DATA ANOMALY DETECTION METHOD AND SYSTEM

说　　明　　书

数据异常检测方法及系统

技术领域

本说明书涉及数据处理技术领域，尤其涉及一种数据异常检测方法及系统。

背景技术

交易成功率的异常检测对于电子商务等涉及交易的诸多领域非常重要。过去的交易成功率的检测主要集中在使用简单的规则或者基于统计的方法进行异常检测。进行异常检测的规则例如基于预设的静态阈值对交易成功率进行异常检测，但是上述方法无法适应交易数据的动态变化。这导致在高波动性的应用环境中，很难准确地检测到交易成功率的异常情况。

因此，需要提供一种可以多尺度进行异常数据检测的方法，来提高现有技术中的异常情况检测的准确性。

背景技术部分的内容仅仅是发明人个人所知晓的信息，并不代表上述信息在本说明书申请日之前已经进入公共领域，也不代表其可以成为本说明书的现有技术。

发明内容

本说明书提供一种数据异常检测方法及系统，能够适用于需要对交易成功率进行异常检测的场景下。

第一方面，本说明书提供一种数据异常检测方法，包括：获取待检测交易成功率序列；其中，所述待检测支付成率序列包括第一子序列和第二子序列；所述第一子序列对

应的交易成功率均值高于所述第二子序列对应的交易成功率均值；根据所述第一子序列对应的第一交易量序列和所述第一子序列，建立交易成功率与交易量之间的统计分布模型；以及基于所述统计分布模型和所述第二子序列对应的第二交易量序列，对所述第二子序列进行检验，确定所述第二子序列是否异常。

第二方面，本说明书还提供一种数据异常检测系统，包括至少一个存储介质以及至少一个处理器，所述至少一个存储介质存储有至少一个指令集，用于进行数据异常检测方法；所述至少一个处理器同所述至少一个存储介质通信连接，其中，所述至少一个处理器运行时读取所述至少一个指令集，并根据所述至少一个指令集的指示执行：获取待检测交易成功率序列；其中，所述待检测支付成率序列包括第一子序列和第二子序列；所述第一子序列对应的交易成功率均值高于所述第二子序列对应的交易成功率均值；根据所述第一子序列对应的第一交易量序列和所述第一子序列，建立交易成功率与交易量之间的统计分布模型；以及基于所述统计分布模型和所述第二子序列对应的第二交易量序列，对所述第二子序列进行检验，确定所述第二子序列是否异常。

第三方面，本说明书还提供了一种计算机可读非易失性存储介质，其中，所述计算机可读非易失性存储介质中存储有至少一个指令集，所述至少一个指令集被至少一个处理器执行时，实现如上述第一方面任一项所述的数据异常检测方法。

由以上技术方案可知，本说明书提供的数据异常检测方法及系统，在对待检测交易成功率序列进行检验时，会综合交易成功率与交易量之间的关系，来对待检测交易成功率序列是否存在异常进行检测。在本说明书提供的方法中，会基于待检测交易成功率序列中第一子序列对应的第一交易量序列和第一子序列，建立交易成功率与交易量之间的统计分布模型。基于统计分布模型和待检测交易成功率序列中的第二子序列对应的第二交易量序列，对第二子序列进行检验，以对第二子序列是否存在异常进行检验。这样的验证方式在验证过程中不但考虑了交易成功率，并且还考虑了交易量与交易成功率之间的关系，大大提高了异常检测的检测结果的准确性和可靠性。

本说明书提供的数据异常检测方法及系统的其他功能将在以下说明中部分列出。本说明书提供的数据异常检测方法及系统的创造性方面可以通过实践或使用下面详细示例中所述的方法、装置和组合得到充分解释。

附图说明

为了更清楚地说明本说明书实施例中的技术方案，下面将对实施例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说明书的一些实施例，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。

图 1 示出了根据本说明书的实施例提供的一种数据异常检测的应用场景示意图；

图 2 示出了根据本说明书的一些实施例提供的一种计算设备的硬件结构示意图；

图 3 示出了根据本说明书的实施例提供的一种数据异常检测方法的流程图；

图 4 示出了根据本说明书的另一实施例提供的一种数据异常检测方法的流程图；

图 5 示出了根据本说明书的实施例提供的建立统计分布模型方法的流程图；

图 6 示出了根据本说明书的实施例提供的建立统计分布模型的结构示意图。

具体实施方式

以下描述提供了本说明书的特定应用场景和要求，目的是使本领域技术人员能够制造和使用本说明书中的内容。对于本领域技术人员来说，对所公开的实施例的各种局部修改是显而易见的，并且在不脱离本说明书的精神和范围的情况下，可以将这里定义的一般原理应用于其他实施例和应用。因此，本说明书不限于所示的实施例，而是与权利要求一致的最宽范围。

这里使用的术语仅用于描述特定示例实施例的目的，而不是限制性的。比如，除非

上下文另有明确说明，这里所使用的，单数形式“一”，“一个”和“该”也可以包括复数形式。当在本说明书中使用时，术语“包括”、“包含”和/或“含有”意思是指所关联的整数，步骤、操作、元素和/或组件存在，但不排除一个或多个其他特征、整数、步骤、操作、元素、组件和/或组的存在或在该系统/方法中可以添加其他特征、整数、步骤、操作、元素、组件和/或组。

考虑到以下描述，本说明书的这些特征和其他特征、以及结构的相关元件的操作和功能、以及部件的组合和制造的经济性可以得到明显提高。参考附图，所有这些形成本说明书的一部分。然而，应该清楚地理解，附图仅用于说明和描述的目的，并不旨在限制本说明书的范围。还应理解，附图未按比例绘制。

本说明书中使用的流程图示出了根据本说明书中的一些实施例的系统实现的操作。应该清楚地理解，流程图的操作可以不按顺序实现。相反，操作可以以反转顺序或同时实现。此外，可以向流程图添加一个或多个其他操作。可以从流程图中移除一个或多个操作。

