(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103810095 B
(45)授权公告日 2018.01.05

(21)申请号 201210459305.X

(22)申请日 2012.11.15

(65)同一申请的已公布的文献号
申请公布号 CN 103810095 A

(43)申请公布日 2014.05.21

(73)专利权人 百度在线网络技术(北京)有限公司
地址 100085 北京市海淀区上地十街10号百度大厦

(72)发明人 沙安澜

(74)专利代理机构 北京鸿德海业知识产权代理事务所(普通合伙) 11412
代理人 倪志华

(51)Int.Cl.
G06F 11/36(2006.01)

(56)对比文件
CN 101452068 A,2009.06.10,
US 6986125 B2,2006.01.10,
CN 102541736 A,2012.07.04,
王宝龙 等.基于依赖矩阵的测试性分析.《计算机测量与控制》.2011,第19卷(第6期),
曹胜玉 等.非负矩阵分解及其在基因表达数据分析中的应用.《北京师范大学学报(自然科学版)》.2007,第43卷(第1期),

审查员 王仕超

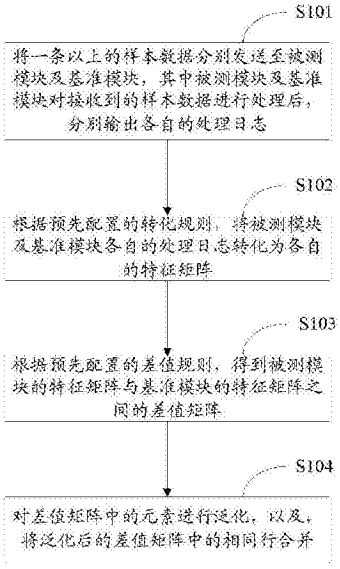
权利要求书2页 说明书9页 附图4页

(54)发明名称

一种数据对比测试的方法及装置

(57)摘要

本发明提供了一种数据对比测试的方法及装置,其中数据对比测试的方法包括:A.将一条以上的样本数据分别发送至被测模块及基准模块,其中所述被测模块及所述基准模块对接收到的样本数据进行处理后,分别输出各自的处理日志;B.根据预先配置的转化规则,将所述被测模块及所述基准模块各自的处理日志转化为各自的特征矩阵;C.根据预先配置的差值规则,得到所述被测模块的特征矩阵与所述基准模块的特征矩阵之间的差值矩阵;D.对所述差值矩阵中的元素进行泛化,以及,将泛化后的差值矩阵中的相同行合并。通过上述方式,能够提高测试的精准度。



CN 103810095 B

1.一种数据对比测试的方法，包括：

A.将一条以上的样本数据分别发送至被测模块及基准模块，其中所述被测模块及所述基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志；

B.根据预先配置的转化规则，将所述被测模块及所述基准模块各自的处理日志转化为各自的特征矩阵；

C.根据预先配置的差值规则，得到所述被测模块的特征矩阵与所述基准模块的特征矩阵之间的差值矩阵；

D.根据预先配置的泛化规则表对所述差值矩阵中的元素进行泛化，以及，将泛化后的差值矩阵中的相同行合并。

2.根据权利要求1所述的方法，其特征在于，所述处理日志的每条记录包含一条样本数据，以及由该样本数据得到的至少一个维度的处理结果。

3.根据权利要求2所述的方法，其特征在于，所述特征矩阵的每个元素表示一条样本数据得到的一个维度的处理结果，并且，同一行的元素对应同一条样本数据，同一列的元素对应同一个维度的处理结果。

4.根据权利要求1所述的方法，其特征在于，“根据预先配置的泛化规则表对所述差值矩阵中的元素进行泛化”的步骤具体包括：

针对所述差值矩阵中的每个元素，查找预先配置的泛化规则表，当所述泛化规则表中有该元素的适用规则时，将该元素按照所述适用规则进行泛化。

5.根据权利要求1所述的方法，其特征在于，“将泛化后的差值矩阵中的相同行合并”的步骤具体包括：

将泛化后的差值矩阵中的同行元素进行拼接；

对拼接后的各行分别计算该行的特征值；

对特征值相同的行进行合并。

6.一种数据对比测试的装置，包括：

日志获取单元，用于将一条以上的样本数据分别发送至被测模块及基准模块，其中所述被测模块及所述基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志；

转化单元，用于根据预先配置的转化规则，将所述被测模块及所述基准模块各自的处理日志转化为各自的特征矩阵；

差值获取单元，用于根据预先配置的差值规则，得到所述被测模块的特征矩阵与所述基准模块的特征矩阵之间的差值矩阵；

泛化单元，用于根据预先配置的泛化规则表对所述差值矩阵中的元素进行泛化；

合并单元，用于将泛化后的差值矩阵中的相同行合并。

7.根据权利要求6所述的装置，其特征在于，所述处理日志的每条记录包含一条样本数据，以及，由该样本数据得到的至少一个维度的处理结果。

8.根据权利要求7所述的装置，其特征在于，所述特征矩阵的每个元素表示一条样本数据得到的一个维度的处理结果，并且，同一行的元素对应同一条样本数据，同一列的元素对应同一个维度的处理结果。

