说　　明　　书

数据处理方法、系统及存储介质

技术领域

本说明书涉及数据处理技术领域，尤其涉及一种数据处理方法、系统及存储介质。

背景技术

目前很多应用场景需要部署风控系统来进行风险防控。以交易场景为例，为了保证交易的安全性，通常在交易平台中部署风控系统，风控系统中集成有风控策略。当交易平台产生交易时，风控系统可以基于上述风控策略识别该交易是否为风险交易。

上述风控策略可以为一系列的风控规则。风控规则通常是由风控人员基于已经发生的风险交易总结风险模式，进而利用自身经验对总结得到的风险模式进行整理得到的。上述风控策略还可以为预先训练得到的风控模型。风控模型的训练过程比如为，从大量历史交易中标注出风险交易和非风险交易，利用这些历史交易及其标注信息进行模型训练，使得训练后的模型具有风险交易的识别能力。

由上述可见，上述风控策略无论是风控规则还是风控模型，都是从已知的风险交易中总结/学习得到的。而从已知的风险交易中总结/学习得到的风险防控经验是具有滞后性的，即滞后于风险交易的发生。这使得风控策略只能防控已知风险，而不具有对未知风险的防控能力。因此，上述风控策略存在漏防未知风险的可能。

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公开申请日之前已经进入公共领域，也不代表其可以成为本公开的现有技术。

发明内容

本说明书提供一种数据处理方法、系统及存储介质，将多个历史事件作为初始种群，通过模拟种群进化过程对初始种群进行多代繁殖得到具有未知风险的目标事件，从而目标事件可用于更新当前风控策略，使得更新后的风控策略具有防控未知风险的能力，改善风控经验滞后的问题。

第一方面，本说明书提供一种数据处理方法，包括：获得多个历史事件，所述多个历史事件中的至少部分事件为具有已知风险的事件，所述已知风险为当前风控策略 能够识别的风险；将所述多个历史事件作为初始种群，通过模拟种群进化过程对所述初始种群进行 N 代繁殖得到目标种群，其中，在所述 N 代繁殖的过程中所述已知风险被进化为所述当前风控策略不能识别的未知风险，所述 N 为大于或等于 1 的整数；以及在所述目标种群中确定出具有所述未知风险的至少一个目标事件，所述至少一个目标事件用于所述当前风控策略的更新过程。

第二方面，本说明书还提供一种数据处理系统，包括：至少一个存储介质，存储有至少一个指令集，用于对数据进行处理；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，所述数据处理系统运行时，所述至少一个处理器读取所述至少一个指令集，并且根据所述至少一个指令集的指示：获得多个历史事件，所述多个历史事件中的至少部分事件为具有已知风险的事件，所述已知风险为当前风控策略能够识别的风险；将所述多个历史事件作为初始种群，通过模拟种群进化过程对所述初始种群进行 N 代繁殖得到目标种群，其中，在所述 N 代繁殖的过程中所述已知风险被进化为所述当前风控策略不能识别的未知风险，所述 N 为大于或等于 1 的整数；以及在所述目标种群中确定出具有所述未知风险的至少一个目标事件，所述至少一个目标事件用于所述当前风控策略的更新过程。

第三方面，本说明书还提供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，实现第一方面中提供的数据处理方法。

本说明书提供的数据处理方法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处理方法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风控场景示意图；

图 2 示出了根据本说明书的实施例提供的一种计算系统 200 的硬件结构图；

图 3 示出了根据本说明书的实施例提供的一种数据处理方法的流程图；

图 4 示出了根据本说明书的实施例提供的一种模拟种群进化过程的流程图；

图 5 示出了根据本说明书的实施例提供的第 i 代的繁殖过程的示意图；

图 6 示出了根据本说明书的实施例提供的一种进行交叉操作的示意图；

图 7 示出了根据本说明书的实施例提供的一种增加属性的变异操作的示意图；

图 8 示出了根据本说明书的实施例提供的一种删除属性的变异操作的示意图；以及

图 9 示出了根据本说明书的实施例提供的一种更改属性的变异操作的示意图。

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。

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。

术语 1：遗传算法。遗传算法是一种基于自然选择和遗传机制的算法，属于进化计算（Evolutionary Computation）的一种。它通过模拟生物进化过程中的选择、交叉和变异等机制来搜索最优解。遗传算法广泛通常应用于各种复杂的优化问题，如函数优化、路径规划、机器学习等。

下面对本说明书的应用场景进行介绍。

本说明书提供的技术方案适用于需要进行风险防控的场景。以交易场景为例，为了保证交易的安全性，通常在交易平台中部署风控系统，风控系统中集成有风控策略。当交易平台产生交易时，风控系统可以基于上述风控策略识别该交易是否为风险交易。

