



(12)发明专利申请

(10)申请公布号 CN 111708701 A

(43)申请公布日 2020.09.25

(21)申请号 202010558247.0

(51)Int.Cl.

(22)申请日 2020.06.18

G06F 11/36(2006.01)

(71)申请人 中国信息通信研究院

地址 100191 北京市海淀区花园北路52号

申请人 腾讯科技(深圳)有限公司

(72)发明人 王妙琼 郑智超 周文乐 魏凯

姜春宇 闫树 刘寒 马鹏玮

张奕卉 王卓 张启 李俊逸

李雨霏 和涛 田稼丰 刘虹男

高倩倩 刘思源 庞伟伟 袁博

钟民菁

(74)专利代理机构 北京康盛知识产权代理有限公司

公司 11331

代理人 张宇峰

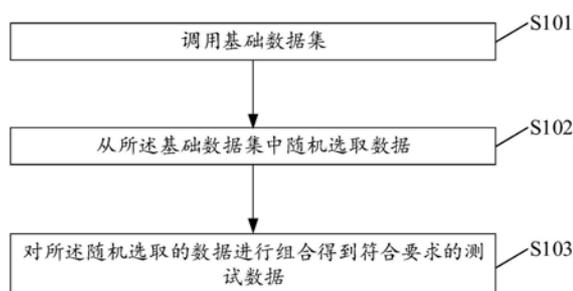
权利要求书1页 说明书7页 附图1页

(54)发明名称

用于数据测试的方法及装置

(57)摘要

本申请涉及数据库技术领域,公开了一种用于数据测试的方法,该方法包括:调用基础数据集;从所述基础数据集中随机选取数据;对所述随机选取的数据进行组合得到符合要求的测试数据。本公开实施例中,基础数据集中的数据为符合测试标准的数据,其中,数据数量小于实际需求的测试数据的数量,从基础数据集中随机选取数据并组合,保证组合后的数据也能够符合标准,且避免了数据存取数量不足或测试数据调用过程耗时长的问题,提高了测试效率。本申请还公开了一种用于数据测试的装置。



1. 一种用于数据测试的方法,其特征在于,包括:  
调用基础数据集;  
从所述基础数据集中随机选取数据;  
对所述随机选取的数据进行组合得到符合要求的测试数据。
2. 根据权利要求1所述的方法,其特征在于,调用基础数据集前,还包括:  
确定初始数据;  
对所述初始数据进行数据格式转换生成基础数据集;  
存储所述基础数据集。
3. 根据权利要求2所述的方法,其特征在于,所述确定初始数据包括:  
从实际生产环境中提取出设定数量的初始数据;或者,  
根据设定规则生成用于模拟实际情况的设定数量的初始数据。
4. 根据权利要求3所述的方法,其特征在于,所述设定数量根据数据生成要求确定。
5. 根据权利要求4所述的方法,其特征在于,所述数据生成要求包括:数据生成量和数据生成时间。
6. 根据权利要求5所述的方法,其特征在于,所述设定数量、所述数据生成量和所述数据生成时间满足如下关系:  
$$y = b \sqrt{\frac{T}{ad}}$$
  
其中,y为所述设定数量;d为数据生成量;T为所述数据生成时间;a和b为所述设定数量的调节系数。
7. 根据权利要求3所述的方法,其特征在于,所述设定规则包括:  
确定初始数据取值范围;  
根据扩展随机数函数在所述初始数据取值范围内生成初始数据。
8. 根据权利要求2所述的方法,其特征在于,对所述初始数据进行数据格式转换生成基础数据集,包括:  
将初始数据转换为整型变量。
9. 根据权利要求1所述的方法,其特征在于,对所述随机选取的数据进行组合得到符合要求的测试数据,包括:  
确定组合规则;  
根据组合规则对所述随机选取的数据进行分组并组合。
10. 一种用于数据测试的装置,包括处理器和存储有程序指令的存储器,其特征在于,所述处理器被配置为在执行所述程序指令时,执行如权利要求1至9任一项所述的用于数据测试的方法。