9.根据权利要求6所述的装置，其特征在于，所述泛化单元根据预先配置的泛化规则表对所述差值矩阵中的元素进行泛化的方式具体包括：

针对所述差值矩阵中的每个元素，查找预先配置的泛化规则表，当所述泛化规则表中有该元素的适用规则时，将该元素按照所述适用规则进行泛化。

10.根据权利要求6所述的装置，其特征在于，所述合并单元具体包括：

拼接单元，用于将泛化后的差值矩阵中的同行元素进行拼接；

计算单元，用于对拼接后的各行分别计算该行的特征值；

行合并单元，用于对特征值相同的行进行合并。

一种数据对比测试的方法及装置

【技术领域】

[0001] 本发明涉及测试技术，特别涉及一种数据对比测试的方法及装置。

【背景技术】

[0002] 对比测试是一种常见的测试方法。其实施方式为：将被测模块及基准模块置于相同的测试环境中，并采用相同的测试数据分别作为被测模块和基准模块的输入，以比较被测模块和基准模块各自输出之间的差异性来验证被测模块是否符合预期的设计。其中，基准模块是用来与被测模块进行对比的模块，例如对发布的模块进行了升级，对升级后的模块进行对比测试时，升级前的模块就是基准模块，升级后的模块就是被测模块。对计算密集型模块和遗留系统采取对比测试方法往往比根据功能设计进行主动验证测试更具可实施性，也更加有效率。

[0003] 对比测试通常需要大量的测试数据，才能使得被测模块的测试覆盖足够充分，但是当测试数据的数量很大时，被测模块或基准模块输出的数据也相当巨大，测试人员对巨大的输出数据进行逐一分析几乎是不可能完成的任务。在现有的对比测试中，针对大量的输出数据，测试人员通常是采用对输出数据进行抽样分析的方法来确定被测模块是否符合预期的。

[0004] 抽样分析由于不能对所有的数据进行分析，有可能存在数据遗漏的问题，统计数据表明，万分之二的数据导致的预期外差异，有可能在长达两年的时间内都难以发现。

[0005] 可以看出，现有的对比测试方法，由于输出的结果难以被有效分析，因此存在测试精准度差的问题。

【发明内容】

[0006] 本发明所要解决的技术问题是提供一种数据对比测试的方法及装置，以提高测试的精准度。

[0007] 本发明为解决技术问题而采用的技术方案是提供一种数据对比测试的方法，包括：A.将一条以上的样本数据分别发送至被测模块及基准模块，其中所述被测模块及所述基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志；B.根据预先配置的转化规则，将所述被测模块及所述基准模块各自的处理日志转化为各自的特征矩阵；C.根据预先配置的差值规则，得到所述被测模块的特征矩阵与所述基准模块的特征矩阵之间的差值矩阵；D.对所述差值矩阵中的元素进行泛化，以及，将泛化后的差值矩阵中的相同行合并。

[0008] 根据本发明之一优选实施例，所述处理日志的每条记录包含一条样本数据，以及由该样本数据得到的至少一个维度的处理结果。

[0009] 根据本发明之一优选实施例，所述特征矩阵的每个元素表示一条样本数据得到的一个维度的处理结果，并且，同一行的元素对应同一条样本数据，同一列的元素对应同一个维度的处理结果。

[0010] 根据本发明之一优选实施例，“对所述差值矩阵中的元素进行泛化”的步骤具体包括：针对所述差值矩阵中的每个元素，查找预先配置的泛化规则表，当所述泛化规则表中有该元素的适用规则时，将该元素按照所述适用规则进行泛化。

[0011] 根据本发明之一优选实施例，“将泛化后的差值矩阵中的相同行合并”的步骤具体包括：将泛化后的差值矩阵中的同行元素进行拼接；对拼接后的各行分别计算该行的特征值；对特征值相同的行进行合并。

[0012] 本发明还提供了一种数据对比测试的装置，包括：日志获取单元，用于将一条以上的样本数据分别发送至被测模块及基准模块，其中所述被测模块及所述基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志；转化单元，用于根据预先配置的转化规则，将所述被测模块及所述基准模块各自的处理日志转化为各自的特征矩阵；

[0013] 差值获取单元，用于根据预先配置的差值规则，得到所述被测模块的特征矩阵与所述基准模块的特征矩阵之间的差值矩阵；泛化单元，用于对所述差值矩阵中的元素进行泛化；合并单元，用于将泛化后的差值矩阵中的相同行合并。