风控策略是风控系统能否准确识别风险交易的关键因素。目前风控策略通常是通过对已知的风险交易进行总结/学习得到的，能够识别存在已知风险的事件，而不具有对未知风险的防控能力。

本说明书提供了一种数据处理方法，可以由数据处理系统执行。数据处理系统将多个历史事件作为初始种群，通过模拟种群进化过程对初始种群进行 N 代繁殖得到目标种群。其中，上述多个历史事件中的至少部分事件具有已知风险（即能够被当前风控策略识别的风险），在上述 N 代繁殖过程中，上述已知风险被进化为未知风险（即不能被当前风控策略识别的风险）。进而，数据处理系统从目标种群中确定具有未知风险的至少一个目标事件。上述至少一个目标事件可以被用于当前风控策略的更新过程。这样，更新后的风控策略具有对未知风险的防控能力。

需要说明的是，上述交易场景只是本说明书提供的多个使用场景中的一个，本说明书提供的数据处理方法不仅可以应用于交易场景，还可以应用于需要进行风险防控的所有场景，比如，病毒防控场景、防火墙防控场景等等。本领域技术人员应当明白，当本说明书提供的数据处理方法应用于其他使用场景时，其实现方式及技术效果是类似的。

图 1 示出了根据本说明书的实施例提供的一种风控场景示意图。如图 1 所示，该场景 100 可以包括数据处理系统 11 以及风控系统 12。场景 100 涉及数据处理阶段和风控阶段。

其中，在数据处理阶段，数据处理系统 11 可以基于多个历史事件，通过模拟种群进化过程生成具有未知风险的至少一个目标事件。上述至少一个目标事件可以被用于更新当前风控策略，使得更新后的风控策略具有对未知风险的防控能力。上述更新后的风控策略可以被部署到风控系统 12 中。在风控阶段，风控系统 12 可以利用更新后的风控

策略进行风险防控。

数据处理系统 11 可以是具有一定计算能力的计算系统。数据处理系统 11 中可以包括一个或者多个计算设备。本说明书提供的数据处理方法可以由数据处理系统 11 执行。数据处理系统 11 可以存储有执行本说明书描述的数据处理方法的数据或指令，并可以执行或用于执行所述数据或指令。在一些实施例中，数据处理系统 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。数据处理系统 11 可以将多个历史事件作为初始种群，通过模拟种群进化过程对初始种群进行 N 代繁殖得到目标种群。其中，上述多个历史事件中的至少部分事件具有已知风险（即能够被当前风控策略识别的风险），在上述 N 代繁殖的过程中，所述已知风险被进化为未知风险（即不能被当前风控策略识别的风险）。进而，数据处理系统 11 可以从目标种群中确定可能具有未知风险的至少一个目标事件。

风控系统 12 可以是具有一定计算能力的计算系统。如图 1 所示，风控系统 12 中可以配置有更新后的风控策略，并利用更新后的风控策略进行风险防控。

在一些实施例中，基于所述至少一个目标事件对当前风控策略进行更新的过程可以由数据处理系统 11 执行。比如，数据处理系统 11 得到所述至少一个目标事件之后，可以基于所述至少一个目标事件对所述当前风控策略进行更新，并将更新后的风控策略提供给风控系统。在一些实施例中，基于所述至少一个目标事件对当前风控策略进行更新的过程可以由风控系统 12 执行。比如，数据处理系统 11 得到所述至少一个目标事件之后，可以将所述至少一个目标事件通过给风控系统 12。风控系统 12 可以根据所述至少一个目标事件，更新风控系统 12 中当前风控策略。

数据处理系统 11 和风控系统 12 可以对应于相同的实体系统，也可以对应于不同的实体系统。当数据处理系统 11 和风控系统 12 对应不同的实体系统时，风控系统 12 可以通过网络与数据处理系统 11 通信连接。当数据处理系统 11 和风控系统 12 对应于同一实体系统时，风控系统 12 可以通过该实体系统中的总线与数据处理系统 11 通信连

接。

应该理解，图 1 中的数据处理系统 11、风控系统 12 的数目仅仅是示意性的。根据实现需要，例如，一个数据处理系统 11 得到的目标事件可以用于对多个风控系统 12 中的风控策略进行更新，这个情况下，可以具有任意数目的风控系统 12。例如，通过分布式的设备组执行数据处理方法时，数据处理方法可以在多个数据处理系统 11 中分布式执行，这个情况下，可以具有任意数目的数据处理系统 11。

图 2 示出了根据本说明书的实施例提供的一种计算系统 200 的硬件结构图。计算系统 200 可以作为图 1 中的数据处理系统 11，执行本说明书描述的数据处理方法。

