(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115062705 A
(43)申请公布日 2022.09.16

(21)申请号 202210689220.4

(22)申请日 2022.06.17

(71)申请人 中国平安人寿保险股份有限公司
地址 518000 广东省深圳市福田区福田街道益田路5033号平安金融中心14、15、16、37、41、44、45、46、54、58、59层

(72)发明人 雷志勇

(74)专利代理机构 广州嘉权专利商标事务所有限公司 44205
专利代理师 廖慧贤

(51)Int.Cl.
G06K 9/62(2022.01)
G06F 17/18(2006.01)

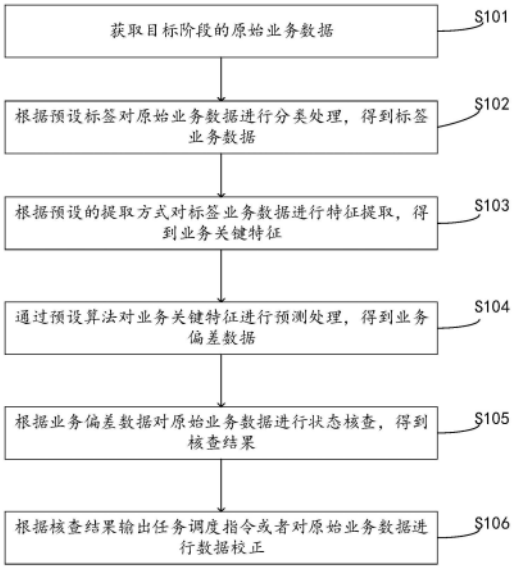
权利要求书2页 说明书14页 附图5页

(54)发明名称

数据核查方法、数据核查装置、电子设备及存储介质

(57)摘要

本申请实施例提供了一种数据核查方法、数据核查装置、电子设备及存储介质，属于人工智能技术领域。该方法包括：获取目标阶段的原始业务数据；根据预设标签对原始业务数据进行分类处理，得到标签业务数据；根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；通过预设算法对业务关键特征进行预测处理，得到业务偏差数据；根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；根据核查结果输出任务调度指令或者对原始业务数据进行数据校正。本申请实施例能够提高数据核查效率。



CN 115062705 A

1.一种数据核查方法,其特征在于,所述方法包括:

获取目标阶段的原始业务数据;

根据预设标签对所述原始业务数据进行分类处理,得到标签业务数据;

根据预设的提取方式对所述标签业务数据进行特征提取,得到业务关键特征;

通过预设算法对所述业务关键特征进行预测处理,得到业务偏差数据;

根据所述业务偏差数据对所述原始业务数据进行状态核查,得到核查结果;

根据所述核查结果输出任务调度指令或者对所述原始业务数据进行数据校正。

2.根据权利要求1所述的数据核查方法,其特征在于,所述根据预设标签对所述原始业务数据进行分类处理,得到标签业务数据的步骤,包括:

获取所述目标阶段的属性数据;

根据所述属性数据对预设标签库中的预设标签进行筛选处理,得到目标标签;

通过预设函数和所述目标标签对所述原始业务数据进行标签概率计算,得到标签概率值;

根据所述标签概率值对所述原始业务数据进行分类处理,得到所述标签业务数据。

3.根据权利要求1所述的数据核查方法,其特征在于,所述目标阶段包括第一阶段,所述提取方式包括第一方式,所述根据预设的提取方式对所述标签业务数据进行特征提取,得到业务关键特征的步骤,包括:

若所述标签业务数据为第一阶段的业务数据,则获取所述第一方式对应的数据类别词段;

根据所述数据类别词段对所述标签业务数据进行关键词提取,得到初始业务关键词;

对所述初始业务关键词进行过滤处理,得到所述业务关键特征。

4.根据权利要求1所述的数据核查方法,其特征在于,所述目标阶段包括第二阶段,所述提取方式包括第二方式,所述根据预设的提取方式对所述标签业务数据进行特征提取,得到业务关键特征的步骤,包括:

若所述标签业务数据为第二阶段的业务数据,则根据预设的数据维度,获取所述第二方式对应的算法配置表,所述算法配置表存储有预设标签和预设算法,所述预设标签和预设算法之间具有映射关系;

从所述预设标签提取出所述标签业务数据对应的目标标签;

根据所述映射关系,从所述算法配置表提取出所述目标标签对应的预设算法;

通过所述预设算法对所述标签业务数据进行字段提取,得到初始关键字段;

对所述初始关键字段进行字段值计算,得到所述业务关键特征。

5.根据权利要求1所述的数据核查方法,其特征在于,所述通过预设算法对所述业务关键特征进行预测处理,得到业务偏差数据的步骤,包括:

通过最小二乘法对所述业务关键特征进行回归预测,得到初始预测数据;

根据预设的偏差参数对所述初始预测数据进行校准处理,得到所述业务偏差数据。

6.根据权利要求1所述的数据核查方法,其特征在于,所述核查结果包括第一结果和第二结果,所述根据所述业务偏差数据对所述原始业务数据进行状态核查,得到核查结果的步骤,包括:

若所述业务偏差数据处于预设的偏差范围之内,则确定所述原始业务数据为正常状

态，输出第一结果；

若所述业务偏差数据处于预设的偏差范围之外，则确定所述原始业务数据为异常状态，输出第二结果。

7.根据权利要求1至6任一项所述的数据核查方法，其特征在于，所述核查结果包括第一结果和第二结果，所述根据所述核查结果输出任务调度指令或者对所述原始业务数据进行数据校正的步骤，包括：

若所述核查结果为第一结果，则输出任务调度指令，以根据所述任务调度指令进行业务数据的核查；

若所述核查结果为第二结果，则对所述原始业务数据进行异常定位，生成异常信息，并根据所述异常信息对所述原始业务数据进行数据校正。

8.一种数据核查装置，其特征在于，所述装置包括：

数据获取模块，用于获取目标阶段的原始业务数据；

分类模块，用于根据预设标签对所述原始业务数据进行分类处理，得到标签业务数据；

特征提取模块，用于根据预设的提取方式对所述标签业务数据进行特征提取，得到业务关键特征；

预测模块，用于通过预设算法对所述业务关键特征进行预测处理，得到业务偏差数据；

核查模块，用于根据所述业务偏差数据对所述原始业务数据进行状态核查，得到核查结果；

处理模块，用于根据所述核查结果输出任务调度指令或者对所述原始业务数据进行数据校正。

9.一种电子设备，其特征在于，所述电子设备包括存储器、处理器、存储在所述存储器上并可在所述处理器上运行的程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线，所述程序被所述处理器执行时实现如权利要求1至7任一项所述的数据核查方法的步骤。

10.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有一个或者多个计算机程序，所述一个或者多个计算机程序可被一个或者多个处理器执行，以实现权利要求1至7中任一项所述的数据核查方法的步骤。

数据核查方法、数据核查装置、电子设备及存储介质

技术领域

[0001] 本申请涉及人工智能技术领域，尤其涉及一种数据核查方法、数据核查装置、电子设备及存储介质。

背景技术

[0002] 目前，在数据核查时，常常是采用人工核查的方式，而这一方式往往需要人工对每一阶段的数据进行一一核查，工作量较大，影响数据核查的效率，因此，如何提高数据核查的效率，成为了亟待解决的技术问题。

发明内容

[0003] 本申请实施例的主要目的在于提出一种数据核查方法、数据核查装置、电子设备及存储介质，旨在提高数据核查的效率。

[0004] 为实现上述目的，本申请实施例的第一方面提出了一种数据核查方法，所述方法包括：

[0005] 获取目标阶段的原始业务数据；

[0006] 根据预设标签对所述原始业务数据进行分类处理，得到标签业务数据；

[0007] 根据预设的提取方式对所述标签业务数据进行特征提取，得到业务关键特征；

[0008] 通过预设算法对所述业务关键特征进行预测处理，得到业务偏差数据；

[0009] 根据所述业务偏差数据对所述原始业务数据进行状态核查，得到核查结果；

[0010] 根据所述核查结果输出任务调度指令或者对所述原始业务数据进行数据校正。

[0011] 在一些实施例，所述根据预设标签对所述原始业务数据进行分类处理，得到标签业务数据的步骤，包括：

[0012] 获取所述目标阶段的属性数据；

[0013] 根据所述属性数据对预设标签库中的预设标签进行筛选处理，得到目标标签；

[0014] 通过预设函数和所述目标标签对所述原始业务数据进行标签概率计算，得到标签概率值；

[0015] 根据所述标签概率值对所述原始业务数据进行分类处理，得到所述标签业务数据。

[0016] 在一些实施例，所述目标阶段包括第一阶段，所述提取方式包括第一方式，所述根据预设的提取方式对所述标签业务数据进行特征提取，得到业务关键特征的步骤，包括：