[0014] 根据本发明之一优选实施例，所述处理日志的每条记录包含一条样本数据，以及，由该样本数据得到的至少一个维度的处理结果。

[0015] 根据本发明之一优选实施例，所述特征矩阵的每个元素表示一条样本数据得到的一个维度的处理结果，并且，同一行的元素对应同一条样本数据，同一列的元素对应同一个维度的处理结果。

[0016] 根据本发明之一优选实施例，所述泛化单元对所述差值矩阵中的元素进行泛化的方式具体包括：针对所述差值矩阵中的每个元素，查找预先配置的泛化规则表，当所述泛化规则表中有该元素的适用规则时，将该元素按照所述适用规则进行泛化。

[0017] 根据本发明之一优选实施例，所述合并单元具体包括：拼接单元，用于将泛化后的差值矩阵中的同行元素进行拼接；计算单元，用于对拼接后的各行分别计算该行的特征值；行合并单元，用于对特征值相同的行进行合并。

[0018] 由以上技术方案可以看出，本发明通过将对比测试中被测模块及基准模块的处理日志分别通过矩阵建模的方式转化为特征矩阵，并通过被测模块的特征矩阵与基准模块的特征矩阵得到差值矩阵，能够对差值矩阵进行自动的数据分析，其中差值矩阵中的各行元素通过数据泛化归类去重，能够有效化简，因此，本发明与传统的对比测试相比，无论测试数据的数量有多大，都可以做到全量的数据分析，而不是通过数据抽样的方式观察被测模块是否符合预期，这不仅极大地降低了测试过程中人工介入的程度，也能够提高测试的精准度。

【附图说明】

[0019] 图1为本发明中数据比对测试的方法的实施例的流程示意图；

[0020] 图2a为本发明中被测模块的特征矩阵的示意图；

[0021] 图2b为本发明中基准模块的特征矩阵的示意图；

[0022] 图3为本发明中差值矩阵的示意图；

[0023] 图4为本发明中差值矩阵泛化后的示意图；

[0024] 图5为本发明中数据比对测试的装置的实施例的结构示意框图；

[0025] 图6为本发明中合并单元205的实施例的结构示意框图。

【具体实施方式】

[0026] 为了使本发明的目的、技术方案和优点更加清楚，下面结合附图和具体实施例对本发明进行详细描述。

[0027] 请参考图1，图1为本发明中数据对比测试的方法的实施例的流程示意图。如图1所示，该方法包括：

[0028] 步骤S101：将一条以上的样本数据分别发送至被测模块及基准模块，其中被测模块及基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志。

[0029] 步骤S102：根据预先配置的转化规则，将被测模块及基准模块各自的处理日志转化为各自的特征矩阵。

[0030] 步骤S103：根据预先配置的差值规则，得到被测模块的特征矩阵与基准模块的特征矩阵之间的差值矩阵；

[0031] 步骤S104：对差值矩阵中的元素进行泛化，以及，将泛化后的差值矩阵中的相同行合并。

[0032] 下面对上述步骤进行具体说明。

[0033] 步骤S101中，一条样本数据是被测模块或基准模块完成一次处理过程所需要的基本数据单元。例如，被测模块或基准模块的功能是对页面进行分类，其基本数据单元是一个页面的URL地址。

[0034] 本发明中，发送至被测模块及基准模块的样本数据是相同的，即同一条样本数据会分别发送至被测模块及基准模块。被测模块对接收到的各条样本数据进行处理后，将输出自己的处理日志，基准模块对接收到的各条样本数据进行处理后，也会输出自己的处理日志。

[0035] 在上述处理日志中，每条记录包含一条样本数据，以及由该样本数据得到的至少一个维度的处理结果。

[0036] 请参考下面的伪代码块：

[0037]

```
A（url）{
redir = process related to url;   //对 url 进行处理得到 redir
type = process related to url;   //对 url 进行处理得到 type
value = process related to url; //对 url 进行处理得到 value
weight = process related to url; //对 url 进行处理得到 weight
x = B（redir,type,value,weight）; //输出 redir、type、value、weight 作为
…                                  被调模块 B 的输入
…
}
```

[0038] 上述的模块A可以是被测模块或基准模块。

[0039] 与上述伪代码对应的，被测模块的处理日志请参考下面日志记录：

[0040] 第1条记录:url=www.sina.com,redir=null,type=A,value=10,weight=30

[0041] 第2条记录:…

[0042] 第3条记录:…

[0043] 与上述伪代码对应的，基准模块的处理日志请参考下面的日志记录：

[0044] 第1条记录：

[0045] url=www.sina.com,redir=www.sina.com/index.html,type=C,value=20,weight=40

[0046] 第2条记录:…

[0047] 第3条记录:…

[0048] 上述日志记录中，“url”字段对应记录的就是样本数据，“redir”、“type”、“value”、“weight”字段对应记录的就是一个维度的处理结果。