如图 2 所示，计算系统 200 可以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，计算系统 200 还可以包括通信端口 250 和内部通信总线 210。计算系统 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。例如，内部通信总线 210 可以连接存储介质 230、处理器 220、通信端口 250 和 I/O 组件 260 等。

I/O 组件 260 支持计算系统 200 和其他组件之间的输入/输出。

通信端口 250 用于计算系统 200 同外界的数据通信。比如，通信端口 250 可以用于计算系统 200 同网络之间的数据通信。通信端口 250 可以是有线通信端口也可以是无线通信端口。

存储介质 230 可以包括数据存储装置。所述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，所述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 还包括存储在所述数据存储装置中的至少一个指令集。所述指令集可以包括计算机程序代码，所述计算机程序代码可以包括程序、例程、对象、组件、数据结构、过程、模块等等。

至少一个处理器 220 可以同至少一个存储介质 230 通信连接。当计算系统 200 运行

时，至少一个处理器 220 读取所述至少一个指令集，并且根据所述至少一个指令集的指示，执行本说明书提供的数据处理方法。处理器 220 可以执行数据处理方法包含的步骤。处理器 220 可以是一个或多个处理器的形式，在一些实施例中，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，附图中计算系统 200 仅示出了一个处理器 220。然而，应当注意，本说明书中计算系统 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，如果在本说明书中描述计算系统 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处理方法的流程图。如前，数据处理系统 11 可以执行本说明书的数据处理方法。

如图 3 所示，数据处理方法可以包括：

S310：获得多个历史事件，多个历史事件中的至少部分事件为具有已知风险的事件，已知风险为当前风控策略能够识别的风险。

历史事件可以是在历史时刻发生的事件。一个历史事件的事件信息可以包括多个属性-属性值的键值对。需要说明的是，在不同的应用场景中，历史事件的指代、以及历史事件中的属性可能有所不同。比如，在交易场景中，历史事件可以为历史时刻发生的交易事件。交易事件的属性可以包括交易参与双方的账号、设备信息、环境信息、以及交易过程信息等。为了方便理解，在后文涉及举例时，均以交易场景为例进行举例说明。

风控系统中可以部署风控策略，并根据当前风控策略对事件进行风控。例如，将一个事件的事件信息（多个属性-属性值的键值对）输入风控系统，风控系统根据上述事件信息识别该事件是否存在风险，或者识别该事件存在何种风险等。

在本说明书中，当前风控策略是指风控系统中当前部署的风控策略。当前风控策略通常是基于已知的风险事件进行总结/学习得到的。当前风控策略的获得方式可以参加前文相关描述，此处不做赘述。当前风控策略可以包括风控规则、风控模型中的一种或多种。另外，本说明书将风控策略能够识别的风险称为已知风险，并将风控策略不能识别的风险称为未知风险。

在 S310 中，数据处理系统获得的多个历史事件中，至少一部分事件为具有已知风险的事件（称为黑样本），剩余事件为不具有已知风险的事件。例如，多个历史事件中，可以包括 1%至 10%的黑样本。

S320：将多个历史事件作为初始种群，通过模拟种群进化过程对初始种群进行 N 代繁殖得到目标种群，其中，在所述 N 代繁殖的过程中所述已知风险被进化为所述当前风控策略不能识别的未知风险，所述 N 为大于或等于 1 的整数。

在一些实施例中，数据处理系统可以通过遗传算法或者其他类似算法来模拟种群进化过程。后续以遗传算法为例进行举例说明。

图 4 示出了根据本说明书的实施例提供的一种模拟种群进化过程的流程图。参考图 4，在遗传算法中，数据处理系统可以将多个历史事件作为初始种群，通过对初始种群进行 N 代繁殖，模拟种群进化过程，最终得到目标种群。例如，将初始种群作为第 1 代种群，数据处理系统在第 1 代种群的基础上，通过模拟种群进化过程生成第 2 代种群。进而，数据处理系统在第 2 代种群的基础上，通过模拟种群进化过程生成第 3 代种群。以此类推，数据处理系统在第 N-1 代种群的基础上，通过模拟种群进化过程生成第 N 代种群。第 N 代种群即为目标种群。由于初始种群中包括了已知风险的事件，因此，在模拟种群进化过程的过程中，已知风险被进化为未知风险，从而目标种群中存在包含未知风

险的事件。

初始种群中包括具有已知风险的事件。通过图 4 所示的 N 代繁殖，数据处理系统可以在繁殖过程中将已知风险进化为未知风险，使得到的目标种群中包括具有未知风险的目标事件。这些具有未知风险的目标事件可以用于更新当前风控策略，从而使得更新后的风控策略能够实现对未知风险的识别，减少漏防未知风险的概率，提高风控效果。