[0017] 若所述标签业务数据为第一阶段的业务数据，则获取所述第一方式对应的数据类别词段；

[0018] 根据所述数据类别词段对所述标签业务数据进行关键词提取，得到初始业务关键词；

[0019] 对所述初始业务关键词进行过滤处理，得到所述业务关键特征。

[0020] 在一些实施例，所述目标阶段包括第二阶段，所述提取方式包括第二方式，所述根

据预设的提取方式对所述标签业务数据进行特征提取，得到业务关键特征的步骤，包括：

[0021] 若所述标签业务数据为第二阶段的业务数据，则根据预设的数据维度，获取所述第二方式对应的算法配置表，所述算法配置表存储有预设标签和预设算法，所述预设标签和预设算法之间具有映射关系；

[0022] 从所述预设标签提取出所述标签业务数据对应的目标标签；

[0023] 根据所述映射关系，从所述算法配置表提取出所述目标标签对应的预设算法；

[0024] 通过所述预设算法对所述标签业务数据进行字段提取，得到初始关键字段；

[0025] 对所述初始关键字段进行字段值计算，得到所述业务关键特征。

[0026] 在一些实施例，所述通过预设算法对所述业务关键特征进行预测处理，得到业务偏差数据的步骤，包括：

[0027] 通过最小二乘法对所述业务关键特征进行回归预测，得到初始预测数据；

[0028] 根据预设的偏差参数对所述初始预测数据进行校准处理，得到所述业务偏差数据。

[0029] 在一些实施例，所述核查结果包括第一结果和第二结果，所述根据所述业务偏差数据对所述原始业务数据进行状态核查，得到核查结果的步骤，包括：

[0030] 若所述业务偏差数据处于预设的偏差范围之内，则确定所述原始业务数据为正常状态，输出第一结果；

[0031] 若所述业务偏差数据处于预设的偏差范围之外，则确定所述原始业务数据为异常状态，输出第二结果。

[0032] 在一些实施例，所述核查结果包括第一结果和第二结果，所述根据所述核查结果输出任务调度指令或者对所述原始业务数据进行数据校正的步骤，包括：

[0033] 若所述核查结果为第一结果，则输出任务调度指令，以根据所述任务调度指令进行业务数据的核查；

[0034] 若所述核查结果为第二结果，则对所述原始业务数据进行异常定位，生成异常信息，并根据所述异常信息对所述原始业务数据进行数据校正。

[0035] 为实现上述目的，本申请实施例的第二方面提出了一种数据核查装置，所述装置包括：

[0036] 数据获取模块，用于获取目标阶段的原始业务数据；

[0037] 分类模块，用于根据预设标签对所述原始业务数据进行分类处理，得到标签业务数据；

[0038] 特征提取模块，用于根据预设的提取方式对所述标签业务数据进行特征提取，得到业务关键特征；

[0039] 预测模块，用于通过预设算法对所述业务关键特征进行预测处理，得到业务偏差数据；

[0040] 核查模块，用于根据所述业务偏差数据对所述原始业务数据进行状态核查，得到核查结果；

[0041] 处理模块，用于根据所述核查结果输出任务调度指令或者对所述原始业务数据进行数据校正。

[0042] 为实现上述目的，本申请实施例的第三方面提出了一种电子设备，所述电子设备

包括存储器、处理器、存储在所述存储器上并可在所述处理器上运行的程序以及用于实现所述处理器和所述存储器之间的连接通信的数据总线，所述程序被所述处理器执行时实现上述第一方面所述的方法。

[0043] 为实现上述目的，本申请实施例的第四方面提出了一种计算机可读存储介质，所述计算机可读存储介质存储有一个或者多个计算机程序，所述一个或者多个计算机程序可被一个或者多个处理器执行，以实现第一方面的数据核查方法的步骤

[0044] 本申请提出的数据核查方法、数据核查装置、电子设备及存储介质，其通过获取目标阶段的原始业务数据；根据预设标签对原始业务数据进行分类处理，得到标签业务数据，能够根据业务需求，获取不同目标阶段的业务数据，从而在不同业务阶段进行数据核查，提高全流程的数据准确性。进一步地，根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；通过预设算法对业务关键特征进行预测处理，得到业务偏差数据，能够对不同目标阶段的业务数据采取不同的提取方式进行提取，从而使得获取到的业务关键特征更为准确。相较于人工核查的方式，本申请实施例通过预设算法进行数据预测，能够有效地缩短数据处理的时间，并且能够减小业务偏差数据的计算误差，从而提高数据核查的效率与核查精度。最后，根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；根据核查结果输出任务调度指令或者对原始业务数据进行数据校正，能够根据业务偏差数据分析出原始业务数据处于正常状态还是异常状态，从而根据核查结果及时地采取相应的措施，从而提高数据处理的时效性。

附图说明

[0045] 图1是本申请实施例提供的数据核查方法的流程图；

[0046] 图2是图1中的步骤S102的流程图；

[0047] 图3是图1中的步骤S103的流程图；

[0048] 图4是图1中的步骤S103的另一流程图；

[0049] 图5是图1中的步骤S104的流程图；

[0050] 图6是图1中的步骤S105的流程图；

[0051] 图7是图1中的步骤S106的流程图；

[0052] 图8是本申请实施例提供的数据核查装置的结构示意图；

[0053] 图9是本申请实施例提供的电子设备的硬件结构示意图。

具体实施方式

[0054] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请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用以解释本申请，并不用于限定本申请。

[0055] 需要说明的是，虽然在装置示意图中进行了功能模块划分，在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于装置中的模块划分，或流程图中的顺序执行所示出或描述的步骤。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。

[0056] 除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的

技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的，不是旨在限制本申请。

[0057] 首先，对本申请中涉及的若干名词进行解析：

[0058] 人工智能(artificial intelligence，AI)：是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学；人工智能是计算机科学的一个分支，人工智能企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能可以对人的意识、思维的信息过程的模拟。人工智能还是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0059] 自然语言处理(natural language processing，NLP)：NLP用计算机来处理、理解以及运用人类语言(如中文、英文等)，NLP属于人工智能的一个分支，是计算机科学与语言学的交叉学科，又常被称为计算语言学。自然语言处理包括语法分析、语义分析、篇章理解等。自然语言处理常用于机器翻译、手写体和印刷体字符识别、语音识别及文语转换、信息意图识别、信息抽取与过滤、文本分类与聚类、舆情分析和观点挖掘等技术领域，它涉及与语言处理相关的数据挖掘、机器学习、知识获取、知识工程、人工智能研究和与语言计算相关的语言学研究等。

[0060] 信息抽取(Information Extraction)：从自然语言文本中抽取指定类型的实体、关系、事件等事实信息，并形成结构化数据输出的文本处理技术。信息抽取是从文本数据中抽取特定信息的一种技术。文本数据是由一些具体的单位构成的，例如句子、段落、篇章，文本信息正是由一些小的具体的单位构成的，例如字、词、词组、句子、段落或是这些具体的单位的组合。抽取文本数据中的名词短语、人名、地名等都是文本信息抽取，当然，文本信息抽取技术所抽取的信息可以是各种类型的信息。

[0061] MP(Mass-Production)：量产。

[0062] 最小二乘法((generalized least squares))：又称最小平方法，是一种数学优化技术。它通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。最小二乘法还可用于曲线拟合。其他一些优化问题也可通过最小化能量或者最大化熵的最小二乘法来表达。

[0063] 数据拟合：又称曲线拟合，俗称拉曲线，是一种把现有数据透过数学方法来代入一条数式的表示方式。科学和工程问题可以通过诸如采样、实验等方法获得若干离散的数据，根据这些数据，我们往往希望得到一个连续的函数(也就是曲线)或者更加密集的离散方程与已知数据相吻合，这过程就叫做拟合(fitting)。