[0049] 步骤S102中，根据被测模块的处理日志，可以转化得到被测模块的特征矩阵，根据基准模块的处理日志，可以转化得到基准模块的特征矩阵。

[0050] 具体地，特征矩阵中的每个元素，表示一条样本数据得到的一个维度的处理结果，并且，同一行的元素对应同一条样本数据，同一列的元素对应同一个维度的处理结果。

[0051] 在将处理日志中的字段转化为特征矩阵中的元素时，根据预先配置的字段类型，查找与该类型对应的转化规则，即可得到该字段转化为特征矩阵中的元素时的结果。假设上述示意的被测模块及基准模块的处理日志中的字段类型均被配置为原生类型，对应原生类型的转化规则为直接将该字段的内容作为特征矩阵中的元素内容，则由上述被测模块的处理日志得到的特征矩阵可参考图2a，由上述基准模块的处理日志得到的特征矩阵可参考图2b。

[0052] 处理日志中的每个字段还可预先分别配置为不同的类型，对应每种类型，均有一种转化规则。本发明中可以采用的转化规则可参见表1：

[0053] 表1

[0054]

字段类型	字段转化规则
原生类型	直接使用原字段内容作为转化结果
枚举类型	根据设定的映射规则进行转化，如 A 转化为 1，B 转化为 2，或者区间（0,10）转化为 1，区间（11,20）转化为 2
整数类型	将原字段取值转化为整数
容量类型	计算原字段内容的字符串长度作为转化结果
签名类型	计算原字段内容的签名作为转化结果，如采用 MD5 算法计算原字段内容的签名
正则类型	根据正则表达式对原字段内容进行转化，如原字段内容符合正则表达式“[a-z]+”则转化为 1，符合正则表达式“[0-9]+”则转化为 2
自定义类型	由自定义的函数对原字段内容进行转化

[0055] 通过步骤S102，得到了两个特征矩阵，在步骤S103中，则会根据差值规则确定这两个特征矩阵之间的差值矩阵。

[0056] 差值矩阵中的每个元素内容是被测模块的特征矩阵和基准模块的特征矩阵对应位置的元素之间的差值。

[0057] 请参考表2，表2是本发明中可采用的差值规则说明：

[0058] 表2

[0059]

差值类型	差值规则
原生差值	直接将对应位置的两个元素拼接得到差值
异同差值	对应位置的两个元素相同则差值为 0 对应位置的两个元素不同则差值为 1
距离差值	将对应位置的两个元素之间的代数差作为差值
汉明差值	将对应位置的两个元素的二进制数按位对比时，存在差异的位数作为差值
自定义差值	由自定义的函数求得差值

[0060] 以图2a和图2b所示的两个矩阵为例，假设对矩阵第一列配置为原生差值，矩阵第二列配置为异同差值，矩阵第三列配置为汉明差值，矩阵第四列配置为距离差值，则图2a和

图2b中第一行的元素取差值如下：

[0061] 第一列元素差值为null|www.sina.com/index.html(由null和www.sina.com/index.html拼接而成)，第二列元素差值为1(因为A和C不同)，第三列元素差值为4(因为10的二进制数为00001010，20的二进制数为00010100，第4至第7位不同，不同的位数为4)，第四列元素差值为-10(因为30-40=-10)。

[0062] 以上举例说明了差值矩阵中一行元素的产生过程，根据类似的过程，差值矩阵可以得到若干行元素。请参考图3，图3为本发明中差值矩阵的示意图。

[0063] 步骤S 104中，对差值矩阵中的元素进行泛化，具体地包括：

[0064] 针对差值矩阵中的每个元素，查找预先配置的泛化规则表，当该泛化规则表中有该元素的适用规则时，将该元素按照适用规则进行泛化。

[0065] 泛化规则可用正则表达式表示，例如通过“[0-9a-zA-Z\./]->SOME_URL”这条正则表达式表示的泛化规则，可将图3的矩阵泛化为图4所示的形式。

[0066] 对差值矩阵进行泛化后，可能存在一些相同的行，由于对比测试中的测试数据数量巨大，因此差值矩阵可能包含千万行的数据，如果差值矩阵的列数也很多时，通过直接对每行的各个元素进行比较的方式合并相同行，耗费的计算资源及时间都是巨大的。