本领域技术人员能够理解，上述 N 代繁殖的实现方式是类似的。下面对第 i 代的繁殖过程为例进行具体说明。

自然界的种群进化过程通常存在基因遗传、基因变异、自然选择等环节。为了模拟自然界的种群进化过程。在一些实施例中，数据处理系统在模拟种群进化过程时，可以采用交叉、变异、淘汰等操作。其中，交叉操作是指根据当前种群中的两个不同事件，生成一个进化事件的操作。交叉操作可以模拟种群进化过程中的基因遗传。当前种群可以是初始种群，可以是繁殖过程中的某一代种群。通过交叉操作可以根据当前种群繁殖得到下一代种群的进化事件。变异操作是指随机对进化事件的属性进行变异操作，模拟种群进化过程中的基因变异。淘汰操作是指将种群中不满足预设的风险进化条件的进化事件淘汰掉，模拟种群进化过程中的自然选择。

图 5 示出了根据本说明书的实施例提供的第 i 代的繁殖过程的示意图。如图 5 所示，第 i 代的繁殖过程可以包括：

（1）对当前种群中的不同事件进行交叉操作，以生成多个进化事件。

其中，当$i = 1$时，当前种群为初始种群。当$1 < i \leq N$时，当前种群为第$i - 1$代繁殖后的种群。

一个事件对应有多个属性。在不同的应用场景下，事件所对应的属性可能有所不同。例如，以交易场景为例，一个事件对应的多个属性可以包括下述中的一项或多项：交易参与双方的账号、设备信息、环境信息、以及交易过程信息。

在一些实施例中，一个事件所对应的多个属性还可以包括风险属性。风险属性对应

的属性值表征该事件是否具有已知风险。

在一些实施例中，风险属性对应的属性值可以作为该事件的风险标签。例如，初始种群包括的多个历史事件中，一部分事件为具有已知风险的事件，则该部分事件中的风险属性对应的属性值为1（表示具有已知风险），剩余事件中的风险属性对应的属性值为0（表示不具有已知风险）。

或者，也可以将多个属性中的某个能够用于表征具有已知风险的属性作为风险属性。例如，当某个交易参与方处于黑名单时，其参与的交易为具有已知风险的交易，则可以将该交易参与方的账号作为风险属性。

进化事件是通过对当前种群中的不同事件进行交叉操作得到的。具体来说，一个进化事件的一部分属性的属性值来自第一事件，其他属性的属性值来自第二事件。第一事件和第二事件为当前种群中的任意两个不同事件。

在一些实施例中，数据处理系统对当前种群执行交叉操作，可以包括：数据处理系统在当前种群中执行 M 次以下操作：数据处理系统从当前种群中选定第一事件和第二事件，第一事件和第二事件为当前种群中的不同事件。数据处理系统根据第一事件和第二事件得到至少一个进化事件，进化事件的部分属性的属性值来自第一事件，进化事件其他属性的属性值来自第二事件。其中，M 为大于或等于 1 的整数。

在一些实施例中，数据处理系统根据第一事件和第二事件得到每个进化事件时，可以参考图 6 示出的方式。图 6 示出了根据本说明书的实施例提供的一种进行交叉操作的示意图。

参考图 6，第一事件包括多个属性，第二事件的属性与第一事件相同，根据第一事件和第二事件生成的进化事件也具有相同的属性。数据处理系统可以随机从第一事件中获取至少一个属性的属性值，从第二事件中获取剩余属性的属性值，并根据获取到的属性值生成一个进化事件。

参考图 6，第一事件的属性包括风险属性、属性 1、属性 2、属性 3、属性 4、属性 5。其中，每个属性包括一个对应的属性值，参考图 5，风险属性的属性值为“属性值 A”，属性 1 的属性值为“属性 1a”，属性 2 的属性值为“属性 2a”，属性 3 的属性值为“属性 3a”，属性 4 的属性值为“属性 4a”，属性 5 的属性值为“属性 5a”。

参考图 6，第二事件的属性与第一事件相同，但是具有不同的属性值。例如，风险属性的属性值为“属性值 B”，属性 1 的属性值为“属性 1b”，属性 2 的属性值为“属性 2b”，属性 3 的属性值为“属性 3b”，属性 4 的属性值为“属性 4b”，属性 5 的属性值为“属性 5b”。

参考图 6，作为示例，数据处理系统可以随机获取第一事件的风险属性的属性值“属性 A”、属性 1 的属性值“属性 1a”以及属性 4 的属性值“属性 4a”；从第二事件中获取属性 2 的属性值“属性 2b”、属性 3 的属性值“属性 3b”、属性 5 的属性值“属性 5b”。数据处理系统可以根据以上获取到的属性值，生成进化事件。也就是说，进化事件的风险属性的属性值为“属性 A”、属性 1 的属性值为“属性 1a”、属性 2 的属性值为“属性 2b”、属性 3 的属性值为“属性 3b”、属性 4 的属性值为“属性 4a”以及属性 5 的属性值为“属性 5b”。