[0064] 目前，对在保险业务场景的数据进行核查时，常常是采用人工核查的方式，而这一方式往往需要人工对每一阶段的数据进行一一核查，工作量较大，影响数据核查的效率，因此，如何提高数据核查的效率，成为了亟待解决的技术问题。

[0065] 基于此，本申请实施例提供了一种数据核查方法、数据核查装置、电子设备及存储介质，旨在提高数据核查的效率。

[0066] 本申请实施例提供的数据核查方法、数据核查装置、电子设备及存储介质，具体通

过如下实施例进行说明，首先描述本申请实施例中的数据核查方法。

[0067] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中，人工智能(Artificial Intelligence，AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0068] 人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0069] 本申请实施例提供的数据核查方法，涉及人工智能技术领域。本申请实施例提供的数据核查方法可应用于终端中，也可应用于服务器端中，还可以是运行于终端或服务器端中的软件。在一些实施例中，终端可以是智能手机、平板电脑、笔记本电脑、台式计算机等；服务器端可以配置成独立的物理服务器，也可以配置成多个物理服务器构成的服务器集群或者分布式系统，还可以配置成提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN以及大数据和人工智能平台等基础云计算服务的云服务器；软件可以是实现数据核查方法的应用等，但并不局限于以上形式。

[0070] 本申请可用于众多通用或专用的计算机系统环境或配置中。例如：个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的消费电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。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，由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的计算机可读存储介质中。

[0071] 图1是本申请实施例提供的数据核查方法的一个可选的流程图，图1中的方法可以包括但不限于包括步骤S101至步骤S106。

[0072] 步骤S101，获取目标阶段的原始业务数据；

[0073] 步骤S102，根据预设标签对原始业务数据进行分类处理，得到标签业务数据；

[0074] 步骤S103，根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；

[0075] 步骤S104，通过预设算法对业务关键特征进行预测处理，得到业务偏差数据；

[0076] 步骤S105，根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；

[0077] 步骤S106，根据核查结果输出任务调度指令或者对原始业务数据进行数据校正。

[0078] 本申请实施例所示意的步骤S101至步骤S106，通过获取目标阶段的原始业务数据；根据预设标签对原始业务数据进行分类处理，得到标签业务数据，能够根据业务需求，获取不同目标阶段的业务数据，从而在不同业务阶段进行数据核查，提高全流程的数据准确性。进一步地，根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；通过预设算法对业务关键特征进行预测处理，得到业务偏差数据，能够对不同目标阶段的业务数据采取不同的提取方式进行提取，从而使得获取到的业务关键特征更为准确。相较

于人工核查的方式，本申请实施例通过预设算法进行数据预测，能够有效地缩短数据处理的时间，并且能够减小业务偏差数据的计算误差，从而提高数据核查的效率与核查精度。最后，根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；根据核查结果输出任务调度指令或者对原始业务数据进行数据校正，能够根据业务偏差数据分析出原始业务数据处于正常状态还是异常状态，从而根据核查结果及时地采取相应的措施，从而提高数据处理的时效性。

[0079] 在一些实施例的步骤S101中，可以通过编写网络爬虫，设置好数据源之后进行有目标性地爬取数据，得到目标阶段的原始业务数据。也可以通过其他方式获取原始业务数据，不限于此。需要说明的是，在保险业务中，目标阶段主要包括保单汇总阶段、计算阶段以及数据压缩阶段，每个阶段都会产生相应的业务结果数据，例如，在保单汇总阶段，原始业务数据包括多个类型的保险数据，保险数据包括保单数、保费等等。

[0080] 请参阅图2，在一些实施例中，步骤S102可以包括但不限于包括步骤S201至步骤S204：

[0081] 步骤S201，获取目标阶段的属性数据；

[0082] 步骤S202，根据属性数据对预设标签库中的预设标签进行筛选处理，得到目标标签；

[0083] 步骤S203，通过预设函数和目标标签对原始业务数据进行标签概率计算，得到标签概率值；

[0084] 步骤S204，根据标签概率值对原始业务数据进行分类处理，得到标签业务数据。

[0085] 在一些实施例的步骤S201中，目标阶段包括保单汇总阶段、计算阶段以及数据压缩阶段，在每一阶段中均会产生大量的业务数据，同时，由于每一阶段的属性数据存在差异，因此，可以通过不同的属性数据来区分不同阶段。例如，当目标阶段为保单汇总阶段，则属性数据包括保单种类等；当目标阶段为数据压缩阶段，则属性数据包括产品种类等等。因此，可以通过编写网络爬虫，设置好数据源之后进行有目标性地爬取数据，得到目标阶段的属性数据。也可以通过其他方式获取属性数据，不限于此。

[0086] 在一些实施例的步骤S202中，预设标签库中存储有不同目标阶段的数据标签，即预设标签，其中，预设标签包括保险类别标签、时间标签、数量标签或者数值标签等等。通过不同阶段的属性数据，可以从多个预设标签中筛选出符合当前阶段需求的预设标签，将这一类预设标签作为目标标签，例如，在保单汇总阶段，属性数据包括保单种类，根据保单种类，可以从预设标签中筛选出数量标签和数值标签，以便根据数量标签获取不同种类的保单数量，根据数值标签获取不同类型的保单的保险金额。

[0087] 在一些实施例的步骤S203中，预设函数为softmax函数，通过softmax函数能够方便地创建出原始业务数据在目标标签上的概率分布，从而得到每一目标标签的标签概率值，该标签概率值能够用于反映原始业务数据属于目标标签的可能性。

[0088] 在一些实施例的步骤S204中，由于标签概率值能够用于反映原始业务数据属于目标标签的可能性，因此，根据标签概率值对原始业务数据进行分类处理，将原始业务数据划分到标签概率值最大的目标标签所在类别，从而将不同阶段的原始业务数据进行区分，得到不同阶段的不同类型的标签业务数据，例如，通过这一方式能够较为方便地将保单汇总阶段汇总属于不同险种的原始业务数据按照保险类别进行划分，得到属于保单汇总阶段的

标签业务数据。

[0089] 请参阅图3,在一些实施例中,目标阶段包括第一阶段,提取方式包括第一方式,步骤S103可以包括但不限于包括步骤S301至步骤S303:

[0090] 步骤S301,若标签业务数据为第一阶段的业务数据,则获取第一方式对应的数据类别词段;

[0091] 步骤S302,根据数据类别词段对标签业务数据进行关键词提取,得到初始业务关键词;

[0092] 步骤S303,对初始业务关键词进行过滤处理,得到业务关键特征。

[0093] 在一些实施例的步骤S301中,第一阶段为保单汇总阶段,若标签业务数据为保单汇总阶段的业务数据,则获取第一方式对应的数据类别词段,第一方式一般是对不同类型的数据设置对应的标签词段,即数据类型词段,其中,数据类型词段具体为保单数量、保单金额等等。

[0094] 在一些实施例的步骤S302中,可以通过TF-IDF算法将标签业务数据处理成若干个句子节点,具体地,通过TF-IDF算法计算标签业务数据中每一句子的数据类别词段出现的频率,得到数据类别词段的词频(Term Frequency,TF),其中,TF＝数据类别词段出现的次数/标签业务数据中的词段数量;进一步地,计算数据类别词段的逆向文件频率(Inverse Document Frequency,IDF),其中,IDF＝log(标签业务数据的词段总数/(包含数据类别词段的标签业务数据的词段数量+1)),最后,根据词频和逆向文件频率计算出词段的综合频率值,综合频率值＝词频*逆向文件频率。在每一目标文本中选取综合频率值最大的词段作为初始业务关键词。

[0095] 在一些实施例的步骤S303中,对初始业务关键词预设字符范围内的其他词段进行提取,将与初始业务关键词的相关性较高的词段保留,相关性较低的词段过滤掉,得到业务关键特征,例如,若初始业务关键词为保单金额,则将保单金额这一字段位置接近的词段进行提取,将提取到的词段中为数值以及金额计量单位的词段保留,得到保单金额对应的业务关键特征:5万元。

[0096] 请参阅图4,在一些实施例中,目标阶段包括第二阶段,提取方式包括第二方式,步骤S103可以包括但不限于包括步骤S401至步骤S405:

[0097] 步骤S401,若标签业务数据为第二阶段的业务数据,则根据预设的数据维度,获取第二方式对应的算法配置表,算法配置表存储有预设标签和预设算法,预设标签和预设算法之间具有映射关系;

[0098] 步骤S402,从预设标签提取出标签业务数据对应的目标标签;

[0099] 步骤S403,根据映射关系,从算法配置表提取出目标标签对应的预设算法;

[0100] 步骤S404,通过预设算法对标签业务数据进行字段提取,得到初始关键字段;

[0101] 步骤S405,对初始关键字段进行字段值计算,得到业务关键特征。

[0102] 在一些实施例的步骤S401中,第二阶段可以是指计算阶段或者数据压缩阶段,第二方式一般为算法配置方式,数据维度可以根据实际业务需求设置,当处于计算阶段时,预设的数据维度为保单级别的数据分析,当处于数据压缩阶段时,预设的数据维度为产品级别的数据分析。当标签业务实际为计算阶段的业务数据或者数据压缩阶段的业务数据。获取第二方式对应的算法配置表,其中,算法配置表存储有预设标签和预设算法,且预设标签

和预设算法之间具有映射关系。需要说明的是，这种映射关系可以是一对一或者一对多的形式。

[0103] 在一些实施例的步骤S402中，由于目标标签是预设标签中的某一个，则可以方便地根据目标标签的标签名称或者标签值从预设标签中提取到相同的标签。

[0104] 在一些实施例的步骤S403中，当确定预设标签中的哪个标签为目标标签之后，则可以直接从算法配置表中提取出和这一预设标签对应的预设算法以及两者的映射关系数据。例如，根据目标标签遍历算法配置表，提取出目标标签对有的预设算法，以及目标标签与预设算法之间的映射数据，其中，该映射数据包括目标标签、预设算法、目标标签和预设算法的相关度。

[0105] 在一些实施例的步骤S404中，预设算法可以是通过slice函数、start函数以及stop函数来实现对标签业务数据的字段提取，得到初始关键字段，其中，预设算法可以是TF-IDF算法等等，不做限制。

[0106] 在一些实施例的步骤S405中，对初始关键字段进行字段值计算时，首先通过tan函数对初始关键字段进行字段值提取，再对所有计算完成的初始关键字段的字段值进行求和或者求平均，从而得到业务关键特征。

[0107] 在一个具体实施例中，当数据核查处于计算阶段时，获取预设的算法配置表，该算法配置表中配置了目标标签以及对应目标标签的预设算法，即通过算法配置表可以明确需要计算的字段以及计算该字段的计算方法。在实际的业务关键特征计算时，能够根据相应的预设算法对标签业务数据进行字段提取以及字段值计算，从而得到业务关键特征。

[0108] 请参阅图5，在一些实施例中，步骤S104可以包括但不限于包括步骤S501至步骤S502：

[0109] 步骤S501，通过最小二乘法对业务关键特征进行回归预测，得到初始预测数据；

[0110] 步骤S502，根据预设的偏差参数对初始预测数据进行校准处理，得到业务偏差数据。

[0111] 在一些实施例的步骤S501中，通过最小二乘法能够最小化误差的平方和寻找数据的最佳函数匹配，因此通过最小二乘法能够对业务关键特征进行回归预测，生成回归线，得到初始预测数据，该初始预测数据可以表示为曲线形式。

[0112] 在一些实施例的步骤S502中，预设的偏差参数可以根据实际业务需求设置，不做限制，例如，偏差参数为±5%，则通过预设的偏差参数对初始预测数据对应的回归预测范围进行校准处理，得到业务偏差数据，该业务偏差数据为回归预测数值区间。

[0113] 请参阅图6，在一些实施例，核查结果包括第一结果和第二结果，步骤S105包括但不限于包括步骤S601至步骤S602：

[0114] 步骤S601，若业务偏差数据处于预设的偏差范围之内，则确定原始业务数据为正常状态，输出第一结果；

[0115] 步骤S602，若业务偏差数据处于预设的偏差范围之外，则确定原始业务数据为异常状态，输出第二结果。

[0116] 在一些实施例的步骤S601中，若业务偏差数据处于预设的偏差范围之内，即标签业务数据对应的字段信息的偏差处于预设的偏差范围之内，则原始业务数据是正常的，输出第一结果，其中，第一结果的具体语义内容包括：当前的原始业务数据为正常状态，核查

通过。从而能够方便地通过第一结果表征出正常的原始业务数据。

[0117] 在一些实施例的步骤S602中，若业务偏差数据处于预设的偏差范围之外，即标签业务数据对应的字段信息的偏差处于预设的偏差范围之外，则原始业务数据是异常的，输出第二结果，其中，第二结果的具体语义内容包括：当前的原始业务数据为异常状态，核查不通过，需要对当前的原始业务数据进行异常分析，确定异常原因并对当前原始业务数据进行数据校正操作。从而能够给方便地通过第二结果表征出存在异常的原始业务数据。

[0118] 请参阅图7，在一些实施例中，核查结果包括第一结果和第二结果，步骤S106可以包括但不限于包括步骤S701至步骤S702：

[0119] 步骤S701，若核查结果为第一结果，则输出任务调度指令，以根据任务调度指令进行业务数据的核查；

[0120] 步骤S702，若核查结果为第二结果，则对原始业务数据进行异常定位，生成异常信息，并根据异常信息对原始业务数据进行数据校正。

[0121] 在一些实施例的步骤S701中，若核查结果为第一结果，则表示当前阶段的原始业务数据正常，可以进行到下一阶段的数据核查，因此，输出任务调度指令，并根据任务调度指令提取下一阶段的原始业务数据，以相同的核查过程对下一阶段的原始业务数据进行核查。

[0122] 在一些实施例的步骤S702中，若核查结果为第二结果，则表示当前阶段的原始业务数据异常，需要对异常数据进行处理，因此，根据原始业务数据与历史业务数据的比对，确定处于异常状态的原始业务数据的位置，生成异常信息，该异常信息包括异常位置、偏差情况以及异常原因等等，从而根据预设的解决方案来对这一原始业务数据进行数据校正，其中，预设的解决方案可以是根据历史排查的异常问题确定的异常原因和解决措施来制定，其中，解决方案可以以配置表的形式进行存储，以便在确定处于异常状态的原始业务数据之后，遍历配置表提取相应的解决方案。

[0123] 在一个具体实施例中，对某一月份的保险业务数据进行核查过程包括三个阶段，即保单汇总阶段、DCS计算阶段以及数据压缩阶段。

[0124] 首先，获取保单汇总阶段的原始业务数据，该原始业务数据一般以结果表数据的形式呈现，对原始业务数据按照保险类型这一属性数据进行分类统计，得到属于不同保险类型的标签业务数据。进一步地，根据保单数量、保单金额等数据类型词段对标签业务数据进行关键字段提取，得到业务关键特征，该业务关键特征主要包括该月份的具体保单数量以及具体保单金额等等。进一步地，通过最小二乘法对这些业务关键特征进行回归预测，构建每一业务关键特征对应的回归曲线，并对每个业务关键特征赋予±5%的偏差，得到每一业务关键特征的业务偏差数据，将业务偏差数据与历史记录中该月份的这一业务关键特征进行比对，若超出了预设的偏差范围，则确定该业务关键特征对应的原始业务数据为异常状态，则针对这一原始业务数据进行异常分析，具体地，根据业务关键特征遍历预设的问题配置表，从问题配置表中调取根据历史排查的异常问题确定的异常原因和解决方案，从而对处于异常状态的原始业务数据进行校正处理，直到保单汇总阶段的所有原始业务数据均为正常状态，输出任务调度指令，并根据任务调度指令提取下一阶段的原始业务数据。

[0125] 进一步地，当保单汇总阶段的原始业务数据核查完毕之后，根据任务调度指令，对DCS计算阶段的原始业务数据进行核查处理。具体地，获取DCS计算阶段的标签业务数据，并