[0067] 下面对本发明步骤S104中合并相同行的方式进行介绍。具体地，将泛化后的差值矩阵中的相同行合并的步骤包括：

[0068] 步骤S1041：将泛化后的差值矩阵中的同行元素进行拼接。

[0069] 步骤S1042：对拼接后的各行分别计算该行的特征值。

[0070] 步骤S1043：对特征值相同的行进行合并。

[0071] 例如图4所示的差值矩阵的第一行，将各元素拼接后得到“null|SOME_URL14-10”，然后对该拼接串采用MD5算法求特征值，并将该特征值存入哈希表。可以理解，包含相同元素的行，其求得的特征值是相同的。依次对差值矩阵中的每行元素进行拼接并求特征值，且在将特征值存入哈希表时，确定表中是否已有该特征值，如果是，则将该特征值对应的当前行丢弃，从而实现对特征值相同的行进行快速合并的目的。

[0072] 将合并后的差值矩阵作为测试结果输出，并且在输出时，高亮显示差异元素(即由被测模块的特征矩阵与基准模块的特征矩阵在对应位置存在差异的元素得到的差值元素)，能够帮助测试人员快速确定引起被测模块及基准模块出现差异的样本数据及该样本数据的差异数据流(即一个维度的处理结果)，这样，测试人员就可以进一步对差异的数据流进行分析，以确定被测模块是否存在缺陷。

[0073] 请参考图5，图5为本发明中数据比对测试的装置的实施例的结构示意框图。如图5所示，该装置包括：日志获取单元201、转化单元202、差值获取单元203、泛化单元204及合并单元205。

[0074] 其中，日志获取单元201，用于将一条以上的样本数据分别发送至被测模块及基准模块，其中被测模块及基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志。

[0075] 一条样本数据是被测模块或基准模块完成一次处理过程所需要的基本数据单元。例如，被测模块或基准模块的功能是对页面进行分类，其基本数据单元是一个页面的URL地址。

[0076] 日志获取单元201发送至被测模块及基准模块的样本数据是相同的，即同一条样本数据会分别发送至被测模块及基准模块。被测模块对接收到的各条样本数据进行处理后，将输出自己的处理日志，基准模块对接收到的各条样本数据进行处理后，也会输出自己的处理日志。

[0077] 日志获取单元201输出的处理日志中，每条记录包含一条样本数据，以及由该样本数据得到的至少一个维度的处理结果。

[0078] 请参考下面的伪代码块：

[0079]

```
A（url）{
redir = process related to url;   //对 url 进行处理得到 redir
type = process related to url;   //对 url 进行处理得到 type
value = process related to url; //对 url 进行处理得到 value
weight = process related to url; //对 url 进行处理得到 weight
x = B（redir,type,value,weight）; //输出 redir、type、value、weight 作为
…                                  被调模块 B 的输入
…
}
```

[0080] 上述的模块A可以是被测模块或基准模块。

[0081] 与上述伪代码对应的，被测模块的处理日志请参考下面日志记录：

[0082] 第1条记录:url=www.sina.com,redir=null,type=A,value=10,weight=30

[0083] 第2条记录:…

[0084] 第3条记录:…

[0085] 与上述伪代码对应的，基准模块的处理日志请参考下面的日志记录：

[0086] 第1条记录：

[0087] url=www.sina.com,redir=www.sina.com/index.html,type=C,value=20,weight=40

[0088] 第2条记录:…

[0089] 第3条记录:…

[0090] 上述日志记录中，“url”字段对应记录的就是样本数据，“redir”、“type”、“value”、“weight”字段对应记录的就是一个维度的处理结果。

[0091] 转化单元202，用于根据预先配置的转化规则，将被测模块及基准模块各自的处理日志转化为各自的特征矩阵。

[0092] 具体地，特征矩阵中的每个元素，表示一条样本数据得到的一个维度的处理结果，并且，同一行的元素对应同一条样本数据，同一列的元素对应同一个维度的处理结果。

[0093] 在将处理日志中的字段转化为特征矩阵中的元素时，根据预先配置的字段类型，

查找与该类型对应的转化规则，即可得到该字段转化为特征矩阵中的元素时的结果。假设上述示意的被测模块及基准模块的处理日志中的字段类型均被配置为原生类型，对应原生类型的转化规则为直接将该字段的内容作为特征矩阵中的元素内容，则由上述被测模块的处理日志得到的特征矩阵可参考图2a，由上述基准模块的处理日志得到的特征矩阵可参考图2b。

[0094] 处理日志中的每个字段还可预先分别配置为不同的类型，对应每种类型，均有一种转化规则。本发明中可以采用的转化规则可参见表1。

[0095] 差值获取单元203，用于根据预先配置的差值规则，得到被测模块的特征矩阵与基准模块的特征矩阵之间的差值矩阵。

[0096] 差值矩阵中的每个元素内容是被测模块的特征矩阵和基准模块的特征矩阵对应位置的元素之间的差值。

[0097] 本发明中可采用的差值规则可参见表2。

[0098] 以图2a和图2b所示的两个矩阵为例，假设对矩阵第一列配置为原生差值，矩阵第二列配置为异同差值，矩阵第三列配置为汉明差值，矩阵第四列配置为距离差值，则图2a和图2b中第一行的元素取差值如下：