上述方案中，通过对当前种群中不同事件进行交叉操作，能够生成全新的进化事件。在生成进化事件的过程中，当前种群中的事件中的风险可能发生进化，从而有助于将已知风险进化为未知风险。

由图 6 所示的交叉操作可知，进化事件可以视为通过将第一事件中的部分属性的属性值替换为第二事件中的对应属性的属性值得到的。例如，结合图 6，进化事件可以视为通过将第一事件中的属性 2、属性 3、属性 5 的属性值替换为第二事件中的属性 2、属性 3、属性 5 的属性值得到的。

或者，进化事件也可以视为通过将第二事件中的部分属性的属性值替换为第一事件中的对应属性的属性值得到的。例如，结合图 6，进化事件可以视为通过将第二事件中

的风险属性、属性 1、属性 4 的属性值替换为第一事件中的风险属性、属性 1、属性 4 的属性值得到的。

在一些实施例中，为了提高已知风险的进化效率，数据处理系统可以通过将具有已知风险的事件中的一部分属性的属性值替换为其他事件中的对应属性的属性值，从而生成多个进化事件。也就是说，数据处理系统在对不同事件中的属性值进行替换时，可以重点将具有已知风险的事件作为被替换的事件（即图 6 中示出的第一事件）。被替换的部分属性包括除风险属性之外的至少一个属性。例如，依然以图 6 中的两个事件为例，在生成进化事件的过程中，第一事件中的风险属性的属性值不被替换，而将属性 2、属性 3、属性 5 的属性值替换为第二事件中的属性 2、属性 3、属性 5 的属性值。

在一些实施例中，在上述替换过程中，第一事件中的被替换属性的数量与第一事件中所有属性的数量之比小于或等于预设阈值。例如，假设第一事件中所有属性的数量为 10，则被替换属性的数量可以小于或等于 5。也就是说，在对具有已知风险的第一事件进行属性值替换时，可以替换较少的属性值，而保留较多的原始属性值。这样的进化过程更加符合自然进化的规律。从而，第一事件中的风险属性的属性值能够被较为准确的遗传到进化事件中。

上述方案中，通过将具有已知风险的第一事件中的部分属性值替换为第二事件中的对应属性的属性值，可以提高已知风险的进化效率。

（2）从多个进化事件中淘汰掉部分事件，部分事件不满足预设的风险进化条件。

在一些实施例中，预设的风险进化条件用于筛选已知风险和未知风险。例如，满足预设的风险进化条件的事件为可能存在未知风险的事件，不满足预设的风险进化条件的事件为具有已知风险的事件。

在一些实施例中，风险进化条件包括：当前风控策略将进化事件识别为无风险事件。

作为示例，当前风控策略为风控模型时，数据处理系统可以将进化事件的事件信息输入风控模型并接收风控模型反馈的识别结果。若风控模型将进化事件识别为风险事件，

则说明进化事件中的风险能够被当前风控策略识别，为具有已知风险的事件；若风控模型将进化事件识别为无风险事件，则说明进化事件无风险或存在无法识别的风险，即进化事件有可能为具有未知风险的事件。将具有已知风险的事件淘汰掉，可以使剩余的事件中可能具有未知风险的事件的比例提高，进一步提高风险进化的效率。

在一些实施例中，由于根据当前种群生成的进化事件数量可能过多，不断繁殖的情况下，可能会导致最终的目标种群数量过于庞大。因此，数据处理系统可以对淘汰事件的数量进行限制，以使得最终目标种群中的进化事件更有可能为具有未知风险的事件。

在一些实施例中，多个进化事件中被淘汰的事件数量可以满足如下要求：淘汰后剩余的事件数量与初始种群中的事件数量之差的绝对值小于预设阈值。其中，预设阈值可以是大于或等于 0 的整数。例如，当预设阈值为 0 时，淘汰后剩余的事件数量与初始种群中的事件数量相等。当预设阈值为 100 时，淘汰后剩余的事件数量可以比初始种群中的事件数量多 100 件，或比初始种群中的事件数量少 100 件。

在一些实施例中，数据处理系统可以在进行淘汰之前，对多个进化事件中的至少部分进化事件进行变异操作，得到包括多个变异事件的进化事件集合。然后，数据处理系统从进化事件集合中淘汰掉部分事件。