在本说明书中，“X 包括 A、B 或 C 中的至少一个”所表达的意思是 X 至少包括 A，或 X 至少包括 B，或 X 至少包括 C。也就是说，X 可以只包括 A、B、C 的任意一个，或者同时包括 A、B、C 的任意组合以及其他可能的内容/元素。所述 A、B、C 的任意组合可以是 A、B、C、AB、AC、BC、或 ABC。

在本说明书中，除非明确说明，否则结构之间产生的关联关系可以是直接的关联关系也可以是间接的关联关系。比如，当描述“A 与 B 连接”时，除非明确说明了 A 与 B 直接连接，否则应当理解成 A 可以与 B 直接连接，也可以与 B 间接地连接；再比如，当描述“A 在 B 之上”时，除非明确说明了 A 直接在 B 的上方（AB 相邻且 A 在 B 的上方），否则应当理解成 A 可以直接在 B 的上方，A 也可以间接地在 B 之上（AB 之间隔着其他元素，且 A 在 B 的上方）。以此类推。

需要说明的是，本说明书获取到的用户数据经过用户授权，不涉及用户隐私。

为了方便描述，首先对本说明书后文中将会出现的术语进行解释。

贝塔分布（Beta Distribution）是一种连续概率分布，常用于描述在区间 [0, 1] 内的随机变量的概率分布。它在统计学和机器学习中非常有用，特别是在处理比例、概率或比率时。贝塔分布有两个形状参数 α（alpha）和 β（beta），这两个参数决定了分布的形状。

在支付监测及异常检测场景中，交易成功率的异常检测对于金融、电商等诸多领域非常重要。交易成功率的下降可能意味着交易风险的增加。或是可能存在潜在的安全威胁，例如欺诈行为或恶意软件的攻击。或是可能存在服务器性能瓶颈或网络延迟等情况。及时监测并检测到交易成功率的异常，可以发现交易中潜在的问题，以及时对异常情况进行排查，消除交易潜在风险，保证交易成功率恢复正常，从而保证用户的体验。

交易成功率是一个高波动的值，对于流量大的情况，可能一段时间内有 100 个交易，一两个失败的交易对交易成功率的影响不大。但是对于流量小的情况，可能一段时间内只有 3 笔交易，一两个失败的交易则会对交易成功率的影响非常大。因此，仅基于静态的交易成功率阈值对交易成功率进行异常检测，无法适应交易数据的动态变化。在高波动性的环境中，很难准确地检测到交易成功率的异常情况。

其次，不同的交易成功率序列可能时间长短也不同，不同的时间尺度可能对应不同的异常检测方式。例如对于短时间内（每分钟或每小时）的交易成功率的检测方式与长时间尺度（每天或每周）的检测方式并不相同，基于现有技术中单一尺度对交易成功率进行检测的方式其准确性并无法保障。因此，本说明书提供了一种数据异常检测的方法，可以解决上述问题。

图 1 示出了根据本说明书的实施例提供的一种数据异常检测的应用场景示意图。

如图 1 所示，该应用场景 100 可以包括数据异常检测系统 11。数据异常检测系统

11 在获取待检测交易成功率序列后，可以根据待检测交易成功率序列中的第一子序列，和第一子序列对应的第一交易量序列，建立待检测交易成功率序列对应的交易成功率与交易量之间的统计分布模型。随后基于统计分布模型对待检测交易成功率序列中的第二子序列进行异常检测，确定第二子序列是否存在异常（异常检测结果）。上述检验方式可以将待检测交易成功率序列分为基线序列（第一子序列）和待检测序列（第二子序列），基于第一子序列和第一交易量序列建立统计分布模型，以建立正常情况（不存在异常的情况）下交易成功率与交易量之间的分布关系。基于正常情况下交易成功率与交易量之间的分布关系以及第二序列对应的第二交易量序列，对第二子序列进行验证，确定第二子序列是否存在异常。由于验证过程考虑了交易成功率与交易量之间的分布关系，不同的交易量可能对应的不同的交易成功率，相对于现有技术中只对成功率这一单一数据、单一尺度进行检测的方式，本说明书提供的方法在检测过程中考虑了不同尺度下的交易数据（交易成功率和交易量），提高了异常检测的准确定和可靠性。

其中，待检测交易成功率序列可以为历史支付成功数据，也可以为实时获取的一段时间之内的历史交易成功率数据。例如可以获取检测时刻之前一段时间内的历史交易成功率数据作为待检测交易成功率序列。可以是从本地数据库获取的，也可以是从云端、服务器等位置的数据库获取的。或者也可以为基于获取的同一时间段对应的交易量序列和成功量序列，生成的上述时间段对应的待检测交易成功率序列。具体待检测交易成功率序列的来源可以根据用户需要灵活调整，在此并不做限制。

数据异常检测系统 11 可以是具有一定计算能力的电子设备。数据异常检测系统 11 可以执行本说明书提供的数据异常检测方法。数据异常检测系统 11 可以存储有执行本说明书描述的数据异常检测方法的数据或指令，并可以执行或用于执行所述数据或指令。数据异常检测系统 11 可以包括具有数据信息处理功能的硬件设备和驱动该硬件设备工作所需必要的程序。

数据异常检测系统 11 可以是单个的计算设备，也可以是由多个计算设备组成的集

群系统。数据异常检测系统 11 中存储的执行本说明书描述的数据异常检测方法的数据或指令可以采用任何形式的系统架构。比如，分层架构、事件驱动架构、微核架构、微服务架构、或者云架构等等。

以提供实时数据异常检测的系统为例进行举例说明，上述系统架构中可以包括终端和服务器。终端可以部署有系统的客户端，服务器可以部署有系统的服务端。终端可以包括移动设备、平板电脑、笔记本电脑或类似内容，或其任意组合。在一些实施例中，移动设备可包括智能家居设备、智能移动设备、虚拟现实设备、增强现实设备或类似设备，或其任意组合。在一些实施例中，智能家居装置可包括智能电视、台式电脑等，或任意组合。在一些实施例中，智能移动设备可包括智能手机、个人数字辅助、游戏设备、导航设备等，或其任意组合。