[0099] 第一列元素差值为null|www.sina.com/index.html(由null和www.sina.com/index.html拼接而成)，第二列元素差值为1(因为A和C不同)，第三列元素差值为4(因为10的二进制数为00001010，20的二进制数为00010100，第4至第7位不同，不同的位数为4)，第四列元素差值为-10(因为30-40=-10)。

[0100] 以上举例说明了差值矩阵中一行元素的产生过程，根据类似的过程，差值矩阵可以得到若干行元素。请参考图3，图3为本发明中差值矩阵的示意图。

[0101] 泛化单元204，用于对差值矩阵中的元素进行泛化。具体地，泛化单元204对差值矩阵中的元素进行泛化的方式包括：针对差值矩阵中的每个元素，查找预先配置的泛化规则表，当该泛化规则表中有该元素的适用规则时，将该元素按照适用规则进行泛化。

[0102] 合并单元205，用于将泛化后的差值矩阵中的相同行合并。对差值矩阵进行泛化后，可能存在一些相同的行，由于对比测试中的测试数据数量巨大，因此差值矩阵可能包含千万行的数据，如果差值矩阵的列数也很多时，通过直接对每行的各个元素进行比较的方式合并相同行，耗费的计算资源及时间都是巨大的。

[0103] 下面给出了合并单元205的一种实施方式。

[0104] 请参考图6，图6是本发明中合并单元205的实施例的结构示意框图。如图6所示，合并单元205包括：拼接单元2051、计算单元2052及行合并单元2053。其中拼接单元2051，用于将泛化后的差值矩阵中的同行元素进行拼接。计算单元2052，用于对拼接后的各行分别计算该行的特征值。行合并单元2053，用于对特征值相同的行进行合并。

[0105] 例如图4所示的差值矩阵的第一行，拼接单元2051将各元素拼接后得到“null|SOME_URL14-10”，然后计算单元2052对该拼接串采用MD5算法求特征值，并由行合并单元2053将该特征值存入哈希表。可以理解，包含相同元素的行，其求得的特征值是相同的。拼接单元2051和计算单元2052依次对差值矩阵中的每行元素进行拼接并求特征值，且在行合并单元2053将特征值存入哈希表时，确定表中是否已有该特征值，如果是，则将该特征值对应的当前行丢弃，从而实现对特征值相同的行进行快速合并的目的。

[0106] 以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，凡在本发明的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本发明保护的范围之内。

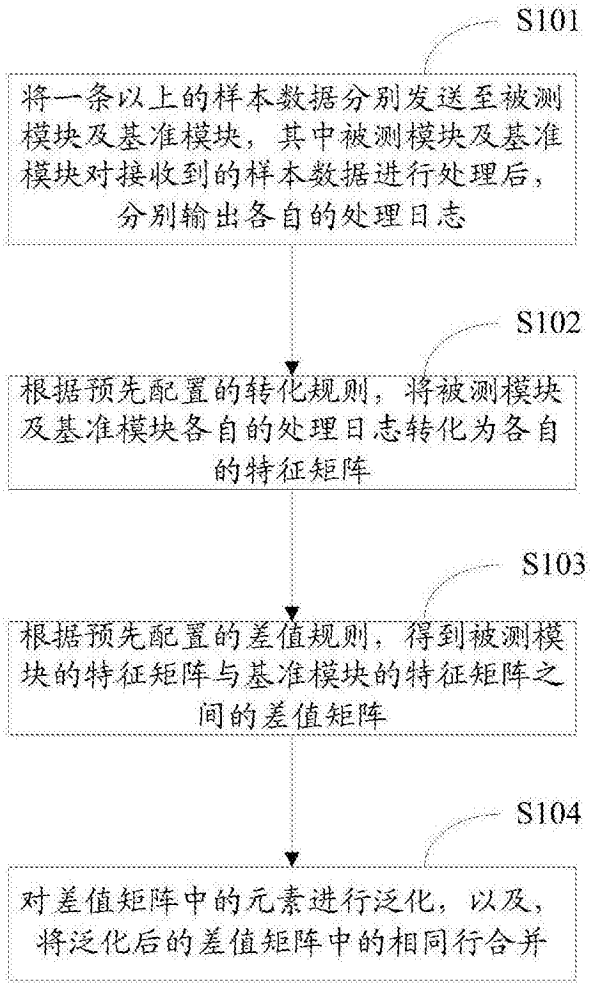

S101
将一条以上的样本数据分别发送至被测模块及基准模块，其中被测模块及基准模块对接收到的样本数据进行处理后，分别输出各自的处理日志
S102
根据预先配置的转化规则，将被测模块及基准模块各自的处理日志转化为各自的特征矩阵
S103
根据预先配置的差值规则，得到被测模块的特征矩阵与基准模块的特征矩阵之间的差值矩阵
S104
对差值矩阵中的元素进行泛化，以及，将泛化后的差值矩阵中的相同行合并

