n=5$ ,j的值可以为2、3、4、5,可以确定 $j=2$ 、 $j=3$ 、 $j=4$ 和 $j=5$ 时的第二预测结果,也可以仅确定 $j=5$ 时的第二预测结果。换言之,该方法可以根据实际场景的需求确定保留的模型的数量。

[0084] 如图3所示,本发明实施例以车险理赔场景为例,对模型合并方法进行详细的说明,该方法包括:

[0085] 步骤301:根据第i个模型对第i个模型的测试样本进行预测,得到第一预测结果;第一预测结果,包括:第i个模型对第i个模型的测试样本的风险评分。

[0086] 步骤302:根据第j个模型对第i个模型的测试样本进行预测,得到第二预测结果;其中, $X_i \subseteq X_{i+1}$ , $X_i$ 用于表征第i个模型包括的变量的集合, $X_{i+1}$ 用于表征第i+1个模型包括的变量的集合, $i < j \leq n$ ,n为模型数量。

[0087] 在本发明实施例中, $n=3$ ,第1个模型为报案模型,第2个模型为查勘模型,第3个模型为核赔模型,其包括的变量的关系请参考图2。

[0088] 在本发明实施例中,该方法可以概括为以下几种情况:

[0089] 情况1:确定是否可以将查勘模型和核赔模型进行合并。

[0090] 此时, $i=2, j=3$ 。

[0091] 根据查勘模型对查勘模型的测试样本进行预测,得到第一预测结果;

[0092] 根据核赔模型对查勘模型的测试样本进行预测,得到第二预测结果。

[0093] 情况2:确定是否可以将报案模型和查勘模型,或报案模型和核赔模型进行合并。

[0094] 此时, $i=1, j=2$ 或 $j=3$ 。

[0095] 根据报案模型对报案模型的测试样本进行预测,得到第一预测结果;

[0096] 根据查勘模型对报案模型的测试样本进行预测,得到第二预测结果(与查勘模型对应);

[0097] 根据核赔模型对报案模型的测试样本进行预测,得到第二预测结果(与核赔模型对应)。

[0098] 需要说明的是,核赔模型变量最全,预测性能一般来说也是最好的,优先保留。

[0099] 步骤303:根据第i个模型对第i个模型的测试样本的风险评分,确定第i个模型的测试样本是否存在风险。

[0100] 步骤304:根据存在风险的第i个模型的测试样本,确定第i个模型对第i个模型的测试样本的预测准确率。

[0101] 步骤305:根据第二预测结果,确定第j个模型对第i个模型的测试样本的预测准确率。

[0102] 步骤306:当第i个模型对第i个模型的测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率满足预设的合并条件时,确定第j个模型为第i个模型与第j个模型合并的结果。

[0103] 在实际应用场景中,可以将通过该方法得到的模型用于识别车险理赔不同的阶段存在的风险,及时发现欺诈行为,保障资金安全。

[0104] 如图4所示,本发明实施例提供了一种模型合并装置,包括:

[0105] 第一预测单元401,用于根据第i个模型对第i个模型的测试样本进行预测,得到第一预测结果;

[0106] 第二预测单元402,用于根据第j个模型对第i个模型的测试样本进行预测,得到第二预测结果;其中,  $X_i \subseteq X_{i+1}$ ,  $X_i$ 用于表征第i个模型包括的变量的集合,  $X_{i+1}$ 用于表征第i+1个模型包括的变量的集合,  $i < j \leq n$ , n为模型数量;

[0107] 合并单元403,用于根据第一预测结果和第二预测结果,将第i个模型与第j个模型合并。

[0108] 在本发明的一个实施例中,合并单元403,用于根据第一预测结果,确定第i个模型对第i个模型的测试样本的预测准确率;根据第二预测结果,确定第j个模型对第i个模型的测试样本的预测准确率;当第i个模型对第i个模型的测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率满足预设的合并条件时,将第i个模型与第j个模型合并。

[0109] 在本发明的一个实施例中,第一预测结果,包括:第i个模型对第i个模型的测试样本的风险评分;

[0110] 合并单元403,用于根据第i个模型对第i个模型的测试样本的风险评分,确定第i个模型的测试样本是否存在风险;根据存在风险的第i个模型的测试样本,确定第i个模型对第i个模型的测试样本的预测准确率。

[0111] 在本发明的一个实施例中,合并单元403,用于当j取不同的值、且不同的值均使得第i个模型对第i个模型的测试样本的预测准确率和第j个模型对第i个模型的测试样本的预测准确率满足合并条件时,将第i个模型与第a个模型合并;其中,a用于表征不同的值中的最大值。