在一些实施例中，数据异常检测系统 11 可以作为上述系统架构中的终端。例如，数据异常检测系统 11 可以接收待检测交易成功率序列，通过执行本说明书提供的数据异常检测方法得到待检测交易成功率序列中的第二子序列是否存在异常，以为支付安全和支付风险提供更可靠的支持。

在一些实施例中，数据异常检测系统 11 可以作为上述系统架构中的服务器。例如，终端接收输入的待检测交易成功率序列，并将检测交易成功率序列发送给数据异常检测系统 11，数据异常检测系统 11 执行本说明书提供的数据异常检测方法对待检测交易成功率序列进行处理，以得到待检测交易成功率序列中的第二子序列对应的异常检测结果。进而数据异常检测系统 11 可以向终端反馈该异常反馈结果。

应当理解，图 1 所示的应用场景只是本说明书提供的多种应用场景中的一个。图 1 中所示的数据异常检测系统 11 的数目仅仅为示例性的。根据实现需要，可以具有任意的数目。

图 2 示出了根据本说明书的一些实施例提供的一种计算设备 200 的硬件结构示意

图。如图 2 所示，计算设备 200 包括至少一个存储介质 230 和至少一个处理器 220。在一些实施例中，计算设备 200 还可以包括内部通信总线 210。在一些实施例中，计算设备 200 还可以包括通信端口 250。在一些实施例中，计算设备 200 还可以包括 I/O 组件 260。

内部通信总线 210 可以连接不同的系统组件，包括存储介质 230 和处理器 220。I/O 组件 260 支持计算设备 200 和其他组件之间的输入/输出。

存储介质 230 可以包括数据存储装置。数据存储装置可以是非暂时性存储介质，也可以是暂时性存储介质。比如，数据存储装置可以包括磁盘 232、只读存储介质（ROM）234 或随机存取存储介质（RAM）236 中的一种或多种。存储介质 230 还包括存储在数据存储装置中的至少一个指令集。指令集是计算机程序代码，计算机程序代码可以包括执行本说明书提供数据异常检测方法的程序、例程、对象、组件、数据结构、过程、模块等等。

通信端口 250 用于计算设备 200 同外界的数据通讯。比如，计算设备 200 可以通过通信端口 250 连接网络。至少一个处理器 220 同至少一个存储介质 230 通过内部通信总线 210 通信连接。至少一个处理器 220 用以执行上述至少一个指令集。系统 11 运行时，至少一个处理器 220 读取至少一个指令集，并且根据至少一个指令集的指示执行本说明书提供的数据异常检测方法。

处理器 220 可以执行数据异常检测方法包含的所有步骤。处理器 220 可以是一个或多个处理器的形式。处理器 220 可以发出执行指令。处理器 220 可以包括一个或多个硬件处理器，例如微控制器，微处理器，精简指令集计算机（RISC），专用集成电路（ASIC），特定于应用的指令集处理器（ASIP），中心处理单元（CPU），图形处理单元（GPU），物理处理单元（PPU），微控制器单元，数字信号处理器（DSP），现场可编程门阵列（FPGA），高级 RISC 机器（ARM），可编程逻辑器件（PLD），能够执行一个或多个

功能的任何电路或处理器等，或其任何组合。

仅仅为了说明问题，在本说明书中计算设备 200 中仅描述了一个处理器 220。然而，应当注意，本说明书中计算设备 200 还可以包括多个处理器，因此，本说明书中披露的操作和/或方法步骤可以如本说明书的由一个处理器执行，也可以由多个处理器联合执行。例如，如果在本说明书中计算设备 200 的处理器 220 执行步骤 A 和步骤 B，则应该理解，步骤 A 和步骤 B 也可以由两个不同处理器 220 联合或分开执行（例如，第一处理器执行步骤 A，第二处理器执行步骤 B，或者第一和第二处理器共同执行步骤 A 和 B）。

图 3 示出了根据本说明书的实施例提供的一种数据异常检测方法的流程图；图 4 示出了根据本说明书的另一实施例提供的一种数据异常检测方法的流程图；该数据异常检测方法 P300 应用于数据异常检测系统 11。如前，系统 11 可以接收待检测交易成功率序列，并执行本说明书提供的数据异常检测方法 P300。如前，计算设备 200 可以执行本说明书的数据异常检测方法 P300。具体地，计算设备 200 中的处理器 220 可以读取存储在其本地存储介质中的指令集，然后根据指令集的规定，执行本说明书的数据异常检测方法 P300。如图 3 和图 4 所示，本说明书提供的方法 P300 可以包括 S310-S350，其中：

S310：获取待检测交易成功率序列。

其中，待检测交易成功率序列为根据一段时间内每一时刻的交易成功率按照时间顺序排列后构成的待检测交易成功率序列。待检测交易成功率序列可以为一段时间内某一支付方式对应的交易成功率序列。待检测交易成功率序列可以为一段时间内某一地区对应的交易成功率序列。待检测成功率序列可以为一段时间内某一支付方式在某一地区对应的交易成功率序列。应当理解，具体待检测交易成功率序列的序列类型可以根据实际情况灵活调整，并不以上述实施例给出的为限。

待检测支付成率序列包括第一子序列和第二子序列。第一子序列和第二子序列共同

构成待检测交易成功率序列，并且第一子序列和第二子序列所包含的成功率数据互不重复，且第一子序列对应的交易成功率均值高于第二子序列对应的交易成功率均值。

在一些可能的实施例中，第一子序列和第二子序列的确定方式可以为基于异常时间点确定的，异常时间点将待检测交易成功率序列切分为第一子序列和第二子序列。待检测交易成功率序列中：在异常时间点之前的一部分序列为第一子序列，在异常时间点之后的一部分序列为第二子序列。

异常时间点的确定方式可以为：在待检测交易成功率序列中确定第一时间点，第一时间点为交易成功率下降的时间点。对于每个第一时间点，确定待检测交易成功率序列在第一时间点前的方差和在第一时间点后的方差的之间的方差和，确定方差和最小的第一时间点为异常时间点。也就是说，第一时间点指示成功率下降的时间点，异常时间点指示成功率发生显著变化的时间点。