## 用于数据测试的方法及装置

### 技术领域

[0001] 本申请涉及数据库技术领域,例如涉及一种用于数据测试的方法及装置。

### 背景技术

[0002] 目前,对于绝大多数软件项目而言,测试都是非常重要的环节,测试成败直接决定了软件项目的成败,而在测试过程中需要大量的符合测试要求的数据。现有技术中生成测试数据的方案主要包括两类:第一,按照规则实时或在需要时才生成数据,可以满足生成速度要求不高,数据生成规则较为简单的情况;第二,事先准备好测试数据,并存储到特定的地方,可以满足生成速度要求较高,且数据使用者跟数据存放在同一个机器或者同一网络的情况。

[0003] 在实现本公开实施例的过程中,发现相关技术中至少存在如下问题:

[0004] 现有技术无法满足生成符合测试要求的海量数据的情况下对于数据生成速率的要求。

### 发明内容

[0005] 为了对披露的实施例的一些方面有基本的理解,下面给出了简单的概括。所述概括不是泛泛评述,也不是要确定关键/重要组成元素或描绘这些实施例的保护范围,而是作为后面的详细描述的序言。

[0006] 本公开实施例提供了一种用于数据测试的方法及装置,以解决现有技术无法满足生成符合测试要求的海量数据的情况下对于数据生成速率的要求技术问题。

[0007] 在一些实施例中,所述方法包括:

[0008] 调用基础数据集;

[0009] 从所述基础数据集中随机选取数据;

[0010] 对所述随机选取的数据进行组合得到符合要求的测试数据。

[0011] 在一些实施例中,所述装置包括:处理器和存储有程序指令的存储器,所述处理器被配置为在执行所述程序指令时,执行上述的用于数据测试的方法。

[0012] 本公开实施例提供的用于数据测试的方法及装置,可以实现以下技术效果:

[0013] 本公开实施例中,基础数据集中的数据为符合测试标准的数据,其中,数据数量小于实际需求的测试数据的数量,从基础数据集中随机选取数据并组合,保证组合后的数据也能够符合标准,且避免了数据存取数量不足或测试数据调用过程耗时过长的问题,提高了测试效率。

[0014] 以上的总体描述和下文中的描述仅是示例性和解释性的,不用于限制本申请。

### 附图说明

[0015] 一个或多个实施例通过与之对应的附图进行示例性说明,这些示例性说明和附图并不构成对实施例的限定,附图中具有相同参考数字标号的元件示为类似的元件,附图不

构成比例限制,并且其中:

[0016] 图1是本公开实施例提供的一个用于数据测试的方法的示意图;

[0017] 图2是本公开实施例提供的一个用于数据测试的装置的示意图。

### 具体实施方式

[0018] 为了能够更加详尽地了解本公开实施例的特点与技术内容,下面结合附图对本公开实施例的实现进行详细阐述,所附附图仅供参考说明之用,并非用来限定本公开实施例。在以下的技术描述中,为方便解释起见,通过多个细节以提供对所披露实施例的充分理解。然而,在没有这些细节的情况下,一个或多个实施例仍然可以实施。在其它情况下,为简化附图,熟知的结构和装置可以简化展示。

[0019] 本公开实施例的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开实施例的实施例。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含。

[0020] 除非另有说明,术语“多个”表示两个或两个以上。

[0021] 本公开实施例中,字符“/”表示前后对象是一种“或”的关系。例如,A/B表示:A或B。

[0022] 术语“和/或”是一种描述对象的关联关系,表示可以存在三种关系。例如,A和/或B,表示:A或B,或,A和B这三种关系。

[0023] 图1是本公开实施例提供的一个用于数据测试的方法的示意图,该方法包括如下步骤:

[0024] S101,调用基础数据集。

[0025] 其中,在实际的数据测试过程中,需要大量的测试数据,基础数据集中的数据量小于测试数据所需的数量,为生成大量的测试数据提供数据组合基础,减少数据生成时间和数据调用时间。基础数据集中的数据为符合测试标准的数据,根据基础数据集中的数据生成的测试数据同样符合测试标准,可以保证数据测试正常进行。