[0112] 在本发明的一个实施例中,合并单元403,用于确定第j个模型为第i个模型与第j个模型合并的结果。

[0113] 在20世纪90年代,对于一个技术的改进可以很明显地区分是硬件上的改进(例如,对二极管、晶体管、开关等电路结构的改进)还是软件上的改进(对于方法流程的改进)。然而,随着技术的发展,当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此,不能说一个方法流程的改进就不能用硬件实体模块来实现。例如,可编程逻辑器件(Programmable Logic Device,PLD)(例如现场可编程门阵列(Field Programmable Gate Array,FPGA))就是这样一种集成电路,其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片PLD上,而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且,如今,取代手工地制作集成电路芯片,这种编程也多半改用“逻辑编译器(logic compiler)”软件来实现,它与程序开发撰写时所用的软件编译器相类似,而要编译之前的原始代码也得用特定的编程语言来撰写,此称之为硬件描述语言(Hardware Description Language,HDL),而HDL也并非仅有一种,而是有许多种,如ABEL(Advanced Boolean Expression Language)、AHDL(Altera Hardware Description Language)、Confluence、CUPL(Cornell University Programming Language)、HDCal、JHDL(Java Hardware Description Language)、Lava、Lola、MyHDL、PALASM、RHDL(Ruby Hardware Description Language)等,目前最普遍使用的是VHDL(Very-High-Speed

Integrated Circuit Hardware Description Language) 与 Verilog。本领域技术人员也应该清楚,只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中,就可以很容易得到实现该逻辑方法流程的硬件电路。

[0114] 控制器可以按任何适当的方式实现,例如,控制器可以采取例如微处理器或处理器以及存储可由该(微)处理器执行的计算机可读程序代码(例如软件或固件)的计算机可读介质、逻辑门、开关、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程逻辑控制器和嵌入微控制器的形式,控制器的例子包括但不限于以下微控制器:ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20以及Silicone Labs C8051F320,存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道,除了以纯计算机可读程序代码方式实现控制器以外,完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件,而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至,可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[0115] 上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的,计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0116] 为了描述的方便,描述以上装置时以功能分为各种单元分别描述。当然,在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0117] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0118] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0119] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0120] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一

个方框或多个方框中指定的功能的步骤。

[0121] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0122] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0123] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0124] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0125] 本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0126] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0127] 以上所述仅为本申请的实施例而已,并不用于限制本申请。对于本领域技术人员来说,本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本申请的权利要求范围之内。

