

说明书

群体特征的挖掘方法、事件识别方法、设备及存储介质

5 技术领域

本说明书涉及互联网技术领域，尤其涉及一种群体特征的挖掘方法、事件识别方法、设备及存储介质。

背景技术

- 10 随着互联网技术的发展，通过网络进行的群体犯罪和网络攻击等问题日益严重，防控平台需要防控的群体数量也日益增加。在实际防控的过程中，防控平台可以依靠防控人员的经验对大量事件的相关信息进行分析，以识别出目标群体的特征。后续防控平台可以依据防控人员识别的目标群体的特征，对目标群体执行的事件进行防控。但是，事件的数据量庞大且复杂度高，防控人员进行人工分析识别的效率低，使得防
- 15 控平台难以做到及时的防控。

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

- 20 本说明书提供一种群体特征的挖掘方法、事件识别方法、设备及存储介质，能够同时针对大量的事件数据挖掘目标群体的特征，提高获取目标群体特征的效率，以便及时防控这些目标群体。

- 第一方面，本说明书提供一种群体特征的挖掘方法，包括：获得多个正常事件的事件信息、以及多个异常事件的事件信息，每个事件的事件信息包括描述所述事件的多个属性值，所述多个异常事件的操作者包括目标群体；基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，所述 N 为大于或等于 1 的整数，其中，每个属性值组合中包括至少一个属性值，每个属性值组合在所述多个异常事件中的出现频次高于第一预设阈值，且在所述多个正常事件中的出现频次为零；以及，基于所述 N 个属性值组合，生成所述目标群体对应的群体特征。
- 25

第二方面，本说明书还提供一种事件识别方法，包括：响应于接收到事件识别请求，获得待识别的目标事件的事件信息，所述事件信息包括描述所述目标事件的多个属性值；获得目标群体对应的群体特征，其中，所述群体特征是采用如权利要求 1 至 8 中任一项所述的方法挖掘得到的，且用于描述所述目标群体所执行的事件满足的属性特征；以及，将所述目标事件的事件信息与所述群体特征进行匹配，以确定所述目标事件是否是由所述目标群体执行的事件。

第三方面，本说明书提供一种群体特征的挖掘设备，包括至少一个存储介质和至少一个处理器，所述至少一个存储介质存储有至少一个指令集，用于进行群体特征的挖掘；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，并且根据所述至少一个指令集的指示：获得多个正常事件的事件信息、以及多个异常事件的事件信息，每个事件的事件信息包括描述所述事件的多个属性值，所述多个异常事件的操作者包括目标群体；基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，所述 N 为大于或等于 1 的整数，其中，每个属性值组合中包括至少一个属性值，每个属性值组合在所述多个异常事件中的出现频次高于第一预设阈值，且在所述多个正常事件中的出现频次为零；以及，基于所述 N 个属性值组合，生成所述目标群体对应的群体特征。

第四方面，本说明书提供一种事件识别设备，包括至少一个存储介质和至少一个处理器，所述至少一个存储介质存储有至少一个指令集，用于进行事件识别；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，并且根据所述至少一个指令集的指示：响应于接收到事件识别请求，获得待识别的目标事件的事件信息，所述事件信息包括描述所述目标事件的多个属性值；获得目标群体对应的群体特征，其中，所述群体特征是采用如第一方面所述的方法挖掘得到的，且用于描述所述目标群体所执行的事件满足的属性特征；以及，将所述目标事件的事件信息与所述群体特征进行匹配，以确定所述目标事件是否是由所述目标群体执行的事件。

第五方面，本说明书提供一种计算机可读非暂时性存储介质，其中，所述计算机可读非暂时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，实现如第一方面所述的群体特征的挖掘方法，或者实现如第二方面所述的事件识别方法。