[0026] 在一些实施例中,调用基础数据集前,还包括:确定初始数据;对初始数据进行数据格式转换生成基础数据集;存储基础数据集。

[0027] 本申请中可以对不同领域的数据进行测试,领域不同对应的测试指标会有所不同,测试指标的取值对应的数据类型也不同。将初始数据进行数据格式转换生成基础数据集,使得数据格式统一,便于对数据的调用和测试。

[0028] 以对电动汽车数据进行测试为例,测试指标包括:车辆状态、速度和总电压,其中,速度和总电压对应的数据类型为整数,车辆状态对应的数据类型为字符串,车辆状态包括:启动,熄火,其他状态,异常和无效。

[0029] 在生成基础数据集后存储该基础数据集,可便于后续测试过程中进行重复调用。

[0030] 在一些实施例中,确定初始数据包括:从实际生产环境中提取出设定数量的初始数据。从实际生产环境中提取数据,数据必然满足测试标准,无需对数据有效性进行校验。

[0031] 在一些实施例中,确定初始数据包括:根据设定规则生成用于模拟实际情况的设定数量的初始数据。根据设定规则生成数据,保证生成后的数据即满足测试标准,无需对数据有效性进行校验。

[0032] 在一些实施例中,设定数量根据数据生成要求确定。

[0033] 在一些实施例中,数据生成要求包括:数据生成量和数据生成时间。根据数据生成要求确定基础数据集中数据量,保证在有限的时间内生成满足数据生成量需求,数据测试能够顺利完成。

[0034] 在一些实施例中,设定数量、数据生成量和数据生成时间满足如下关系:

$$[0035] \quad y = b \sqrt[b]{\frac{T}{ad}}$$

[0036] 其中, $y$ 为设定数量; $d$ 为数据生成量; $T$ 为数据生成时间; $a$ 和 $b$ 为设定数量的调节系数。可选的, $a < 1$ , $b > 2$ 。 $a$ 和 $b$ 根据实际机器性能进行调整。

[0037] 根据上述公式可知,数据生成量为 $d$ ,数据生成时间为 $T$ ,则单个数据生成时间如下:

$$[0038] \quad t = \frac{T}{d} = ay^b$$

[0039] 其中, $t$ 为单个数据生成时间; $T$ 为数据生成时间; $d$ 为数据生成量; $y$ 为设定数量; $a$ 和 $b$ 为设定数量的调节系数。

[0040] 如上可知,单个数据的生成时间 $t$ 与设定数量 $y$ 之间的关系是呈斜率小于1且过原点的单调递增抛物线。根据上述关系系确定出设定数量,并从实际生产环境中提取出设定数量的初始数据能够满足数据生成需求,以在数据生成时间 $T$ 内成功生成 $d$ 个数据满足数据生成量需求。

[0041] 在一具体实施例中,对电动汽车数据进行测试,测试指标包括:车辆状态、速度和总电压。数据生成要求为在0.1S内生成120万个指标数据,即数据生成量为120万,生成时间为0.1S。按照本发明方法,先针对每个测试指标获取100个数据作为初始数据,可以保证在0.1S内生成120万个指标数据以完成测试。具体的,可以从实际生产环境中针对每个测试指标提取出100个初始数据,或者根据设定规则针对每个测试指生成100个初始数据。

[0042] 在一些实施例中,设定规则包括:确定初始数据取值范围;根据扩展随机数函数在初始数据取值范围内生成初始数据。

[0043] 可选的,扩展随机数函数为计算机程序设计语言中random库下的扩展随机数函数randint()。

[0044] 在一具体实施例中,初始数据取值范围为 $[0, p]$ ,根据扩展随机数函数randint()进行取值时,初始数据值取值如下:

$$[0045] \quad X_n = \text{randint}\left(\left(n-1\right)\frac{p}{m}, n\frac{p}{m}\right)$$