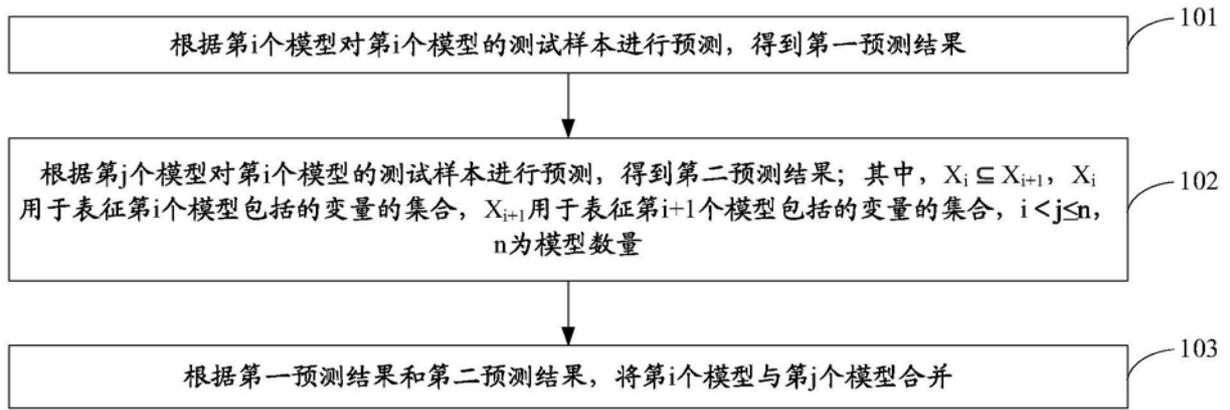


图1

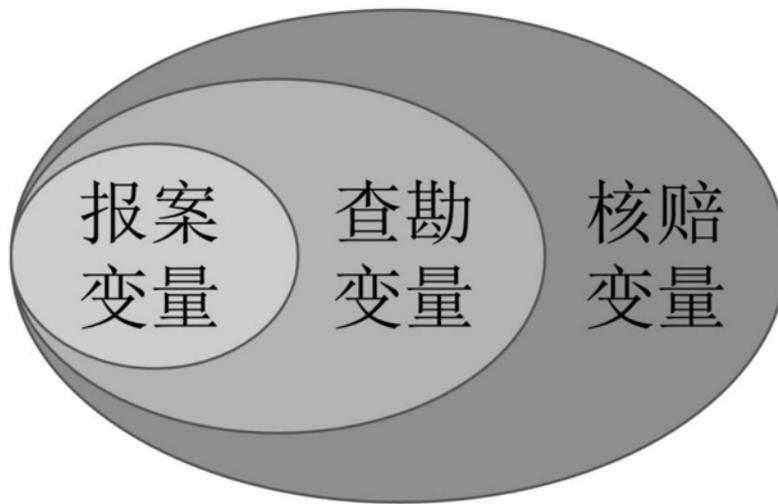


图2

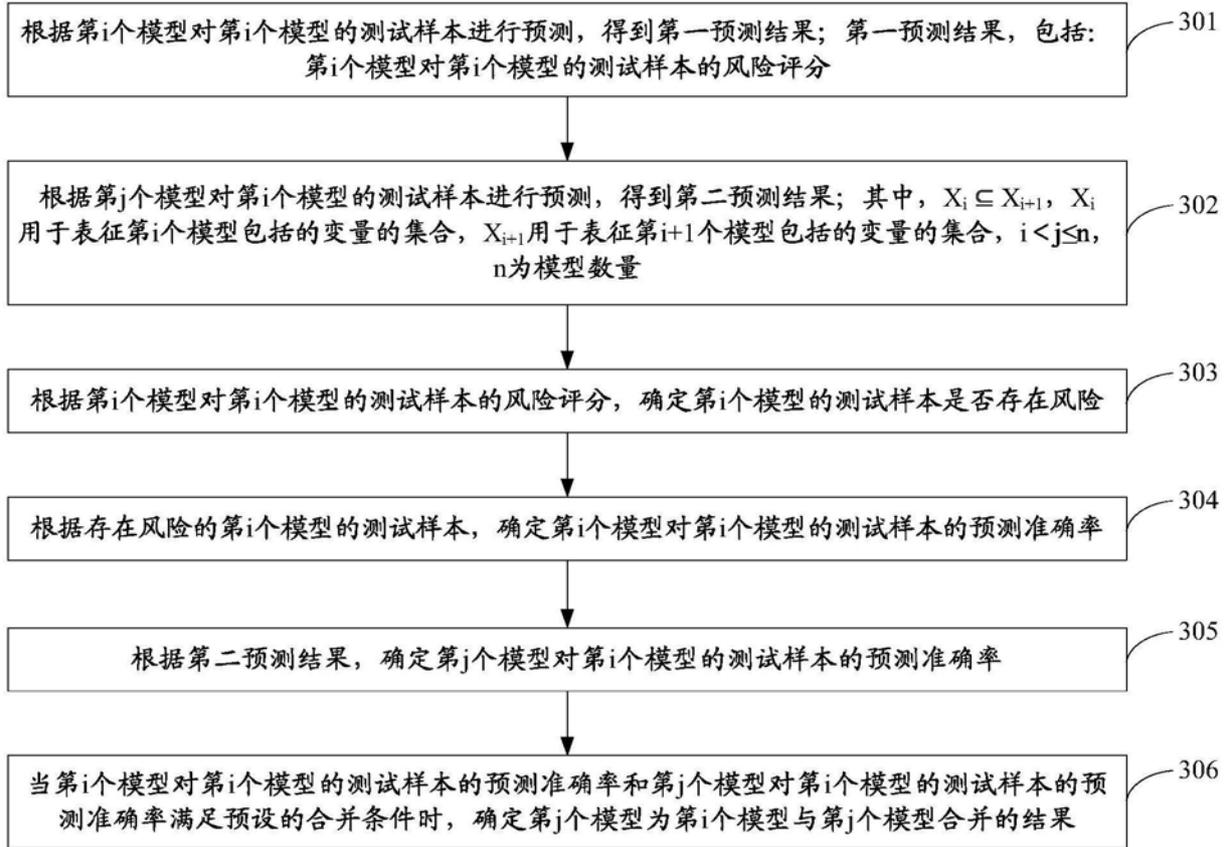


图3

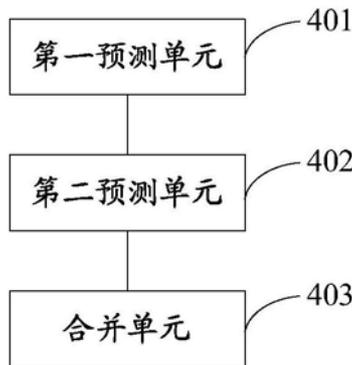


图4