图1

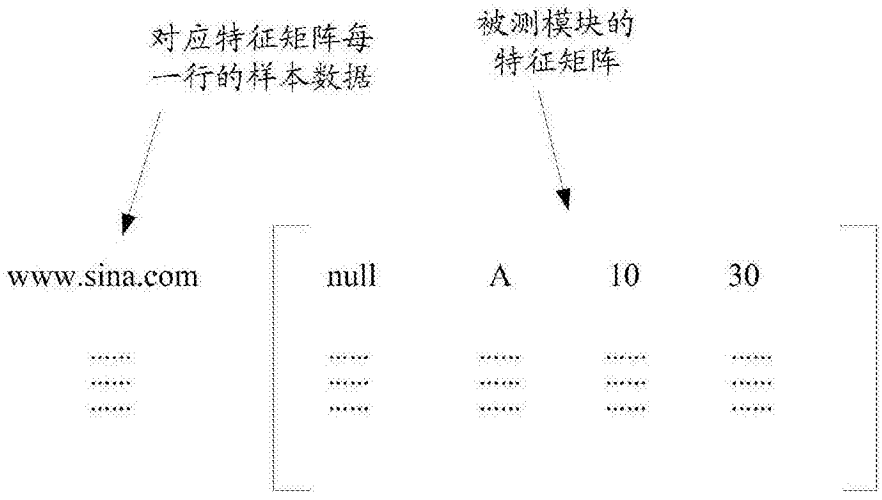

对应特征矩阵每一行的样本数据
被测模块的特征矩阵
www.sina.com
null
A
10
30

图2a


对应特征矩阵每一行的样本数据
基准模块的特征矩阵
www.sina.com
www.sina.com/index.html
C
20
40

图2b

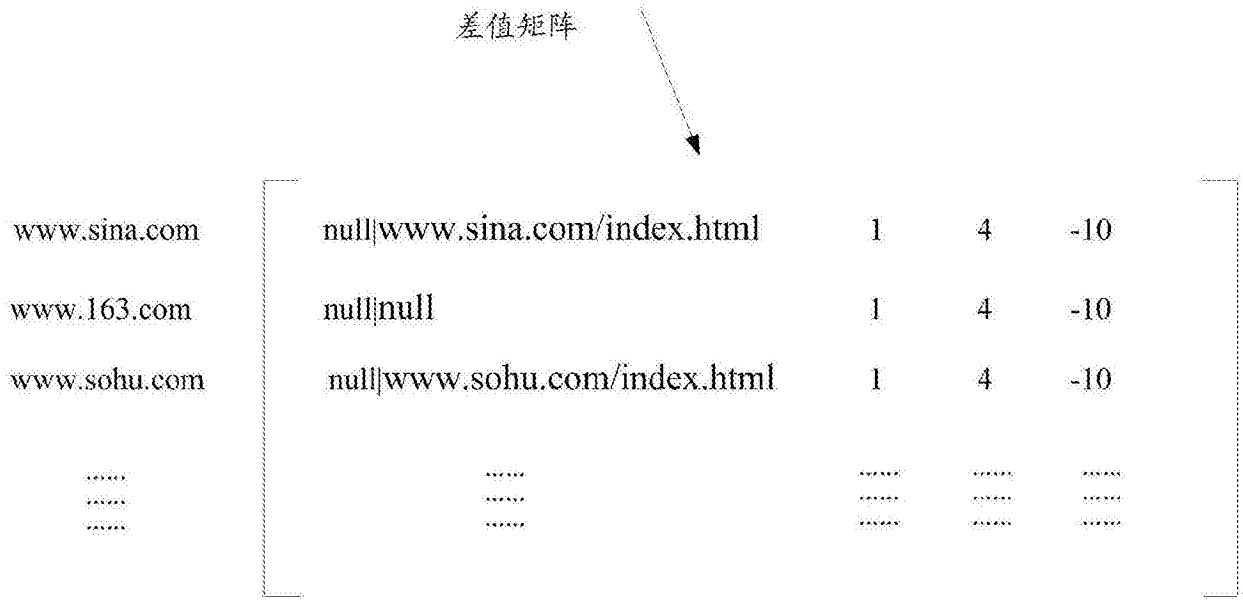

差值矩阵
www.sina.com
null|www.sina.com/index.html
1
4
-10
www.163.com
null|null
1
4
-10
www.sohu.com
null|www.sohu.com/index.html
1
4
-10

图3

泛化后的差值矩阵

www.sina.com	null|SOME_URL	1	4	-10
www.163.com	null|null	1	4	-10
www.sohu.com	null|SOME_URL	1	4	-10
.....	.....	.....	.....	.....