在一些实施例中，一个进化事件对应的变异操作包括下述中的至少一种：随机增加进化事件的至少一个属性的属性值；随机删除进化事件除风险属性以外的至少一个属性的属性值；随机更改进化事件除风险属性以外的至少一个属性的属性值。

图 7 示出了根据本说明书的实施例提供的一种增加属性的变异操作的示意图。

在一些实施例中，参考图 7，进化事件的属性包括风险属性、属性 1、属性 2、属性 3、属性 4 以及属性 5。数据处理系统执行变异操作时，可以对进化事件随机增加至少一个属性，该属性的属性值可以是随机生成的，也可以是从其他进化事件中获取的。

作为示例，数据处理系统对进化事件随机增加两个属性，即属性 6（对应的属性值为属性值 6c）和属性 7（对应的属性值为属性值 7c）。在增加两个属性后，原来的进化

事件即变异为变异事件。变异事件具有风险属性、属性 1、属性 2、属性 3、属性 4、属性 5、属性 6 以及属性 7。

图 8 示出了根据本说明书的实施例提供的一种删除属性的变异操作的示意图。

在一些实施例中，参考图 8，进化事件的属性包括风险属性、属性 1、属性 2、属性 3、属性 4 以及属性 5。数据处理系统执行变异操作时，可以对进化事件随机删除风险属性以外的至少一个属性。

作为示例，数据处理系统随机删除一个属性，例如，属性 5。在删除一个属性后，原来的进化事件即变异为变异事件，变异事件具有风险属性、属性 1、属性 2、属性 3 以及属性 4。

图 9 示出了根据本说明书的实施例提供的一种更改属性的变异操作的示意图。

在一些实施例中，参考图 9，进化事件的属性包括风险属性、属性 1、属性 2、属性 3、属性 4 以及属性 5。数据处理系统执行变异操作时，可以对进化事件随机更改风险属性以外的至少一个属性的属性值。

作为示例，数据处理系统随机删除两个属性的属性值。例如，参考图 9，可以将属性 1 的属性值从“属性值 1b”更改为“属性值 1c”、将属性 4 的属性值从“属性值 4b”更改为“属性值 4c”，在更改两个属性的属性值后，原来的进化事件即变异为变异事件

需要说明的是，对于一个进化事件，数据处理系统可以仅执行增加、删除或变更中的一种变异操作，也可以执行其中任意两个变异操作，或者将三种变异操作均执行一遍。

对进化事件进行变异操作，可以提高进化事件进化为未知风险的概率，进一步提高生成未知风险的效率。

（3）将多个进化事件中剩余的事件作为第*i*代繁殖后的种群。

上文中的（1）-（3）描述的是第 i 代繁殖的过程。数据处理系统按照上述繁殖过程执行第 1 代繁殖、第 2 代繁殖、第 3 代繁殖、…、第 N 代繁殖。从而，得到的第 N 代繁殖后的种群即为目标种群。

通过 N 代繁殖，数据处理系统可以在繁殖过程中将初始种群中的已知风险进化为未知风险，使得到的目标种群中可能会包括具有未知风险的事件，将这些具有未知风险的事件用于当前风控策略的更新过程，以实现对未知风险进行识别，减少漏放未知风险的概率，提高风控效果。

S330：在目标种群中确定出具有未知风险的至少一个目标事件，至少一个目标事件用于当前风控策略的更新过程。

在一些实施例中，在 N 代繁殖的过程中，多个历史事件各自对应的风险属性的属性值被逐代遗传至目标种群中的事件。风险属性的属性值用于表示该事件是否具有已知风险（或者说表示该事件是否由具有已知风险的事件进化而成的事件）。数据处理系统可以基于风险属性的属性值，在目标种群中筛选可能具有未知风险的至少一个目标事件。

在一些实施例中，在目标种群中确定出具有未知风险的至少一个目标事件，包括：针对目标种群中的每个事件，若事件对应的风险属性的属性值表征具有已知风险，则将事件确定为目标事件。

在一些实施例中，风险属性的属性值可以具体表征发起该历史事件的账号是否属于黑名单账号。作为示例，风险属性的属性值可以包括“属于黑名单账号”或“不属于黑名单账号”。在目标种群中，当一个事件的账号属性的属性值为“属于黑名单账号”时，则认为该事件可能存在未知风险。这个情况下，数据处理系统可以将目标种群中，风险属性的属性值为“属于黑名单账号”的事件确定为目标事件。

在上述方案中，通过将具有已知风险的事件（风险属性的属性值表征具有已知风险）中的部分属性的属性值替换为其他事件中的对应属性的属性值来生成进化事件，一方面能够提高已知风险的进化效率，另一方面使得具有已知风险的事件中的风险属性被逐代遗传至目标种群。进而，数据处理系统可以基于风险属性的属性值在目标种群中确定目标事件，从而能够较为方便快捷的在目标种群中定位出具有未知风险的目标事件。