由以上技术方案可知，本说明书提供的群体特征的挖掘方法、设备及存储介质中，可以根据获得的正常事件和异常事件的事件信息确定属性值组合，后续根据该属性值组合生成表征目标群体的群体特征，该属性值组合可以高度概括异常事件且不概括正常事件，即属性值组合在异常事件的事件信息中频繁出现，且在正常事件的事件信息中不出现。可见，该方案在生成目标群体的群体特征的过程中考虑了正常事件的事件信息，尽量避免确定的群体特征在正常事件的事件信息中出现过，以便准确的防控目标群体，且不影响正常用户或群体。并且，本方案在挖掘目标群体的群体特征时，可以同时处理多个正常事件以及多个异常事件的事件信息，多个异常事件的事件信息可以由多个目标群体执行生成的，因此采用本方案可以提高群体特征的挖掘效率，使得对目标群体的防控更加及时。

进一步的，本说明书提供的事件识别方法、设备及存储介质中，能够利用所述挖掘方法中生成的群体特征识别目标事件，在识别到目标事件是由目标群体执行的事件后，可以通过手段阻碍目标事件的执行，从而达到防控目标群体的目的。采用所述事件识别的方案可以提高防控的及时性和准确性。

本说明书提供的群体特征的挖掘方法、事件识别方法、设备及存储介质的其他功能将在以下说明中部分列出。本说明书提供的群体特征的挖掘方法、事件识别方法、设备及存储介质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、设备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群体特征的挖掘以及事件识别的应用场景示意图；

图2示出了根据本说明书的实施例提供的一种电子设备的硬件结构图；

图3示出了根据本说明书的实施例提供的一种群体特征的挖掘方法的流程示意图；

图4示出了采用方式1确定N个属性值组合的流程示意图；

图 5 示出了方式 1 建立的项头表的示意图；

图 6 示出了方式 1 建立的属性值树的示意图；

图 7 示出了采用方式 2 确定 N 个属性值组合的流程示意图；

图 8 示出了方式 2 建立的项头表的示意图；以及

5 图 9 示出了根据本说明书的实施例提供的一种事件识别方法的流程示意图。

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，本说明书将对以下描述将会出现的术语进行如下解释。

目标群体：在需要防控的领域，目标群体可以理解为防控的组织或者团伙。在其他领域，目标群体可以理解为该领域需要针对的组织或者团伙。一个目标群体中可以

包括一个或者多个对象。对象可以通过用户账号、用户名、设备标识（ID，identifier）、卡号等来表征，或者，对象可以通过上述中两项或者两项以上的组合来表征。

群体特征：群体执行的事件所满足的属性特征。

属性值：包括属性和值，即属性值中包含了属性以及值两个维度的信息。例如属性值为卡号 1111，其中卡号为属性，1111 为值。

正常事件：除上述目标群体之外的对象或者群体执行的事件，为后文描述方便，将除上述目标群体之外的对象或者群体称为正常群体。一般来说，描述正常事件的属性值都为正常值，没有异常情况，或者其中有部分属性值有些微异常，但不影响事件的定性。

异常事件：上述目标群体执行的事件，或者，由正常群体执行的部分事件，描述该部分事件的属性值有异常，或者描述该部分事件的属性值的组合有异常。

如背景技术中所述，在防控目标群体的过程中，防控人员可以人为地对事件信息进行分析，从而挖掘出目标群体对应的群体特征。上述分析以及挖掘过程非常依赖防控人员积累的专业经验，依防控人员的专业程度，分析和挖掘的准确性有所不同。并且一次分析挖掘过程只能针对一个目标群体，效率比较低，难以做到及时有效的防控。

鉴于此，本说明书提供一种群体特征的挖掘方法，能够同时处理大量事件信息，提高群体特征的挖掘效率。并且，上述挖掘方法能够将正常事件和异常事件作区分，以正常事件为参照，得到高度表征由目标群体执行的异常事件的群体特征，提高群体特征挖掘的准确性。

进一步的，本说明书还提供了一种事件识别方法，能够利用所述挖掘方法挖掘的群体特征对目标事件进行识别，即，识别目标事件是否是由目标群体执行的事件。如果目标事件是由目标群体执行的事件，则避免该目标事件执行完成，从而达到防控目标群体的目的。所述事件识别方法能够提高防控目标群体的及时性和准确性。