每个第一时间点均可以将待检测交易成功率序列切分为两个子序列，在第一时间点之前的子序列对应的交易成功率均值大于第一时间点之后的子序列对应的交易成功率均值。对于一个待检测交易成功率序列，第一时间点可能存在多个，也可能仅存在一个，具体第一时间点的个数本说明书在此不做任何限制，根据具体待检测交易成功率序列的情况确定。

也就是说，在确定待检测交易成功率序列中的异常时间点时，需要遍历待检测交易成功率序列中的每个时间点，在待检测交易成功率序列中确定交易成功率下降的第一时间点 t。对于每个第一时间点 t，计算时间点 0 到时间点 t 之间的滑动均值 mu(t)和方差和 sigma(t)，并计算时间点 t 到待检测交易成功率序列末尾 T 之间的滑动均值 r_mu(t) 和方差和 r_sigma(t)。基于上述计算结果，在待检测交易成功率序列中的各第一时间点中确定异常时间点 t*，使得异常时间点 t*满足 mu(t*) > r_mu(t*)且 r_sigma(t*) + sigma(t*)最小。由此得到的异常时间点可以将待检测交易成功率序列切分为两个子序列（第一子

序列和第二子序列），并且第一子序列对应的方差和第二子序列对应的方差之间的方差和，是第一时间点中前后序列之间的方差和最小的时间点。

上述在第一时间点中确定异常时间点的方式，限定了首先通过设置预设条件来对待检测交易成功率序列中的第一时间点进行筛选（第一时间点之前的成功率均值高于第一时间点之后的成功率均值）。上述筛选第一时间点的方式可以减少因随机波动导致的假阳性结果，只有当待检测交易成功率序列出现明显的下降趋势时，才会被识别为第一时间点，这有助于过滤掉非实质性的变化，并提高后续识别异常时间点的效率。此外，上述通过寻找第一时间点的两侧方差和最小的时间点，可以更准确地确定待检测交易成功率序列的发生显著变化的时间点为异常时间点。

采用上述方法可以准确在待检测交易成功率序列中确定第一时间点，并在第一时间点中进一步确定待检测交易成功率序列中最有可能发生显著变化的异常时间点，其中，异常时间点为交易成功率下降开始的时间点，不同的待检测交易成功率序列可能对应不同的异常时间点。也即，第二子序列的序列长度是根据异常时间点确定的，不同的待检测交易成功率序列对应的第二子序列进行异常检测的尺度可能是不同的，可能对于某些交易成功率序列对应的第二子序列的序列长度较长，可能交易成功率序列对应的第二子序列的序列长度较短。基于异常时间点对待检测交易成功率序列进行拆分，可以根据待检测交易成功率序列的实际情况确定异常时间点，以基于异常时间点准确地将待检测交易成功率序列拆分为交易正常的第一子序列和交易可能异常的第二子序列，实现了待检测交易成功率的自适应拆分。

为了从待检测交易成功率序列中获取有意义的信息，并减少噪声的影响。在本说明书的实施例中，确定第一时间点之前，需要基于预设聚合窗口对待检测交易成功率序列进行聚合，得到聚合后的平滑序列，并在平滑序列中确定第一时间点。

其中，对于待检测交易成功率序列中的任意时间点 t，重新根据预设聚合窗口的大

小计算预设聚合窗口内的交易成功率：SR = sum(SV) / sum(OV)，其中，SV 为交易成功量，OV 为交易量，sum(SV) 为预设聚合窗口内交易成功量之和， sum(OV)为预设聚合窗口内交易量之和。也就是说，对于一个预设聚合窗口内的交易成功率的确定方式，是根据该聚合窗口内的交易成功量总和与交易量总和的比值确定的，基于聚合后的结果确定聚合后的平滑的待检测交易成功率序列。相对于未聚合的序列，聚合后的平滑序列可以去除短期波动，使得待检测交易成功率序列的趋势更加明显。

在本说明书的实施例中，为了保证聚合效果，不同的交易量可能对应不同的预设聚合窗口的窗口大小。在对不同的待检测交易成功率序列进行聚合时，可以根据待检测交易成功率序列对应的交易量序列，确定预设聚合窗口的窗口大小。例如交易量较大的时候，每个时间点可能包含更多有价值的信息。因此在交易量较大的场景下，可以选择将预设聚合窗口的窗口大小设置的较大，以较大的预设聚合窗口来对待检测交易成功率序列进行聚合，从而更好地平滑交易成功率数据，减少噪声的影响。交易量较小的时候，每个时间点可能更容易收到偶然事件的影响。因此在交易量较小的场景下，可以将预设聚合窗口的窗口大小设置的较小，以较小的预设聚合窗口来更精确地捕捉上述事件。具体聚合窗口的窗口大小的设置可以根据实际情况灵活调整，并不以上述实施例给出的为限。

S330：根据第一子序列对应的第一交易量序列和第一子序列，建立交易成功率与交易量之间的统计分布模型。

其中，第一子序列为正常状态下的交易成功率的序列，可以将第一子序列作为基线序列，基于第一子序列和第一交易量序列，确定正常状态下的交易成功率与交易量之间的统计分布模型。

在本说明书的实施例中，第一交易量序列为交易量序列中的一部分，交易量序列为待检测成功率序列对应的交易量序列，也即交易量序列对应的时间跨度与待检测成功率

序列一致。异常时间点将交易量序列也切分为两个交易量序列，在异常时间点之前的为第一交易量序列，在异常时间点之后的为第二交易量序列。也就是说，第一交易量序列与第一子序列相对应，第二交易量序列与第二子序列相对应。

图 5 示出了根据本说明书的实施例提供的建立统计分布模型的流程图，如图 5 所示 S330 可包括 S331-S337：

S331：基于第一子序列和第一交易量序列，确定第一交易成功量序列。

其中，第一交易成功量序列为基于第一子序列和第一交易量序列的乘积确定的。举例说明：例如第一子序列和第一交易量序列对应的时间长度均为 0-t*，其中时刻 0 为序列开始的时刻，时刻 t*为异常时间点对应的时刻。在 0-t*内的 t1 时刻下，可以在第一子序列中获取 t1 时刻的交易成功率为 sr_t1，在第一交易量序列中获取 t1 时刻的交易量为 ov_t1，则 t1 时刻的交易成功量为 sv_t1=sr_t1*ov_t1。在确定了 t*-T 的每个时刻下的支付成功量之后，即可确定第一交易成功量序列。