获取预设的算法配置表，该算法配置表中配置了目标标签以及对应目标标签的预设算法，即通过算法配置表可以明确需要计算的标签业务数据以及计算该标签业务数据的计算方法，通过算法配置表来提取出每一标签业务数据对应的预设算法，根据相应的预设算法对标签业务数据进行字段提取以及字段值计算，从而得到业务关键特征，其中，该业务关键特征。进而，通过最小二乘法对这些业务关键特征进行回归预测，构建每一业务关键特征对应的回归曲线，并对每个业务关键特征赋予±5%的偏差，得到每一业务关键特征的业务偏差数据，将业务偏差数据与历史记录中该月份的这一业务关键特征进行比对，若超出了预设的偏差范围，则确定该业务关键特征对应的原始业务数据为异常状态，则针对这一原始业务数据进行异常分析，具体地，根据业务关键特征遍历预设的问题配置表，从问题配置表中调取根据历史排查的异常问题确定的异常原因和解决方案，从而对处于异常状态的原始业务数据进行校正处理，直到保单汇总阶段的所有原始业务数据均为正常状态，输出任务调度指令，并根据任务调度指令提取下一阶段的原始业务数据。

[0126] 进一步地，当DSC计算阶段的原始业务数据核查完毕之后，根据任务调度指令，对数据压缩阶段的原始业务数据进行核查处理。具体地，在数据压缩阶段的数据核查主要为对同一类型的数据进行核查，首先按照预设维度对同一类型的标签业务数据进行group by操作，实现标签业务数据的聚合，得到标签业务数据对应的字段数值。从而对这些字段数据进行求和或者求平均处理，得到该阶段的业务关键特征，同样地，采用最小二乘法对该阶段的业务关键特征进行异常核查，从而对处于异常状态的原始业务数据进行校正处理。

[0127] 需要说明的是，在本申请实施例中，若当前阶段的原始业务数据未核查完毕，不会对下一阶段的原始业务数据进行核查，当当前阶段的原始业务数据核查完毕，可以通过计算机程序自动验证合格进入到下一阶段的数据核查，通过这一方式能够对每一阶段的原始业务数据进行逐一排查，能够有效地节约业务人员对结果数据的验证时间，提高数据核查的效率，同时各个阶段可以通过计算机程序自动验证，当验证合格之后能够自动进入下一阶段的数据核查，实现每一阶段数据核查的无缝连接。此外，当发现处于异常状态的原始业务数据时，可以通过计算机程序根据参数与历史问题相结合的方式，从预设的问题配置表上查询到可能出现问题的原因，从而方便业务人员对异常问题进行快速定位，提高解决问题的时效性。

[0128] 本申请实施例的数据核查方法，通过获取目标阶段的原始业务数据；根据预设标签对原始业务数据进行分类处理，得到标签业务数据，能够根据业务需求，获取不同目标阶段的业务数据，从而在不同业务阶段进行数据核查，提高全流程的数据准确性。进一步地，根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；通过预设算法对业务关键特征进行预测处理，得到业务偏差数据，能够对不同目标阶段的业务数据采取不同的提取方式进行提取，从而使得获取到的业务关键特征更为准确。相较于人工核查的方式，本申请实施例通过预设算法进行数据预测，能够有效地缩短数据处理的时间，并且能够减小业务偏差数据的计算误差，从而提高数据核查的效率与核查精度。最后，根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；根据核查结果输出任务调度指令或者对原始业务数据进行数据校正，能够根据业务偏差数据分析出原始业务数据处于正常状态还是异常状态，从而根据核查结果及时地采取相应的措施，从而提高数据处理的时效性。

[0129] 请参阅图8，本申请实施例还提供一种数据核查装置，可以实现上述数据核查方

法，该装置包括：

[0130] 数据获取模块801，用于获取目标阶段的原始业务数据；

[0131] 分类模块802，用于根据预设标签对原始业务数据进行分类处理，得到标签业务数据；

[0132] 特征提取模块803，用于根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；

[0133] 预测模块804，用于通过预设算法对业务关键特征进行预测处理，得到业务偏差数据；

[0134] 核查模块805，用于根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；

[0135] 处理模块806，用于根据核查结果输出任务调度指令或者对原始业务数据进行数据校正。

[0136] 在一些实施例中，分类模块802包括：

[0137] 属性数据获取单元，用于获取目标阶段的属性数据；

[0138] 筛选单元，用于根据属性数据对预设标签库中的预设标签进行筛选处理，得到目标标签；

[0139] 概率计算单元，用于通过预设函数和目标标签对原始业务数据进行标签概率计算，得到标签概率值；

[0140] 数据分类单元，用于根据标签概率值对原始业务数据进行分类处理，得到标签业务数据。

[0141] 在一些实施例中，目标阶段包括第一阶段，提取方式包括第一方式，特征提取模块803包括：

[0142] 词段获取单元，用于若标签业务数据为第一阶段的业务数据，则获取第一方式对应的数据类别词段；

[0143] 关键词提取单元，用于根据数据类别词段对标签业务数据进行关键词提取，得到初始业务关键词；

[0144] 过滤单元，用于对初始业务关键词进行过滤处理，得到业务关键特征。

[0145] 在一些实施例中，目标阶段包括第二阶段，提取方式包括第二方式，特征提取模块803还包括：

[0146] 配置表获取单元，用于若标签业务数据为第二阶段的业务数据，则根据预设的数据维度，获取第二方式对应的算法配置表，算法配置表存储有预设标签和预设算法，预设标签和预设算法之间具有映射关系；

[0147] 标签提取单元，用于从预设标签提取出标签业务数据对应的目标标签；

[0148] 算法提取单元，用于根据映射关系，从算法配置表提取出目标标签对应的预设算法；

[0149] 字段提取单元，用于通过预设算法对标签业务数据进行字段提取，得到初始关键字段；

[0150] 字段值计算单元，用于对初始关键字段进行字段值计算，得到业务关键特征。

[0151] 在一些实施例中，预测模块804包括：

[0152] 回归预测单元，用于通过最小二乘法对业务关键特征进行回归预测，得到初始预测数据；

[0153] 校准单元，用于根据预设的偏差参数对初始预测数据进行校准处理，得到业务偏差数据。

[0154] 在一些实施例中，核查结果包括第一结果和第二结果，核查模块805包括：

[0155] 第一判断单元，用于若业务偏差数据处于预设的偏差范围之内，则确定原始业务数据为正常状态，输出第一结果；

[0156] 第二判断单元，用于若业务偏差数据处于预设的偏差范围之外，则确定原始业务数据为异常状态，输出第二结果。

[0157] 在一些实施例中，核查结果包括第一结果和第二结果，处理模块806包括：

[0158] 输出单元，用于若核查结果为第一结果，则输出任务调度指令，以根据任务调度指令进行业务数据的核查；

[0159] 异常处理单元，用于若核查结果为第二结果，则对原始业务数据进行异常定位，生成异常信息，并根据异常信息对原始业务数据进行数据校正。

[0160] 该数据核查装置的具体实施方式与上述数据核查方法的具体实施例基本相同，在此不再赘述。

[0161] 本申请实施例还提供了一种电子设备，电子设备包括：存储器、处理器、存储在存储器上并可在处理器上运行的程序以及用于实现处理器和存储器之间的连接通信的数据总线，程序被处理器执行时实现上述数据核查方法。该电子设备可以为包括平板电脑、车载电脑等任意智能终端。

[0162] 请参阅图9，图9示意了另一实施例的电子设备的硬件结构，电子设备包括：

[0163] 处理器901，可以采用通用的CPU(CentralProcessingUnit，中央处理器)、微处理器、应用专用集成电路(ApplicationSpecificIntegratedCircuit，ASIC)、或者一个或多个集成电路等方式实现，用于执行相关程序，以实现本申请实施例所提供的技术方案；

[0164] 存储器902，可以采用只读存储器(ReadOnlyMemory，ROM)、静态存储设备、动态存储设备或者随机存取存储器(RandomAccessMemory，RAM)等形式实现。存储器902可以存储操作系统和其他应用程序，在通过软件或者固件来实现本说明书实施例所提供的技术方案时，相关的计算机程序代码保存在存储器902中，并由处理器901来调用执行本申请实施例的数据核查方法；

[0165] 输入/输出接口903，用于实现信息输入及输出；

[0166] 通信接口904，用于实现本设备与其他设备的通信交互，可以通过有线方式(例如USB、网线等)实现通信，也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信；