下面结合图 1 对本说明书提供的群体特征的挖掘以及事件识别方案的应用场景进行介绍。

图 1 示出了根据本说明书的实施例提供的一种群体特征的挖掘以及事件识别的应用场景示意图。如图 1 所示，场景 001 可以包括：目标群体 100，终端设备 200，服务器 300，网络 400，挖掘设备 500，正常群体 600，终端设备 700。

其中，目标群体 100 可以通过终端设备 200 执行异常事件，正常群体可以通过终端设备 700 执行正常事件。本领域技术人员能够理解的，异常事件和正常事件在不同的业务场景下对应不同的指代。例如，在支付场景下，正常事件可以是指合法用户使用自己的卡/账号进行的支付事件，异常事件可以是指不法分子盗用合法用户的卡/账号进行的支付事件。

终端设备 200 或 700 可以向用户呈现交互界面，以使用户通过该交互界面执行一些事件，例如，支付事件、转账事件等。终端设备 200 或 700 可以包括移动设备、平板电脑、笔记本电脑、机动车辆的内置设备或类似内容，或其任意组合。在一些实施例中，所述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，所述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，所述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，所述虚拟现实设备或增强现实设备可能包括虚拟现实头盔、虚拟现实眼镜、虚拟现实补丁、增强现实头盔、增强现实眼镜、增强现实补丁或类似内容，或其中的任何组合。例如，所述虚拟现实设备或所述增强现实设备可能包括谷歌眼镜、头戴式显示器、VR 等。在一些实施例中，所述机动车中的内置装置可包括车载计算机、车载电视等。在一些实施例中，终端设备 200 或 700 可以是具有定位技术的设备，用于定位终端设备 200 或 700 的位置。

在一些实施例中，终端设备 200 或 700 可以安装有一个或多个应用程序（APP）。所述 APP 能够为用户提供执行某些事件（例如支付事件、转账事件等）的能力。所述 APP 包括但不限于：网页浏览器类 APP 程序、搜索类 APP 程序、聊天类 APP 程序、购物类 APP 程序、视频类 APP 程序、理财类 APP 程序等等。

如图 1 所示，终端设备 200 和终端设备 700 可以与服务器 300 进行通信连接。在一些实施例中，服务器 300 还可以与其他终端设备进行通信连接，并接收终端设备发送的数据。在一些实施例中，终端设备 200 和终端设备 700 可以通过网络 400 与服务器 300 交互，以接收或发送消息等。服务器 300 可以是提供各种服务的服务器，例如对终端设备 200 安装的境外购物平台提供支持的后台服务器。此时，服务器 300 可以存储有目标群体 100 执行异常交易事件的事件信息以及正常群体 600 执行正常交易事件的事件信息。

网络 400 用以在终端设备和服务器 300 之间提供通信连接的介质，或者用以在挖掘设备 500 和服务器 300 之间提供通信连接的介质。网络 400 可以促进信息或数据的交换。在一些实施例中，网络 400 可以是任何类型的有线或无线网络，也可以是其组合。比如，网络 400 可以包括电缆网络，有线网络、光纤网络、电信通信网络、内联网、互联网、局域网（LAN）、广域网（WAN）、无线局域网（WLAN）、大都市市区网（MAN）、公用电话交换网（PSTN）、蓝牙网络、ZigBee 网络、近场通信（NFC）网络或类似网络。在一些实施例中，网络 400 可以包括一个或多个网络接入点。例如，网络 400 可以包括有线或无线网络接入点，如基站或互联网交换点，通过该接入点，终端设备 200、服务器 300 和挖掘设备 500 的一个或多个组件可以连接到网络 400 以交换数据或信息。