在一些实施例中，数据处理系统可以通过人工、大模型、机器学习等方式，从目标种群中筛选得到具有未知风险的目标事件。

目标事件可以用于当前风控策略的更新过程，使得更新后的风控策略具有识别未知风险的能力。例如，当前风控策略包括风控模型时，数据处理系统或者风控系统可以至少利用至少一个目标事件对风控模型进行训练，得到训练后的风控模型，训练后的风控模型具有识别未知风险的能力。例如，当前风控策略包括风控规则时，数据处理系统或者风控系统或者风控人员可以基于所述至少一个目标事件进行风险模式的分析，以总结得到能够识别未知风险的风控规则。

综上所述，本说明书提供的数据处理方法和系统中，将历史事件作为初始种群，通过模拟种群进化过程对初始种群进行多代繁殖得到目标种群，在繁殖过程中将多个历史事件中的已知风险进化为未知风险，并在目标种群中确定出具有未知风险的至少一个目标事件用于当前风控策略的更新过程。更新后的当前风控策略具有识别未知风险的能力，可以改善风控经验滞后于风险事件，对于未知风险存在漏放的问题。

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，用来进行数据处理。当所述至少一个指令集被处理器执行时，所述至少一个指令集指导所述处理器实施本说明书所述的数据处理方法的步骤。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。当所述程序产品在计算系统 200 上运行时，所述程序代码用于使计算系统 200 执行本说明书描述的数据处理方法的步骤。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，并可以在计算系统 200 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。所述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。所述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，所述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。程序代码可以完全地在计算系统 200 上执行、部分地在计算系统 200 上执行、作为一个独立的软件包执行、部分在计算系统 200 上部分在远程计算设备上执行、或者完全在远程计算设备上执行。

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。

综上所述，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. 一种数据处理方法，包括：

获得多个历史事件，所述多个历史事件中的至少部分事件为具有已知风险的事件，所述已知风险为当前风控策略能够识别的风险；

将所述多个历史事件作为初始种群，通过模拟种群进化过程对所述初始种群进行 N 代繁殖得到目标种群，其中，在所述 N 代繁殖的过程中所述已知风险被进化为所述当前风控策略不能识别的未知风险，所述 N 为大于或等于 1 的整数；以及

在所述目标种群中确定出具有所述未知风险的至少一个目标事件，所述至少一个目标事件用于所述当前风控策略的更新过程。

2. 根据权利要求 1 所述的方法，其中，第i代的繁殖过程包括：

对当前种群中的不同事件进行交叉操作，以生成多个进化事件；

从所述多个进化事件中淘汰掉部分事件，所述部分事件不满足预设的风险进化条件；以及

将所述多个进化事件中剩余的事件作为第i代繁殖后的种群，其中

当$i = 1$时，所述当前种群为所述初始种群，当$1 < i \leq N$时，所述当前种群为第$i - 1$代繁殖后的种群，所述目标种群为第 N 代繁殖后的种群。

3. 根据权利要求 2 所述的方法，其中，一个事件对应有多个属性，所述进化事件的部分属性的属性值来自第一事件，所述进化事件其他属性的属性值来自第二事件，所述第一事件和所述第二事件为所述当前种群中的不同事件。

4. 根据权利要求 3 所述的方法，其中，所述多个属性至少包括风险属性，所述风险属性的属性值表征是否具有已知风险；

所述进化事件是通过将所述第一事件中的部分属性的属性值替换为所述第二事件的对应属性的属性值得到的，其中，所述第一事件中的所述风险属性的属性值表征具有已知风险，所述部分属性包括所述多个属性中除所述风险属性之外的至少一个属性。

5. 根据权利要求 2 所述的方法，其中，所述风险进化条件包括：所述当前风控策略将所述进化事件识别为无风险事件。

6. 根据权利要求 2 所述的方法，其中，所述多个进化事件中被淘汰的事件数量满足如下要求：淘汰后剩余的事件数量与所述初始种群中的事件数量之差的绝对值小于预设阈值。