[0046] 其中, $X_n$ 为第 $n$ 个生成的初始数据; $m$ 为设定数量; $P$ 为初始数据取值范围的上限值。

[0047] 在一些实施例中,对初始数据进行数据格式转换生成基础数据集,包括:将初始数据转换为整型变量。将所有的初始数据转换为整型变量,使得数据格式统一,便于对数据的调用和测试。

[0048] 在不同实施例中,将初始数据转换为整型变量的形式有多种。

[0049] 在一些实施例中,使用正整数表示测试数据中可穷举的状态。以对电动汽车数据

进行测试为例,对车辆状态进行格式转换。可选的,整数“1”指代启动状态,整数“2”指代熄火状态,整数“3”指代其他状态,整数“4”指代异常状态,整数“5”指代无效状态。可选的,使用八进制整数或十六进制整数表示测试数据中可穷举的状态。

[0050] 在一些实施例中,使用正整数表示测试数据中包含小数部分的数据。以对电动汽车数据进行测试为例,当车速的有效范围为0km/h~220km/h,最小计量单位为0.1km/h时。可选的,使用十进制整数0~2200来表示所有速度值。1~10指代0.1km/h~1km/h,2200指代220km/h。

[0051] 当车速的有效范围为0km/h~220km/h,最小计量单位为0.01km/h时。可选的,使用十进制整数0~22000来表示所有速度值。1~10指代0.01km/h~0.1km/h,100指代1km/h,2200指代22km/h,22000指代220km/h。

[0052] 可选的,使用八进制整数或十六进制整数表示测试数据中包含小数部分的数据。

[0053] 在一些实施例中,使用正整数表示测试数据中的负数。以对电动汽车数据进行测试为例,当电机控制器温度取值范围在-40℃~210℃之间,最小计量单位为1℃时。可选的,使用十进制整数0~250来表示所有温度值。其中,0指代-40℃,1指代-39℃,250指代210℃。

[0054] 当电机控制器温度取值范围在-40℃~210℃之间,最小计量单位为0.1℃时。可选的,使用十进制整数0~2500来表示所有温度值。其中,0指代-40℃,1指代-39.9℃,10指代-39℃,2500指代210℃。

[0055] 可选的,使用八进制整数或十六进制整数表示测试数据中表示测试数据中的负数。

[0056] S102,从基础数据集中随机选取数据。基础数据集中的数据符合测试标准的数据,从中选取数据使得生成的测试数据符合测试标准,有利于数据测试的顺利进行。

[0057] 在不同实施例中,从基础数据集中随机选取数据的方式不同。

[0058] 在一些实施例中,从基础数据集中随机选取数据,包括:确定基础数据集中包含的指标类型和每类指标对应的取值范围;根据取值范围进行平均取值。

[0059] 在一些实施例中,从基础数据集中随机选取数据,包括:确定基础数据集中包含的指标类型和每类指标对应的取值范围;根据取值范围和设定函数进行取值。

[0060] 根据前述选取策略进行数据选取,保证选取出的数据进行组合后可以得到不同的测试数据,提高测试过程中样本数据的多样性。

[0061] 以对电动汽车数据进行测试为例,测试指标包括:车辆状态、速度和总电压,车辆状态、速度和总电压对应的取值范围不同,对车辆状态、速度和总电压进行单独选取,保证选取出不同的车辆状态、速度和总电压进行组合。

[0062] S103,对随机选取的数据进行组合得到符合要求的测试数据。

[0063] 在一些实施例中,对随机选取的数据进行组合得到符合要求的测试数据,包括:确定组合规则;根据组合规则对随机选取的数据进行分组并组合。

[0064] 在一些实施例中,各组测试数据之间没有关联关系,则对选取出的数据进行随机组合。

[0065] 以对电动汽车数据进行测试为例,测试指标包括车速和累积里程,两者之间无直接关系,因此,在对数据进行组合时,可以对两者进行任意组合。

[0066] 在一些实施例中,各组测试数据之间具有关联关系,则对选取出的数据根据该关

联关系进行组合。

[0067] 以对电动汽车数据进行测试为例,测试指标包括车速和车辆状态,当“车辆状态”为“熄火”时,“车速”只能为0,当组合后的测试数据中包含“车辆状态”为“熄火”但车速不为0的数据时,则该测试数据为无效数据。因此,依据关联关系对相关的数据进行组合提高了测试数据的有效性。