挖掘设备 500 是具有一定数据处理能力的电子设备。挖掘设备 500 可以执行本说明书描述的群体特征的挖掘方法的数据和指令，并可以执行或用于执行所述数据和指令。在一些实施例中，挖掘设备 500 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。挖掘设备 500 可以和服务器 300 进行通信，从服务器 300 获得正常事件的事件信息以及异常事件的事件信息。例如，挖掘设备 500 可以通过网络 400 向服务器 300 获取事件信息，或者，技术人员可以人为的将服务器 300 中的事件信息导入挖掘设备 500。

在一些实施例中，挖掘设备 500 通过本说明书提供的群体特征挖掘方法得到群体特征后，将所述群体特征通过网络 400 传到服务器 300 中，服务器 300 可以基于所述群体特征实现本说明书提供的事件识别方法。服务器 300 可以执行本说明书描述的事件识别方法的数据和指令，并可以执行或用于执行所述数据和指令。在另一些实施例中，挖掘设备 500 通过本说明书提供的群体特征挖掘方法得到群体特征后，将所述群体特征通过网络 400 传到终端设备 200 或 700 中，或者，将所述群体特征通过网络 400 先传到服务器 300 中，再由服务器 300 分发到终端设备 200 或 700 中。终端设备 200 或 700 可以基于所述群体特征实现本说明书提供的事件识别方法。终端设备 200 或 700 可以执行本说明书描述的事件识别方法的数据和指令，并可以执行或用于执行所述数据和指令。

应该理解，图 1 中的终端设备、服务器 300、网络 400 和挖掘设备 500 的数目仅仅是示意性的。根据实现需要，可以具有任意数目的终端设备、服务器 300、网络 400 和挖掘设备 500。

需要说明的是，所述群体特征的挖掘方法可以完全在挖掘设备 500 上执行，也可以

完全在服务器 300 上执行，还可以部分在挖掘设备 500 上执行，部分在服务器 300 上执行。后面的描述中将以群体特征挖掘方法在挖掘设备 500 上执行为例进行说明。所述事件识别方法可以完全在终端设备 200 上执行，也可以完全在服务器 300 上执行，还可以部分在终端设备 200 上执行，部分在服务器 300 上执行。后面的描述中将以事件识别方法在服务器 300 上执行为例进行说明。

图 2 示出了根据本说明书的实施例提供的一种电子设备 800 的硬件结构图。电子设备 800 可以作为图 1 中的挖掘设备 500，也可以作为图 1 中的服务器 300，还可以作为图 1 中的终端设备 200 或 700。

如图 2 所示，电子设备 800 可以包括至少一个存储介质 830 和至少一个处理器 820。在一些实施例中，电子设备 800 还可以包括通信端口 850 和内部通信总线 810。同时，电子设备 800 还可以包括 I/O 组件 860。

内部通信总线 810 可以连接不同的系统组件，包括存储介质 830、处理器 820 和通信端口 850。

I/O 组件 860 支持电子设备 800 和其他组件之间的输入/输出。

通信端口 850 用于电子设备 800 同外界的数据通信，比如，通信端口 850 可以用于电子设备 800 同网络 400 之间的数据通信。通信端口 850 可以是有线通信端口也可以是无线通信端口。

存储介质 830 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 832、只读存储介质（ROM）834 或随机存取存储介质（RAM）836 中的一种或多种。存储介质 830 还包括存储在所述数据存储装置中的至少一个指令集。所述指令集可以是计算机程序代码。当电子设备 800 为挖掘设备 500 时，所述计算机程序代码可以包括执行本说明书提供的群体特征的挖掘方法的程序、例程、对象、组件、数据结构、过程、模块等等。当电子设备 800 作为服务器 300 或者终端设备 200 或 700 时，所述计算机程序代码可以包括执行本说明书提供的事件识别方法的程序、例程、对象、组件、数据结构、过程、模块等等。

处理器 820 可以同存储介质 830 以及通信端口 850 通过内部通信总线 810 通信连接。处理器 820 用以执行上述至少一个指令集。当电子设备 800 作为挖掘设备 500 运行时，处理器 820 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