在相同的时间段内，第一子序列、第一交易量序列和第一交易成功量序列之间可以相互推导，推导依据为：sv=sr*ov，其中，sv 为交易成功量，sr 为交易成功率，ov 为交易量。也即，已知相同的时间段内，第一子序列、第一交易量序列和第一交易成功量序列之中的任意两个序列，均可对另一个未知序列进行推导。例如，已知第一子序列和第一交易量序列，即可确定第一交易成功量序列：sv=sr*ov。已知第一交易成功量序列和第一子序列，即可确定第一交易量序列：ov=sv/sr。已知第一交易量序列和第一交易成功量序列，即可确定第一子序列：sr=sv/ov。

S333：基于第一交易成功量序列的总支付成功量和第一交易序列的总交易量的比值，确定第一子序列的总交易成功率，将第一子序列的总交易成功率作为统计分布模型的均值。

在本说明书的实施例中，统计分布模型的均值 mu=sum(sv)/sum(ov)。其中，sv 是第

一交易成功量序列中各个时间段的交易成功量（即交易成功的数量），ov 是第一交易量序列中的各个时间段的总交易量。sum(sv)为第一交易成功量序列的总支付成功量（各时刻的交易成功量之和），sum(ov)是第一交易量序列的总交易量（各时刻的交易量之和）。也就是说，mu 是第一子序列中所有交易成功率的加权平均值。

S335：基于第一子序列和第一子序列的总交易成功率确定统计分布模型的方差。

在本说明书的实施例中，统计分布模型的方差 sigma=sqrt(var(sr)) 。其中，sigma 表示第一子序列中交易成功率的标准差。var(sr)是 sr 的方差。

S337：基于统计分布模型的均值和方差，确定统计分布模型。

其中，统计分布模型包括贝塔（Beta）分布。图 6 示出了根据本说明书的实施例提供的建立统计分布模型的结构示意图，如图 6 所示，基于上述计算得到的 mu 和 sigma，可以计算出 Beta 分布的参数：第一分布参数 alpha 和第二分布参数 beta。Beta 分布是区间[0，1]上的连续概率分布，常用于建模随机变量，例如：成功率或比例。alpha 和 beta 是 Beta 分布的两个形状参数，它们决定了 Beta 分布的形态。本说明书中的第一子序列为基于异常时间点确定的，待检测成功率序列中在异常时间点之前序列为第一子序列，因此第一子序列表示的为正常情况下的成功率分布情况。将第一子序列和第一交易量作为基线数据，建立贝塔分布。贝塔分布将第一交易量序列和第一子序列耦合在了一起，建立了正常情况下成功率与交易量之间的统计关系。后续在对第二子序列进行异常检验时，可以基于统计分布的进行假设统计检验，以确定第二子序列在正常情况下应该对应怎样的成功率分布情况，而实际第二子序列的成功率分布情况是怎样的。基于确定结果，确定第二子序列是否存在异常。

每次交易是否成功可以看作是一次伯努利试验，多次独立的伯努利试验的累积结果遵循二项分布。因此在上述 Beta 分布明确了在给定时间 t 时的交易量 ov_t 和交易成功率 p 之间的先验分布，则在 t 时刻的交易成功量 sv_t 将服从二项分布。

在本说明书的实施例中，基于 Beta 分布的原理，由于 Beta 分布是二项分布的共轭先验，这意味着在对于新引入的待检测数据（第二子序列）后，后验分布仍保持 Beta 分布的形式，只是参数有所更新。在已知给定时间 t 时刻的交易量 ov_t 和 Beta 分布的参数（alpha 和 beta）时，就能计算出在 t 时刻的成功量 sv_t 的概率分布。所以在基于第一子序列和第一子序列的总交易成功率确定统计分布模型的参数之后，可以基于统计分布模型的参数和第二交易量序列，确定符合统计分布模型的交易成功量的预期概率分布，并基于第二子序列中的概率分布和预期概率分布，确定第二子序列是否异常。

S350：基于统计分布模型和第二子序列对应的第二交易量序列，对第二子序列进行检验，确定第二子序列是否异常。

其中，第二子序列的总交易成功率（sum（sr））包括第二子序列对应的总支付成功量（sum(sv)）与第二交易量序列对应的总交易量（sum(ov)）的比值 sum（sr）=sum(sv)/sum(ov)。

在本说明书的实施例中，确定第二子序列是否存在异常的方式可以为：基于预设的置信度阈值、统计分布模型和第二交易量序列，确定成功率阈值。基于第二子序列的总交易成功率以及成功率阈值，确定第二子序列是否异常。也即，本说明书中的成功率阈值是基于预设的置信度阈值、统计分布模型和第二交易量序列综合确定的，不但考虑了正常情况下的交易成功率与交易量之间的关系，并且综合考虑了第二子序列对应的第二交易量序列的交易量情况。不同的交易量可能对应不同的交易成功率，不同的第二交易量序列可能对应不同的成功率阈值。上述确定成功率阈值的方式，通过成功率阈值的设置限定了确定异常的边界，若第二子序列的总交易成功率低于边界，则认为第二子序列为异常子序列。并且上述确定成功率阈值的方式，实现了成功率阈值不是一成不变的，而是根据第二交易量序列的具体情况进行动态调整，使得成功率阈值能够更加灵活地适应交易量数据的变化，从而提高了异常检测的准确性。

其中，在第二子序列的总交易成功率大于成功率阈值的情况下，则说明第二子序列的总交易成功率符合统计分布模型，确定第二子序列没有异常。或者在第二子序列的总交易成功率小于成功率阈值的情况下，则确定第二子序列存在异常。

在一些可能的实施例中，确定成功率阈值的方式可以为：基于统计分布模型对第二子序列进行采样，获取样本，样本包括在第一预设数量的交易成功率对应的第二预设数量的交易量下的样本。确定预设置信度阈值在样本中对应的样本分位数，基于样本分位数对应的交易成功率和交易量，确定样本分位数对应的支付成功量。随后，基于样本分位数对应的支付成功量和第二交易量序列对应的总交易量的比值，确定成功率阈值。