[0167] 总线905，在设备的各个组件(例如处理器901、存储器902、输入/输出接口903和通信接口904)之间传输信息；

[0168] 其中处理器901、存储器902、输入/输出接口903和通信接口904通过总线905实现彼此之间在设备内部的通信连接。

[0169] 本申请实施例还提供了一种计算机可读存储介质，计算机可读存储介质存储有一个或者多个计算机程序，一个或者多个计算机程序可被一个或者多个处理器执行，以实现上述数据核查方法。

[0170] 存储器作为一种非暂态计算机可读存储介质，可用于存储非暂态软件程序以及非暂态性计算机可执行程序。此外，存储器可以包括高速随机存取存储器，还可以包括非暂态存储器，例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。在一些实施方式中，存储器可选包括相对于处理器远程设置的存储器，这些远程存储器可以通过网络连接至该处理器。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0171] 本申请实施例提供的数据核查方法、数据核查装置、电子设备及存储介质，其通过获取目标阶段的原始业务数据；根据预设标签对原始业务数据进行分类处理，得到标签业务数据，能够根据业务需求，获取不同目标阶段的业务数据，从而在不同业务阶段进行数据核查，提高全流程的数据准确性。进一步地，根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征；通过预设算法对业务关键特征进行预测处理，得到业务偏差数据，能够对不同目标阶段的业务数据采取不同的提取方式进行提取，从而使得获取到的业务关键特征更为准确。相较于人工核查的方式，本申请实施例通过预设算法进行数据预测，能够有效地缩短数据处理的时间，并且能够减小业务偏差数据的计算误差，从而提高数据核查的效率与核查精度。最后，根据业务偏差数据对原始业务数据进行状态核查，得到核查结果；根据核查结果输出任务调度指令或者对原始业务数据进行数据校正，能够根据业务偏差数据分析出原始业务数据处于正常状态还是异常状态，从而根据核查结果及时地采取相应的措施，从而提高数据处理的时效性。

[0172] 本申请实施例描述的实施例是为了更加清楚的说明本申请实施例的技术方案，并不构成对于本申请实施例提供的技术方案的限定，本领域技术人员可知，随着技术的演变和新应用场景的出现，本申请实施例提供的技术方案对于类似的技术问题，同样适用。

[0173] 本领域技术人员可以理解的是，图1-7中示出的技术方案并不构成对本申请实施例的限定，可以包括比图示更多或更少的步骤，或者组合某些步骤，或者不同的步骤。

[0174] 以上所描述的装置实施例仅仅是示意性的，其中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，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模块来实现本实施例方案的目的。

[0175] 本领域普通技术人员可以理解，上文中所公开方法中的全部或某些步骤、系统、设备中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。

[0176] 本申请的说明书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)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，以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。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，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0177] 应当理解，在本申请中，“至少一个(项)”是指一个或者多个，“多个”是指两个或两个以上。“和/或”，用于描述关联对象的关联关系，表示可以存在三种关系，例如，“A和/或B”可以表示：只存在A，只存在B以及同时存在A和B三种情况，其中A，B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项(个)”或其类似表达，是指

这些项中的任意组合，包括单项(个)或复数项(个)的任意组合。例如，a，b或c中的至少一项(个)，可以表示：a，b，c，“a和b”，“a和c”，“b和c”，或“a和b和c”，其中a，b，c可以是单个，也可以是多个。

[0178] 在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，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，上述单元的划分，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。另一点，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。

[0179] 上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方案的目的。

[0180] 另外，在本申请各个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。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

[0181] 集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取计算机可读存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或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一个计算机可读存储介质中，包括多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)执行本申请各个实施例的方法的全部或部分步骤。而前述的计算机可读存储介质包括：U盘、移动硬盘、只读存储器(Read-Only Memory，简称ROM)、随机存取存储器(Random Access Memory，简称RAM)、磁碟或者光盘等各种可以存储计算机程序的介质。

[0182] 以上参照附图说明了本申请实施例的优选实施例，并非因此局限本申请实施例的权利范围。本领域技术人员不脱离本申请实施例的范围和实质内所作的任何修改、等同替换和改进，均应在本申请实施例的权利范围之内。

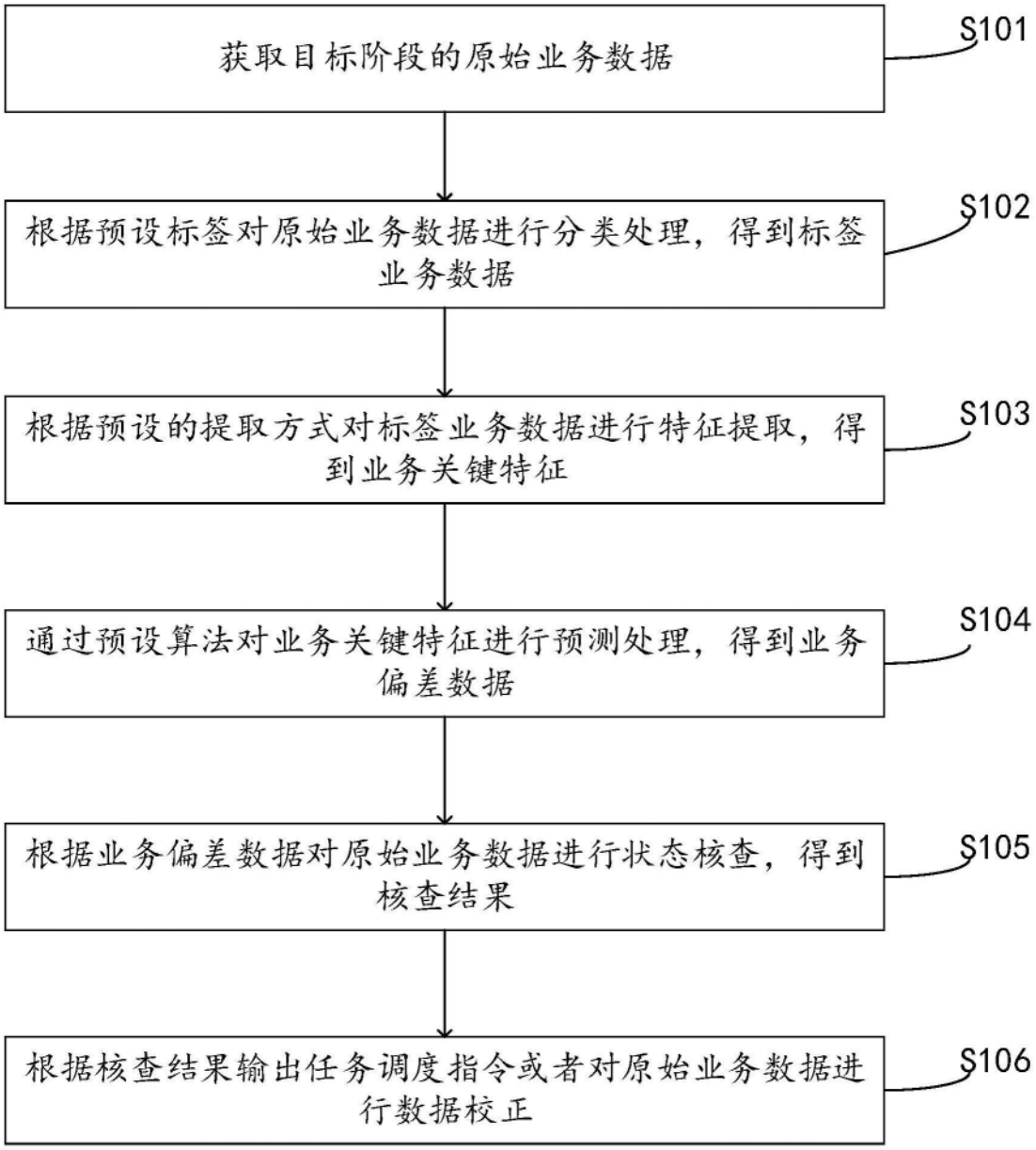

获取目标阶段的原始业务数据
S101
根据预设标签对原始业务数据进行分类处理，得到标签业务数据
S102
根据预设的提取方式对标签业务数据进行特征提取，得到业务关键特征
S103
通过预设算法对业务关键特征进行预测处理，得到业务偏差数据
S104
根据业务偏差数据对原始业务数据进行状态核查，得到核查结果
S105
根据核查结果输出任务调度指令或者对原始业务数据进行数据校正
S106