图4

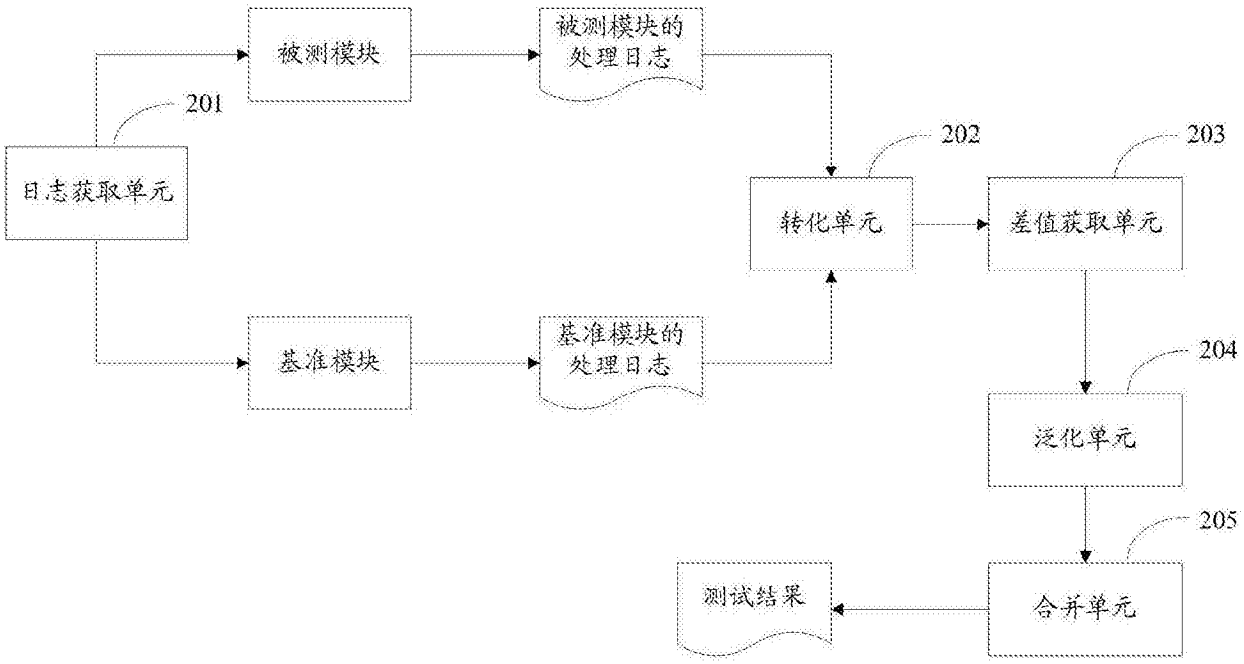


图5

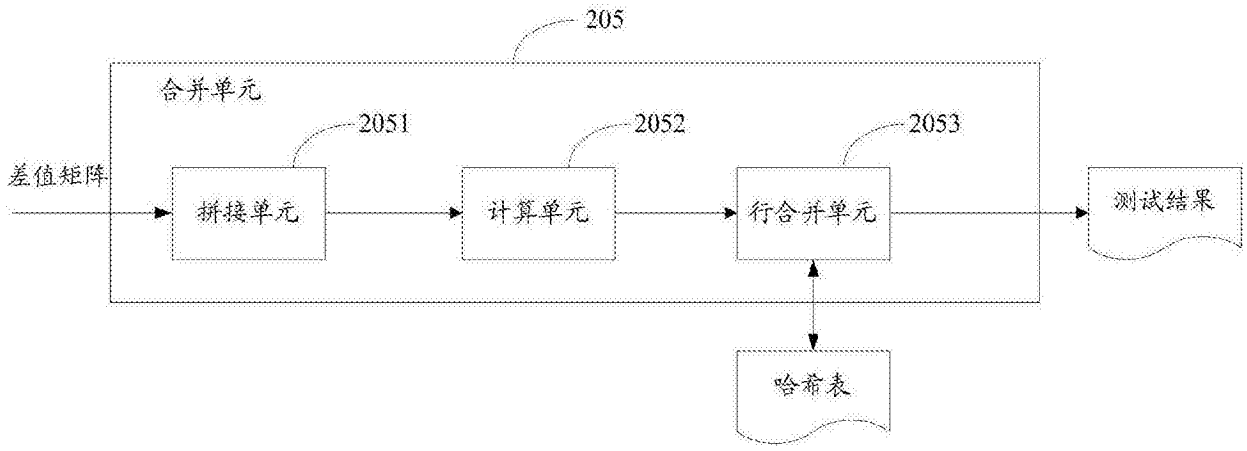


图6