在一些可能的实施例中，获取样本的方式可以为：基于统计分布模型进行第一预设数量（N）次采样，得到 N 个 p 的值，记为 p 1，p 2，……， p N。上述 p 的值是从贝塔分布中随机抽取的，代表了这段时间内交易成功率不同的可能值。对于每个 pi，从二项式分布（Binomial (pi，ovt)）分布中进行第二预设数量（ K）次采样，以得到 N×K 个交易成功量的采样值 sv 1，sv 2，……，sv N×K。其中，pi 表示每次实验成功的概率，ovt 表示从 t*到 T 的总交易量，也即第二子序列的总交易量。其中， t*为异常时间点，也即第二子序列的起始时间点。T 为待检测成功率序列的结束时间点，也即第二子序列的结束时间点。采样出来的 sv 的值是从二项分布中随机抽取的，代表了在不同交易成功率 pi 下可能的成功交易数量。随后，基于 sr*=sum(sr*ov)/sum(ov)确定 t*到 T 的总成功率，也即第二子序列的总成功率。

确定成功率阈值的方式可以为：预设置信度阈值为预设的常数值 a，例如 a 可以预设为 0.05，当然 a 也可以预设为 0.03、0.04 或任意其他预设置信度阈值。具体预设置信度阈值的设置数值本说明书在次不做任何限制，只需将预设置信度阈值设置的较低即可，低置信度表示当前需要关注的是序列分布的尾部，即低置信度区域。计算 N×K 个成功量采样值的 a 分位数 q，得到当前第二子序列对应的成功率阈值为 sr_threshold=q*ovt，其中 ovt 为第二子序列的总交易量。随后比较第二子序列的总成功率（sr*）与成功率阈

值（sr_threshold）之间的大小关系，如果 sr* < sr*threshold 则认为第二子序列是异常的。如果 sr* > sr*threshold 则认为第二子序列是正常的。

上述确定第二子序列是否存在异常的方式中，首先基于贝塔分布采样了第一预设数量个（N）可能的交易成功率的值，对于上述采样得到的每个交易成功率的值，均从二项式分布中采样第二预设数量个（K）对应的交易成功量的采样值。基于上述采样得到的 N*K 个交易成功量的采样值和预设置信度阈值，确定分位数 q，并基于分位数 q 和第二子序列对应的第二交易量序列确定成功率阈值，以基于第二子序列的总成功率和成功率阈值，确定第二子序列内的交易成功率是否异常。

综上，本说明书提供的数据异常检测方法 P300 和系统 11 中，在对待检测交易成功率序列进行检验时，会综合交易成功率与交易量之间的关系，来对待检测交易成功率序列是否存在异常进行检测。在本说明书提供的方法中，会基于待检测支付成功率序列中第一子序列对应的第一交易量序列和第一子序列，建立交易成功率与交易量之间的统计分布模型。基于统计分布模型和待检测交易成功率序列中的第二子序列对应的第二交易量序列，对第二子序列进行检验，以对第二子序列是否存在异常进行检验。这样的验证方式在验证过程中不但考虑了交易成功率，并且还考虑了交易量与交易成功率之间的关系，大大提高了检测的准确性和可靠性。

在本说明书一些可能的实施例中，在确定第二子序列的异常检测结果之后，还可以将第二子序列的检验结果发送至对应终端，以使对应终端即使对异常数据进行处理，保证系统。

在本说明书的实施例中，终端例如可以为运维终端或客户终端，具体终端的确定方式可以为基于异常数据的异常类型确定的。若异常类型指示由于商家侧原因导致的数据异常，则可以将异常数据发送至商家客户终端，以提醒商家及时对异常进行排查。若异常类型指示由于系统侧原因导致的数据异常，则可以将异常数据发送至运维终端，以提示运维人员即使对数据进行排查。具体异常数据的发送方式以及发送对象均可根据用户

需要灵活调整，并不以上述实施例给出的为限。

在本说明书的实施例中，进行数据异常检测的方法例如可以为分钟级的，也即每分钟均获取一段预设时间的历史支付成功率序列作为待检测支付成功率序列（例如获取过去十分钟的历史支付成功率序列作为待检测支付成功率序列，当然，也可以获取过去半小时或过去一小时、两小时的历史支付成功率序列作为待检测支付成功率序列，本说明书在此不做任何限制），并对待检测支付成功率序列中是否存在异常进行检测，随后将检测结果输出。并在下一分钟获取新的一段预设时间的历史支付成功率序列作为待检测支付成功率序列，再次进行检测。应当理解，上述实施例仅为示例性说明，具体数据异常检测的方法也可以为十分钟级的，也即每十分钟获取一段预设时间的历史支付成功率序列作为待检测支付成功率序列进行检测。当然，也可以为半小时级的，或一小时级的，具体数据异常检测的检测频率可以根据用户需要灵活调整，并不以上述实施例给出的为限。

本说明书提供的数据异常检测方法 P300 和系统 11 中，在对待检测交易成功率序列进行检验时，会综合考虑交易量对交易成功率的影响，来准确判断待检测交易成功率序列是否存在异常。在本说明书提供的方法中，会基于异常时间点将待检测支付成功率序列拆分为第一子序列和第二子序列。其中，异常时间点为根据待检测交易成功率序列的真实情况计算得到的，不同的待检测交易成功率序列可能对应不同的异常时间点。在基于异常时间点对待检测交易成功率序列进行切分后，基于第一子序列对应的第一交易量序列和第一子序列，建立交易成功率与交易量之间的统计分布模型。并基于预设置信度阈值和交易量确定当前第二子序列对应的成功率阈值，随后基于第二子序列的总成功率与成功率阈值之间的关系对第二子序列进行检验，以确定第二子序列是否存在异常。这样的验证方式在验证过程中统计分布模型是基于待检测支付成功率序列中的第一子序列确定的，更符合当前待检测支付成功率的分布规则。并且不同的交易量对应不同的成功率阈值，使得对第二子序列的检验综合考虑了第二交易量序列，使得本说明书提供的