[0068] 本公开实施例中,基础数据集中的数据为符合测试标准的数据,其中,数据数量小于实际需求的测试数据的数量,从基础数据集中随机选取数据并组合,保证组合后的数据也能够符合标准,且避免了数据存取数量不足或测试数据调用过程耗时长的问题,提高了测试效率。

[0069] 本公开实施例还提供一种用于数据测试的装置,包括处理器和存储有程序指令的存储器,处理器被配置为在执行程序指令时,执行上述实施例提供的用于数据测试的方法。

[0070] 结合图2所示,本公开实施例提供一种用于数据测试的装置,包括处理器(processor)200和存储器(memory)201。可选地,该装置还可以包括通信接口(Communication Interface)202和总线203。其中,处理器200、通信接口202、存储器201可以通过总线203完成相互间的通信。通信接口202可以用于信息传输。处理器200可以调用存储器201中的逻辑指令,以执行上述实施例的用于数据测试的方法。

[0071] 此外,上述的存储器201中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。

[0072] 存储器201作为一种计算机可读存储介质,可用于存储软件程序、计算机可执行程序,如本公开实施例中的方法对应的程序指令/模块。处理器200通过运行存储在存储器201中的程序指令/模块,从而执行功能应用以及数据处理,即实现上述实施例中用于数据测试的方法。

[0073] 存储器201可包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序;存储数据区可存储根据终端设备的使用所创建的数据等。此外,存储器201可以包括高速随机存取存储器,还可以包括非易失性存储器。

[0074] 本公开实施例提供了一种计算机可读存储介质,存储有计算机可执行指令,所述计算机可执行指令设置为执行上述用于数据测试的方法。

[0075] 本公开实施例提供了一种计算机程序产品,所述计算机程序产品包括存储在计算机可读存储介质上的计算机程序,所述计算机程序包括程序指令,当所述程序指令被计算机执行时,使所述计算机执行上述用于数据测试的方法。

[0076] 上述的计算机可读存储介质可以是暂态计算机可读存储介质,也可以是非暂态计算机可读存储介质。

[0077] 本公开实施例的技术方案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一个存储介质中,包括一个或多个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)执行本公开实施例所述方法的全部或部分步骤。而前述的存储介质可以是非暂态存储介质,包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等多种可以存储程序代码的介质,也可以是暂态存储介质。

[0078] 以上描述和附图充分地示出了本公开的实施例,以使本领域的技术人员能够实践

它们。其他实施例可以包括结构的、逻辑的、电气的、过程的以及其他的改变。实施例仅代表可能的变化。除非明确要求,否则单独的部件和功能是可选的,并且操作的顺序可以变化。一些实施例的部分和特征可以被包括在或替换其他实施例的部分和特征。而且,本申请中使用的用词仅用于描述实施例并且不用于限制权利要求。如在实施例以及权利要求的描述中使用的,除非上下文清楚地表明,否则单数形式的“一个”(a)、“一个”(an)和“所述”(the)旨在同样包括复数形式。类似地,如在本申请中所使用的术语“和/或”是指包含一个或一个以上相关联的列出的任何以及所有可能的组合。另外,当用于本申请中时,术语“包括”(comprise)及其变型“包括”(comprises)和/或包括(comprising)等指陈述的特征、整体、步骤、操作、元素,和/或组件的存在,但不排除一个或一个以上其它特征、整体、步骤、操作、元素、组件和/或这些的分组的存在或添加。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、方法或者设备中还存在另外的相同要素。本文中,每个实施例重点说明的可以是与其他实施例的不同之处,各个实施例之间相同相似部分可以互相参见。对于实施例公开的方法、产品等而言,如果其与实施例公开的方法部分相对应,那么相关之处可以参见方法部分的描述。