本说明书提供的群体特征的挖掘方法。当电子设备 800 作为服务器 300 或者终端设备 200 或 700 运行时，处理器 820 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事件识别方法。

5 处理器 820 可以是一个或多个处理器的形式，在一些实施例中，处理器 820 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。

10 仅仅为了说明问题，在附图中电子设备 800 中仅示出一个处理器 820。然而，应当注意，本说明书中电子设备 800 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所述的由一个处理器执行，也可以由多个处理器联合执行。例如，如果在本说明书中电子设备 800 的处理器 820 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 820 联合或分开执行（例如，第一处
15 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群体特征的挖掘方法 P100 的流程图。如前，挖掘设备 500 可以执行本说明书的挖掘方法 P100。具体地，挖掘设备 500 中的处理器 820 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
20 本说明书的挖掘方法 P100。方法 P100 可应用于交易领域的盗卡团伙的特征挖掘、网络安全领域的网络攻击团伙的特征挖掘、金融领域的欺诈团伙的特征挖掘等。为清楚、完整的描述方法 P100，后文以购物平台的盗卡团伙（目标群体）的特征挖掘为例描述，方法 P100 在其他领域的应用可以此类推。

如图 3 所示，群体特征的挖掘方法 P100 可以包括：

25 S110：获得多个正常事件的事件信息、以及多个异常事件的事件信息，每个正常事件的事件信息包括描述所述正常事件的多个属性值，每个异常事件的事件信息包括描述所述异常事件的多个属性值，所述多个异常事件的操作者包括目标群体。

在购物平台的群体特征挖掘场景中，事件可以理解为交易，一个事件即一笔交易。挖掘设备 500 可以从购物平台的服务器中获得已完成交易的交易数据，例如正常交易的

交易信息和异常交易的交易信息。通常，在交易完成后，购物平台会通过标识信息标识该交易是正常交易还是异常交易。

一个事件的事件信息可以包括描述该事件的多个属性值。具体而言，事件可以存在多个属性、以及每个属性对应的值。在本说明书中，每个属性及其对应的值被称为一个属性值。举例而言，假设事件 x 为支付事件，事件 x 的事件信息可以表示为如下形式：

{[属性 1，值 1]，[属性 2，值 2]，[属性 3，值 3]，.....}

或者，事件 x 的事件信息可以表示为如下表 1 所示的形式。

表 1

事件标识	属性 1	属性 2	属性 3	
事件 x	值 1	值 2	值 3	

每笔交易的交易信息中描述交易的属性值包括基础属性值和统计属性值。其中，基础属性值为交易完成后在交易明细中展示的属性值，例如购物账号、付款卡号、购物金额、购物时间、收货地址、购物设备 ID、以及买家和卖家的交互等等。统计属性值为一笔交易完成后对该笔交易近期一段时间内的交易统计得到的属性值，或者对该笔交易中的基础属性值做处理后得到的属性值。例如该笔交易的交易时间为 1 月 1 日，统计属性值可以是该购物账号在 1 月 1 日完成的所有交易次数、交易金额等等。

另外，获得的多个异常交易不一定都是由目标群体执行的，也可能有部分异常交易是正常群体操作有误生成的。这里分两种情况：第一种情况，获得的多个异常交易都是由目标群体执行的；第二种情况，获得的多个异常交易中的部分异常交易是由目标群体执行的，部分异常交易是由正常群体生成的。

第一种情况，挖掘设备 500 初始获得异常交易的交易信息后，可以获取每个异常交易的异常标签，异常标签可以表征异常交易的异常程度。因此挖掘设备 500 可通过每个异常交易的异常标签从初始获得的异常交易中筛选出所述多个异常交易。可以理解，在交易完成后购物平台会给每个交易设置一个异常标签，或者购物平台对异常程度达到一定程度的交易设置异常标签。所述异常标签可以理解为上述区分交易是正常交易还是异常交易的标识信息，该标识信息对应的值可以量化。例如，交易的标识信息对应的值大于或等于阈值#1，则说明该交易为异常交易。交易的标识信息对应的值大于或等于阈值#2，则说明该交易为异常交易，且该交易是由目标群体执行的交易。或者，所述异常标签和所述标识信息是两个独立的信息，作为异常交易，既包括标识该交易