方法可以根据第二交易量序列灵活调整成功率阈值，大大提高了检测的准确性和可靠性，提高了异常检测的自适应能力。

本说明书另一方面提供一种计算机可读非暂时性存储介质，存储有至少一个指令集，用来进行数据异常检测方法的可执行指令。当至少一个指令集被处理器执行时，至少一个指令集指导处理器实施本说明书的数据异常检测方法 P300 的步骤。在一些可能的实施方式中，本说明书的各个方面还可以实现为一种程序产品的形式，其包括程序代码。当程序产品在系统 11 上运行时，程序代码用于使系统 11 执行本说明书描述的方法 P300 的步骤。用于实现上述方法的程序产品可以采用便携式紧凑盘只读存储器（CD-ROM）包括程序代码，并可以在系统 11 上运行。然而，本说明书的程序产品不限于此，在本说明书中，可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统使用或者与其结合使用。程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。可读存储介质的更具体的例子包括：具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。计算机可读存储介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了可读程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。可读存储介质还可以是可读存储介质以外的任何可读介质，该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。可读存储介质上包含的程序代码可以用任何适当的介质传输，包括但不限于无线、有线、光缆、RF 等等，或者上述的任意合适的组合。可以以一种或多种程序设计语言的任意组合来编写用于执行本说明书操作的程序代

码，程序设计语言包括面向对象的程序设计语言—诸如 Java、C++等，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。

综上，本说明书提供的数据异常检测方法 P300 以及系统 11 能够实现根据第二交易量序列灵活调整成功率阈值，在对第二子序列的异常检测的过程中综合考虑了第二交易量序列的真实情况，使得本说明书提供的方法 P300 以及系统 11 可大大提高了检测的准确性和可靠性，提高了异常检测的自适应能力。

上述对本说明书特定实施例进行了描述。其他实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者是可能有利的。

综上，在阅读本详细公开内容之后，本领域技术人员可以明白，前述详细公开内容可以仅以示例的方式呈现，并且可以不是限制性的。尽管这里没有明确说明，本领域技术人员可以理解本说明书需求囊括对实施例的各种合理改变，改进和修改。这些改变，改进和修改旨在由本说明书提出，并且在本说明书的示例性实施例的精神和范围内。

此外，本说明书中的某些术语已被用于描述本说明书的实施例。例如，“一个实施例”，“实施例”和/或“一些实施例”意味着结合该实施例描述的特定特征，结构或特性可以包括在本说明书的至少一个实施例中。因此，可以强调并且应当理解，在本说明书的各个部分中对“实施例”或“一个实施例”或“替代实施例”的两个或更多个引用不一定都指代相同的实施例。此外，特定特征，结构或特性可以在本说明书的一个或多个实施例中适当地组合。

应当理解，在本说明书的实施例的前述描述中，为了帮助理解一个特征，出于简化本说明书的目的，本说明书将各种特征组合在单个实施例、附图或其描述中。然而，这

并不是说这些特征的组合是必须的，本领域技术人员在阅读本说明书的时候完全有可能将其中一部分设备标注出来作为单独的实施例来理解。也就是说，本说明书中的实施例也可以理解为多个次级实施例的整合。而每个次级实施例的内容在少于单个前述公开实施例的所有特征的时候也是成立的。

本文引用的每个专利，专利申请，专利申请的出版物和其他材料，例如文章、书籍、说明书、出版物、文件、物品等，除了与本文件不一致或相冲突的、或者对权利要求的最宽范围具有限制性影响的文件，均可以通过引用结合于此，并用于现在或以后与本文件相关联的所有目的。此外，在任何材料对相关术语的描述、定义和/或使用与本文件对相关术语的描述、定义和/或使用之间存在任何不一致或冲突时，以本文件中的术语为准。

最后，应理解，本文公开的申请的实施方案是对本说明书的实施方案的原理的说明。其他修改后的实施例也在本说明书的范围内。因此，本说明书披露的实施例仅仅作为示例而非限制。本领域技术人员可以根据本说明书中的实施例采取替代配置来实现本说明书中的申请。因此，本说明书的实施例不限于申请中被精确地描述过的实施例。

权　利　要　求　书

1.一种数据异常检测方法，所述方法包括：

获取待检测交易成功率序列；其中，所述待检测支付成率序列包括第一子序列和第二子序列；所述第一子序列对应的交易成功率均值高于所述第二子序列对应的交易成功率均值；

根据所述第一子序列对应的第一交易量序列和所述第一子序列，建立交易成功率与交易量之间的统计分布模型；以及

基于所述统计分布模型和所述第二子序列对应的第二交易量序列，对所述第二子序列进行检验，确定所述第二子序列是否异常。

2.如权利要求1所述的数据异常检测方法，其中，所述获取待检测交易成功率序列对应的第一子序列和第二子序列，包括：

在所述待检测交易成功率序列中确定第一时间点，其中，第一时间点为交易成功率下降的时间点；

对于每个第一时间点，确定所述待检测交易成功率序列在所述第一时间点前的方差和在所述第一时间点后的方差的方差和，确定方差和最小的第一时间点为所述异常时间点；以及

确定所述待检测交易成功率序列中在所述异常时间点之前的序列为所述第一子序列，在所述异常时间点之后的序列为所述第二子序列。

3.如权利要求2所述的数据异常检测方法，其中，所述在所述待检测交易成功率序列中确定第一时间点，包括：

基于预设聚合窗口对所述待检测交易成功率序列进行聚合，得到聚合后的平滑序列；以及

在所述平滑序列中确定所述第一时间点。

4.如权利要求3所述的数据异常检测方法，其中，根据所述待检测交易成功率序列对应的交易量序列，确定所述预设聚合窗口的窗口大小。

5.如权利要求1所述的数据异常检测方法，其中，所述基于所述统计分布模型和第二子序列对应的第二交易量序列，对所述第二子序列进行检验，确定所述第二子序列是否异常，包括：