[0079] 本领域技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,可以取决于技术方案的特定应用和设计约束条件。所述技术人员可以对每个特定的应用来使用不同方法以实现所描述的功能,但是这种实现不应认为超出本公开实施例的范围。所述技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0080] 本文所披露的实施例中,所揭露的方法、产品(包括但不限于装置、设备等),可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,可以仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例。另外,在本公开实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0081] 附图中的流程图和框图显示了根据本公开实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这可以依所涉及的功能而定。在附图中的流程图和框图所对应的描述中,不同的方框所对应的操作或步骤也可以以不同于描述中所披露的顺序发

生,有时不同的操作或步骤之间不存在特定的顺序。例如,两个连续的操作或步骤实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这可以依所涉及的功能而定。框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

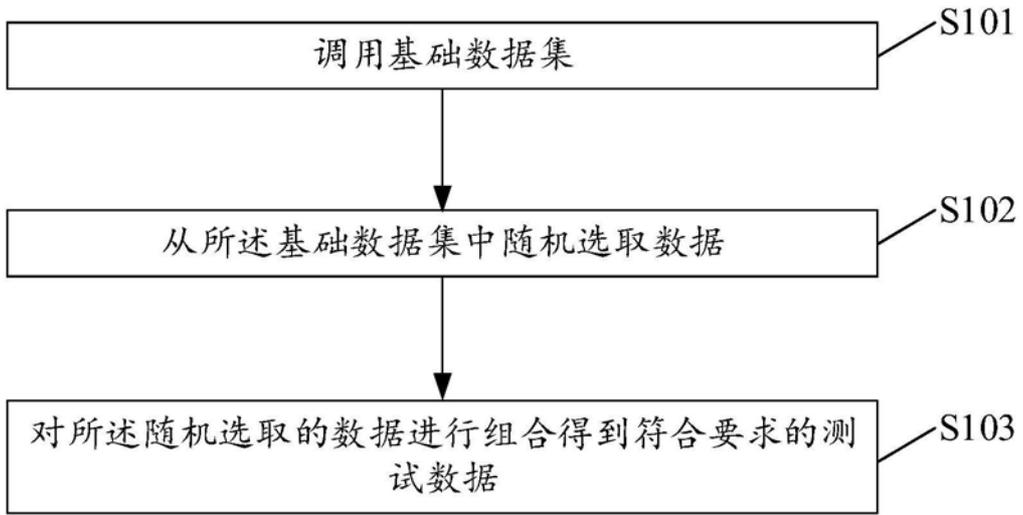


图1

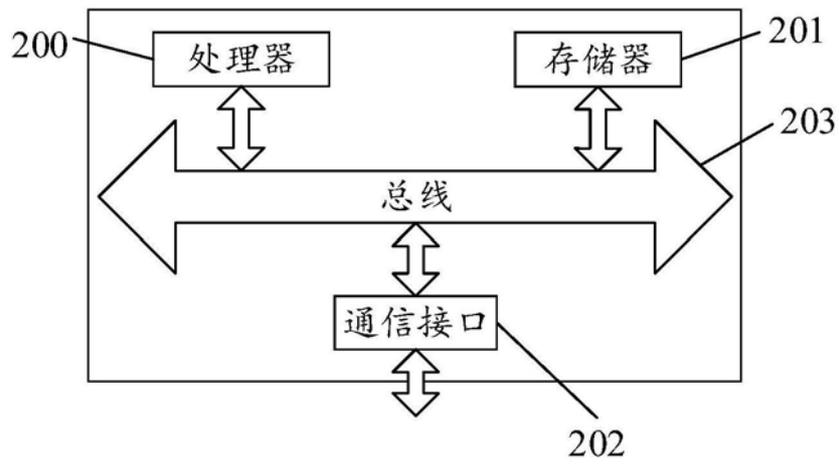


图2