为异常交易的标识信息，也包括标识该交易是否是由目标群体执行的交易的异常标签信息。

第二种情况，挖掘设备 500 初始获得异常交易的交易信息后，不区分异常交易到底是不是由目标群体执行的，将初始获得的异常交易都确定为所述多个异常交易。例如

5 挖掘设备 500 初始获得异常交易集合 A，则挖掘设备 500 在 S110 中获得的所述多个异常交易即异常交易集合 A 中的所有异常交易。

针对上述两种不同的情况，后续步骤的实现会有所不同，具体区别在后文描述。

S120：基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，所述 N 为大于或等于 1 的整数，其中，每个属性值组合中包括
10 至少一个属性值，每个属性值组合在所述多个异常事件中的出现频次高于第一预设阈值，且在所述多个正常事件中的出现频次为零。

挖掘设备 500 根据多个正常交易的交易信息和多个异常交易的交易信息确定 N 个属性值组合可以有两种方式。

方式 1

15 挖掘设备 500 基于所述多个异常交易的交易信息确定 M 个候选属性值组合，其中，每一个候选属性值组合在所述多个异常交易中的出现频次高于预设的阈值#3，每一个候选属性值组合包括一个或多个属性值。

其中，M 个候选属性值组合包括候选属性值组合#1，候选属性值组合#1 包括属性值 a、属性值 b 和属性值 c。若一个异常交易的交易信息中包括属性值 a、属性值 b 和属性值 c 这个整体，则认为候选属性值组合#1 在所述多个异常交易中的出现了 1 次。若属性值 a、属性值 b 和属性值 c 作为一个整体在所述多个异常交易中出现了 m 次，则认为
20 候选属性值组合#1 在所述多个异常交易中的出现了 m 次，m 为大于 1 的整数。

所述出现频次可以是次数，此时预设的阈值#3 也为次数；或者，所述出现频次为出现的次数占异常交易数量的百分比，此时预设的阈值#3 也为百分比。

25 应当理解，属性值 a、属性值 b 和属性值 c 作为一个整体在一个异常交易的交易信息中出现，并不意味属性值 a、属性值 b 和属性值 c 是相邻的属性值。凡是一个异常交易的交易信息中包括属性值 a、属性值 b 和属性值 c，则认为该异常交易出现了属性值 a、属性值 b 和属性值 c 这个整体。

在一些实施例，如图 4 所示，挖掘设备 500 可以对多个异常交易（图 4 示例的 n1
30 个异常交易，n1 大于 1）的交易信息中的属性值进行去重，得到多个目标属性值。然

后，针对所述多个目标属性值中的每一目标属性值，挖掘设备 500 可以基于所述目标属性值在所述多个异常交易中的出现频次，确定所述目标属性值的支持度。进而，挖掘设备 500 基于所述多个异常交易的交易信息、以及所述多个目标属性值的支持度，确定所述 M 个候选属性值组合。后续基于多个正常交易（图 4 示例的 n2 个正常交易，n2 大于 1）的交易信息从 M 个候选属性值组合中排除部分候选属性值组合得到所述 N 个属性值组合。