7. 根据权利要求 2 所述的方法，其中，所述从所述多个进化事件中淘汰掉部分事件，包括：

对所述多个进化事件中的至少部分进化事件进行变异操作，得到包括多个变异事件的进化事件集合；以及

从所述进化事件集合中淘汰掉所述部分事件。

8. 根据权利要求 7 所述的方法，其中，一个事件对应有多个属性，所述多个属性至少包括风险属性；

一个进化事件对应的所述变异操作包括下述中的至少一种：

随机增加所述进化事件的至少一个属性的属性值；

随机删除所述进化事件除所述风险属性以外的至少一个属性的属性值；

随机更改所述进化事件除所述风险属性以外的至少一个属性的属性值。

9. 根据权利要求 1 所述的方法，其中，每个历史事件对应有风险属性，所述风险属性的属性值表征所述历史事件是否具有所述已知风险；

在所述 N 代繁殖的过程中，所述多个历史事件各自对应的所述风险属性的属性值被逐代遗传至所述目标种群中的事件；

所述在所述目标种群中确定出具有所述未知风险的至少一个目标事件，包括：针对所述目标种群中的每个事件，若所述事件对应的所述风险属性的属性值表征具有所述已知风险，则将所述事件确定为所述目标事件。

10. 根据权利要求 1 所述的方法，其中，所述当前风控策略包括风控模型，所述方法还包括：

至少利用所述至少一个目标事件对所述风控模型进行训练，得到训练后的风控模型，所述训练后的风控模型具有识别所述未知风险的能力。

11. 一种数据处理系统，包括：

至少一个存储介质，存储有至少一个指令集，用于对数据进行处理；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，所述数据处理系统运行时，所述至少一个处理器读取所述至少一个指令集，并且根据所述至少一个指令集的指示：

获得多个历史事件，所述多个历史事件中的至少部分事件为具有已知风险的事件，所述已知风险为当前风控策略能够识别的风险；

将所述多个历史事件作为初始种群，通过模拟种群进化过程对所述初始种群进行 N 代繁殖得到目标种群，其中，在所述 N 代繁殖的过程中所述已知风险被进化为所述当前风控策略不能识别的未知风险，所述 N 为大于或等于 1 的整数；以及

在所述目标种群中确定出具有所述未知风险的至少一个目标事件，所述至少一个目标事件用于所述当前风控策略的更新过程。

12. 根据权利要求 11 所述的系统，其中，为了繁殖第代，所述至少一个处理器：

对当前种群中的不同事件进行交叉操作，以生成多个进化事件；

从所述多个进化事件中淘汰掉部分事件，所述部分事件不满足预设的风险进化条件；以及

将所述多个进化事件中剩余的事件作为第i代繁殖后的种群，其中

当$i = 1$时，所述当前种群为所述初始种群，当$1 < i \leq N$时，所述当前种群为第$i-1$代繁殖后的种群，所述目标种群为第 N 代繁殖后的种群。

13. 根据权利要求 12 所述的系统，其中，一个事件对应有多个属性，所述进化事件的部分属性的属性值来自第一事件，所述进化事件其他属性的属性值来自第二事件，所述第一事件和所述第二事件为所述当前种群中的不同事件。

14. 根据权利要求 13 所述的系统，其中，所述多个属性至少包括风险属性，所述风险属性的属性值表征是否具有已知风险；

所述进化事件是通过将所述第一事件中的部分属性的属性值替换为所述第二事件的对应属性的属性值得到的，其中，所述第一事件中的所述风险属性的属性值表征具有已知风险，所述部分属性包括所述多个属性中除所述风险属性之外的至少一个属性。

15. 根据权利要求 12 所述的系统，其中，所述风险进化条件包括：所述当前风控策略将所述进化事件识别为无风险事件。

16. 根据权利要求 12 所述的系统，其中，所述多个进化事件中被淘汰的事件数量满足如下要求：淘汰后剩余的事件数量与所述初始种群中的事件数量之差的绝对值小于预设阈值。

17. 根据权利要求 12 所述的系统，其中，为了从所述多个进化事件中淘汰掉部分事件，所述至少一个处理器：

对所述多个进化事件中的至少部分进化事件进行变异操作，得到包括多个变异事件的进化事件集合；以及

从所述进化事件集合中淘汰掉所述部分事件。

18. 根据权利要求 17 所述的系统，其中，一个事件对应有多个属性，所述属性至少包括风险属性；

一个进化事件对应的所述变异操作包括下述中的至少一种：

随机增加所述进化事件的至少一个属性的属性值；

随机删除所述进化事件除所述风险属性以外的至少一个属性的属性值；

随机更改所述进化事件除所述风险属性以外的至少一个属性的属性值。

19. 根据权利要求 11 所述的系统，其中，每个历史事件对应有风险属性，所述风险属性的属性值表征所述历史事件是否具有所述已知风险；

在所述 N 代繁殖的过程中，所述多个历史事件各自对应的所述风险属性的属性值被逐代遗传至所述目标种群中的事件；

为了在所述目标种群中确定出具有所述未知风险的至少一个目标事件，所述至少一个处理器：

针对所述目标种群中的每个事件，若所述事件对应的所述风险属性的属性值表征具有所述已知风险，则将所述事件确定为所述目标事件。

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 1-10 中任一项所述的方法。