图1

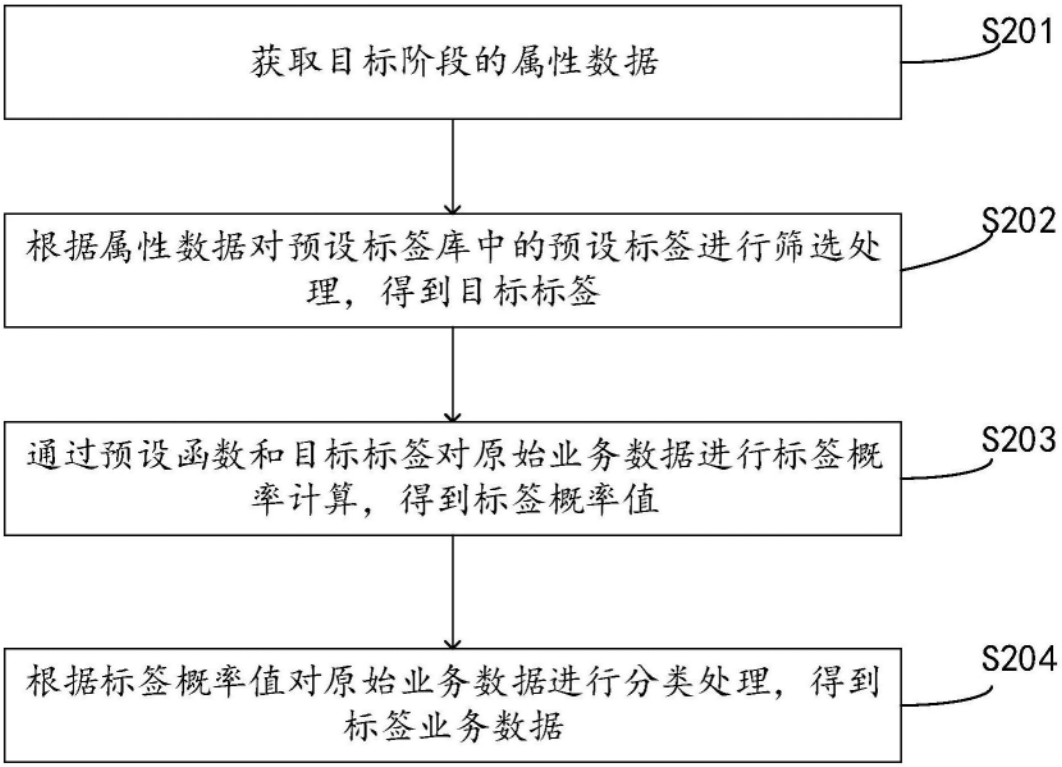

获取目标阶段的属性数据
S201
根据属性数据对预设标签库中的预设标签进行筛选处理，得到目标标签
S202
通过预设函数和目标标签对原始业务数据进行标签概率计算，得到标签概率值
S203
根据标签概率值对原始业务数据进行分类处理，得到标签业务数据
S204

图2

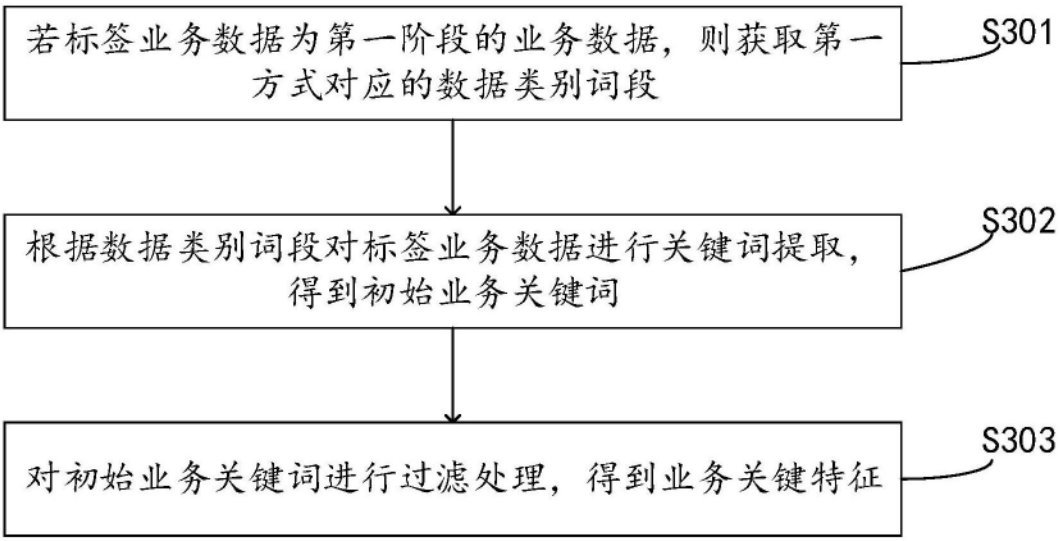

若标签业务数据为第一阶段的业务数据，则获取第一方式对应的数据类别词段
S301
根据数据类别词段对标签业务数据进行关键词提取，得到初始业务关键词
S302
对初始业务关键词进行过滤处理，得到业务关键特征
S303

图3

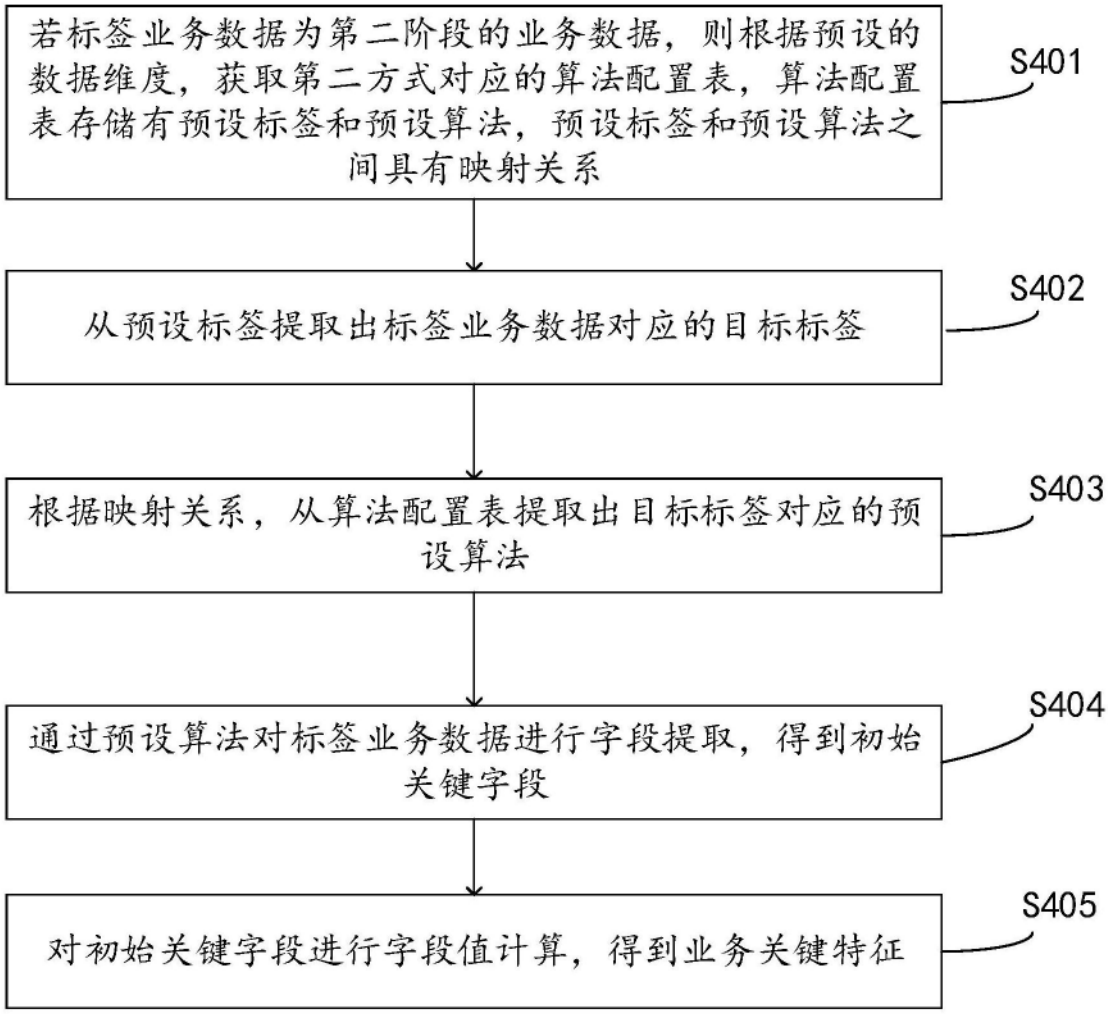

若标签业务数据为第二阶段的业务数据，则根据预设的数据维度，获取第二方式对应的算法配置表，算法配置表存储有预设标签和预设算法，预设标签和预设算法之间具有映射关系
S401
从预设标签提取出标签业务数据对应的目标标签
S402
根据映射关系，从算法配置表提取出目标标签对应的预设算法
S403
通过预设算法对标签业务数据进行字段提取，得到初始关键字段
S404
对初始关键字段进行字段值计算，得到业务关键特征
S405