基于预设的置信度阈值、所述统计分布模型和所述第二交易量序列，确定成功率阈值；

基于所述第二子序列的总交易成功率以及所述成功率阈值，确定所述第二子序列是否异常。

6.如权利要求5所述的数据异常检测方法，其中，所述基于所述第二子序列的总交易成功率以及所述成功率阈值，确定所述第二子序列是否异常，包括：

在所述第二子序列的总交易成功率大于所述成功率阈值的情况下，则确定所述第二子序列没有异常；或者

在所述第二子序列的总交易成功率小于所述成功率阈值的情况下，则确定所述第二子序列存在异常。

7.如权利要求5所述的数据异常检测方法，其中，所述基于预设置信度阈值、所述统计分布模型和所述第二交易量序列，确定成功率阈值，包括：

基于所述统计分布模型对所述第二子序列进行采样，获取样本，所述样本包括在第一预设数量的交易成功率对应的第二预设数量的交易量下的样本；

确定所述预设置信度阈值在所述样本中对应的样本分位数；

基于所述样本分位数对应的交易成功率和交易量，确定所述样本分位数对应的支付成功量；以及

基于所述样本分位数对应的支付成功量和所述第二交易量序列对应的总交易量的比值，确定所述成功率阈值。

8.如权利要求 5 所述的数据异常检测方法，其中，所述第二子序列的总交易成功率包括所述第二子序列对应的总支付成功量与所述第二交易量序列对应的总交易量的比值。

9.如权利要求 1 所述的数据异常检测方法，其中，所述根据所述第一子序列对应的第一交易量序列和所述第一子序列，建立交易成功率与交易量之间的统计分布模型，包括：

基于所述第一子序列和所述第一交易量序列，确定第一交易成功量序列；

基于所述第一交易成功量序列的总支付成功量和所述第一交易序列的总交易量的比值，确定所述第一子序列的总交易成功率，将所述第一子序列的总交易成功率作为所述统计分布模型的均值；

基于所述第一子序列和所述第一子序列的总交易成功率确定所述统计分布模型的方差；以及

基于所述统计分布模型的均值和方差，确定所述统计分布模型。

10.如权利要求 9 所述的数据异常检测方法，其中，所述统计分布模型包括贝塔分布。

11.如权利要求 1 所述的数据异常检测方法，其中，所述方法还包括：

将所述第二子序列的检验结果发送至对应终端。

12.一种数据异常检测系统，包括：

至少一个存储介质，存储有至少一个指令集，用于进行数据异常检测方法；以及

至少一个处理器，同所述至少一个存储介质通信连接，其中，所述至少一个处理器运行时读取所述至少一个指令集，并根据所述至少一个指令集的指示执行：

获取待检测交易成功率序列；其中，所述待检测支付成率序列包括第一子序列和第二子序列；所述第一子序列对应的交易成功率均值高于所述第二子序列对应的交易成功率均值；

根据所述第一子序列对应的第一交易量序列和所述第一子序列，建立交易成功率与交易量之间的统计分布模型；以及

基于所述统计分布模型和所述第二子序列对应的第二交易量序列，对所述第二子序列进行检验，确定所述第二子序列是否异常。

13.如权利要求 12 所述的数据异常检测系统，其中，为了获取待检测交易成功率序列对应的第一子序列和第二子序列，所述至少一个处理器：

在所述待检测交易成功率序列中确定第一时间点，其中，第一时间点为交易成功率下降的时间点；

对于每个第一时间点，确定所述待检测交易成功率序列在所述第一时间点前的方差和在所述第一时间点后的方差的方差和，确定方差和最小的第一时间点为所述异常时间点；以及

确定所述待检测交易成功率序列中在所述异常时间点之前的序列为所述第一子序列，在所述异常时间点之后的序列为所述第二子序列。

14.如权利要求 13 所述的数据异常检测系统，其中，为了在所述待检测交易成功率序列中确定第一时间点，所述至少一个处理器：

基于预设聚合窗口对所述待检测交易成功率序列进行聚合，得到聚合后的平滑序列；以及

在所述平滑序列中确定所述第一时间点。

15.如权利要求 14 所述的数据异常检测系统，其中，根据所述待检测交易成功率序列对应的交易量序列，确定所述预设聚合窗口的窗口大小。

16.如权利要求 12 所述的数据异常检测系统，其中，为了基于所述统计分布模型和第二子序列对应的第二交易量序列，对所述第二子序列进行检验，确定所述第二子序列是否异常，所述至少一个处理器：

基于预设的置信度阈值、所述统计分布模型和所述第二交易量序列，确定成功率阈值；

基于所述第二子序列的总交易成功率以及所述成功率阈值，确定所述第二子序列是否异常。

17.如权利要求 16 所述的数据异常检测系统，其中，为了基于所述第二子序列的总交易成功率以及所述成功率阈值，确定所述第二子序列是否异常，所述至少一个处理器：

在所述第二子序列的总交易成功率大于所述成功率阈值的情况下，则确定所述第二子序列没有异常；或者

在所述第二子序列的总交易成功率小于所述成功率阈值的情况下，则确定所述第二子序列存在异常。

18.如权利要求 16 所述的数据异常检测系统，其中，为了基于预设置信度阈值、所述统计分布模型和所述第二交易量序列，确定成功率阈值，所述至少一个处理器：

基于所述统计分布模型对所述第二子序列进行采样，获取样本，所述样本包括在第一预设数量的交易成功率对应的第二预设数量的交易量下的样本；

确定所述预设置信度阈值在所述样本中对应的样本分位数；

基于所述样本分位数对应的交易成功率和交易量，确定所述样本分位数对应的支付成功量；以及

基于所述样本分位数对应的支付成功量和所述第二交易量序列对应的总交易量的比值，确定所述成功率阈值。

19.如权利要求 16 所述的数据异常检测系统，其中，所述第二子序列的总交易成功

率包括所述第二子序列对应的总支付成功量与所述第二交易量序列对应的总交易量的比值。

20.一种计算机可读非易失性存储介质，其中，所述计算机可读非易失性存储介质中存储有至少一个指令集，所述至少一个指令集被至少一个处理器执行时，实现如权利要求 1-11 中任一项所述的数据异常检测方法。