图4

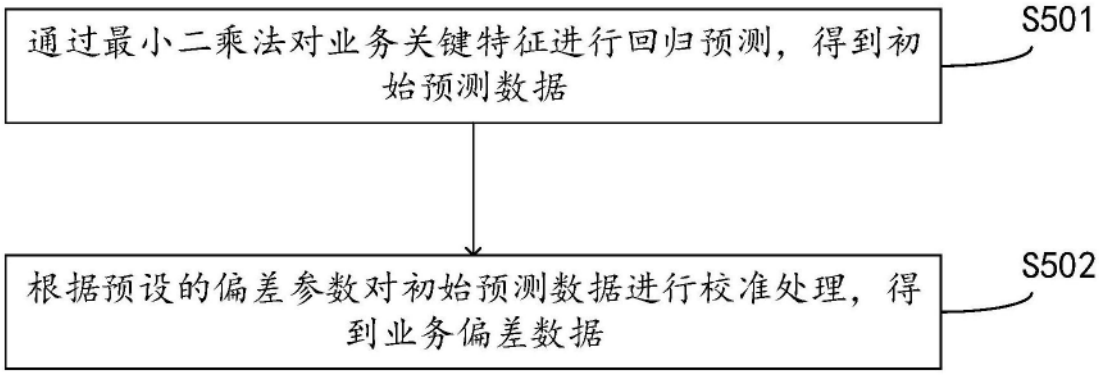

通过最小二乘法对业务关键特征进行回归预测，得到初始预测数据
S501
根据预设的偏差参数对初始预测数据进行校准处理，得到业务偏差数据
S502

图5

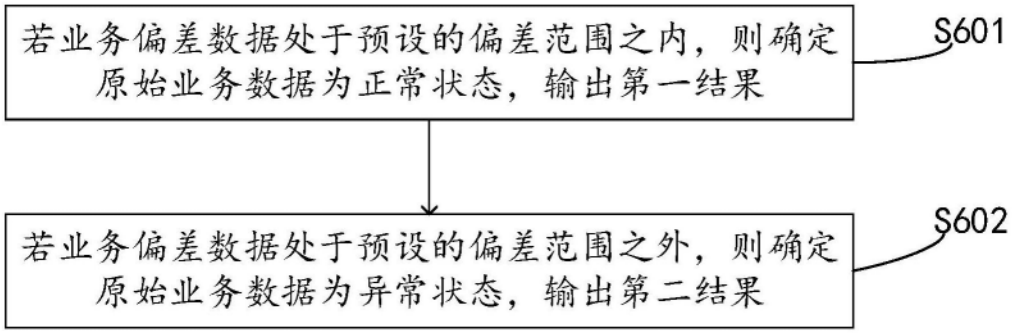

若业务偏差数据处于预设的偏差范围之内，则确定原始业务数据为正常状态，输出第一结果
S601
若业务偏差数据处于预设的偏差范围之外，则确定原始业务数据为异常状态，输出第二结果
S602

图6

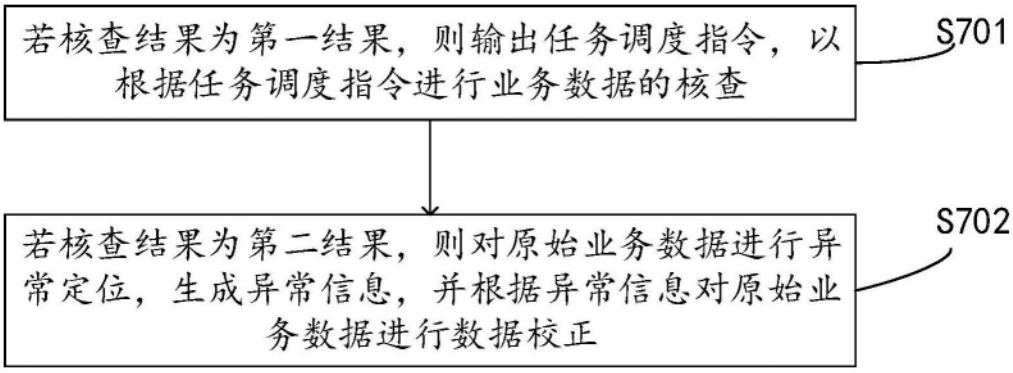

若核查结果为第一结果，则输出任务调度指令，以根据任务调度指令进行业务数据的核查
S701
若核查结果为第二结果，则对原始业务数据进行异常定位，生成异常信息，并根据异常信息对原始业务数据进行数据校正
S702

图7

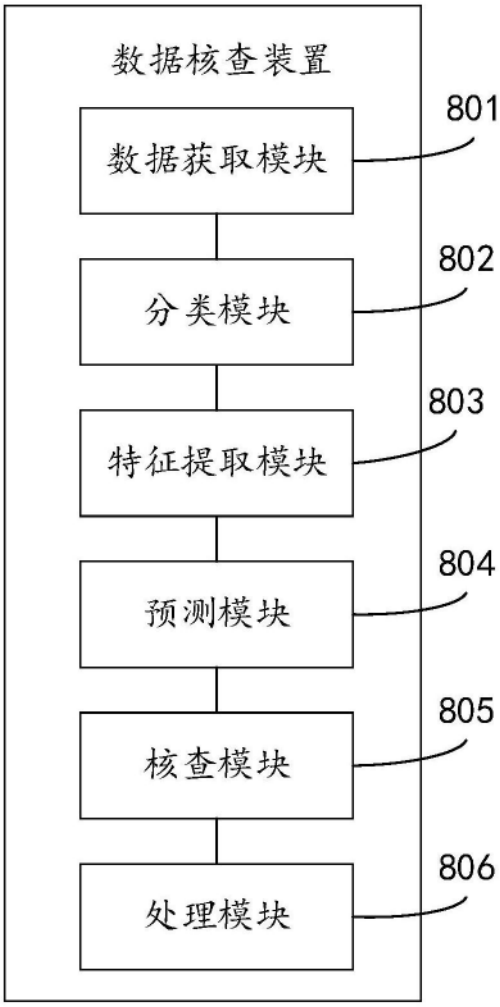

数据核查装置
数据获取模块
801
分类模块
802
特征提取模块
803
预测模块
804
核查模块
805
处理模块
806

图8

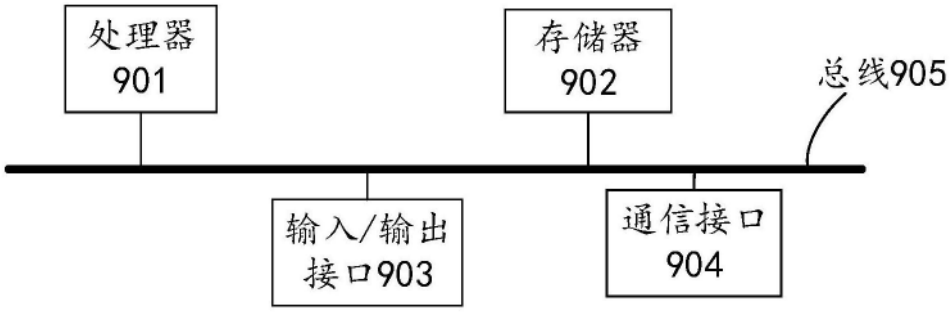

处理器
901
存储器
902
总线905
输入/输出
接口903
通信接口
904

图9