例如，多个异常交易为 10 个异常交易。其中，异常交易 1 的交易信息中包括属性值 a、b、c、e、f、o，异常交易 2 的交易信息中包括属性值 a、c、g，异常交易 3 的交易信息中包括属性值 e、i，异常交易 4 的交易信息中包括属性值 a、c、d、e、g，异常交易 5 的交易信息中包括属性值 a、c、e、g、l，异常交易 6 的交易信息中包括属性值 e、j，异常交易 7 的交易信息中包括属性值 a、b、c、e、f、p，异常交易 8 的交易信息中包括属性值 a、c、d，异常交易 9 的交易信息中包括属性值 a、c、e、g、m，异常交易 10 的交易信息中包括属性值 a、c、e、g、n。对所述 10 个异常交易的交易信息中的属性值去重得到的目标属性值为 a、b、c、d、e、f、g、i、j、l、m、n、o、p。去重可以理解不同异常交易中相同的属性值作为一个目标属性值，比如多个异常交易中都有属性值 a，则目标属性值中只包括一个属性值 a，其他属性值 a 被去重。得到多个目标属性值后，对每个目标属性值都计算支持度，比如，支持度=目标属性值的出现频次/异常交易的数量，或者支持度=目标属性值的出现频次。对于属性值 a 而言，属性值 a 在异常交易中出现了 8 次，异常交易的数量为 10，因此属性值 a 的支持度=8/10=80%，或者属性值 a 的支持度为 8。其他目标属性值的支持度以此类推确定，不再赘述。后文以支持度=目标属性值的出现频次为例介绍方案。

一种确定 M 个候选属性值组合的方式，挖掘设备 500 按照目标属性值的支持度大小建立项头表，并按照目标属性值的支持度大小对每一个异常交易中的目标属性值进行排序。进一步的，根据项头表和排序后的异常交易建立属性值树，属性值树建立完成后，从项头表的底部依次向上挖掘。在挖掘的过程中，在所述属性值树中找到项头表中每一项目标属性值对应的属性值子树，根据每一项目标属性值对应的属性值子树挖掘得到所述 M 个候选属性值组合。

例如，确定 a、b、c、d、e、f、g、i、j、l、m、n、o、p 中每一个目标属性值的支持度，将支持度小于阈值#4 的目标属性值删除。比如目标属性值 i、j、l、m、n、o、p 的支持度都小于阈值#4，因此 i、j、l、m、n、o、p 不会出现在项头表中，并根据剩余

目标属性值 a、b、c、d、e、f、g 的支持度大小从高到低建立项头表，建立的项头表如图 5 所示。并按照图 5 所示，在每一个异常交易的交易信息中将支持度小于阈值#4 的目标属性值删除，针对每一个异常交易，将每一个异常交易中包括的目标属性值按照支持度大小排序，形成排序后的异常交易的交易信息。

5 然后，根据项头表和排序后的异常交易的交易信息建立属性值树，即依次在树中插入排序后的异常交易的交易信息。根据项头表和排序后的异常交易的交易信息建立的属性值树如图 6 所示，图 6 中的数字表示在树中插入异常交易的交易信息过程中某目标属性值重叠的总次数。图 6 中的箭头表示项头表中目标属性值到树中节点的链接。

 接着，根据属性值树从项头表的底部依次向上挖掘，一次挖掘针对一个目标属性值。针对每个目标属性值，找到对应目标属性值的条件模式基，即找到目标属性值作为叶子节点的子树。得到子树后，将子树中每个属性值节点的计数设置为叶子节点的计数，并删除计数低于支持度的节点。挖掘完项头表中的所有目标属性值即可得到 M 个候选属性值组合。比如，从项头表最底部的目标属性值 f 开始，先寻找 f 的条件模式基，由于 f 在树中只有一个节点，因此候选就只有一条路径，即{a:8,c:8,e:6,b:2, f:2}。

15 接着将路径上其他节点的计数设置为 f 节点的计数，即 f 节点对应的子树变为 {a:2,c:2,e:2,b:2, f:2}。根据 f 节点对应的子树，得到 2 项、3 项...更多项的目标属性值的组合。以此类推，挖掘项头表中的每个目标属性值后，得到多个目标属性值的组合，作为 M 个候选属性值组合。

 另一种确定 M 个候选属性值组合的方式，挖掘设备 500 确定所有目标属性值的支持度后，将支持度小于阈值#4 的目标属性值删除，即后续确定候选属性值组合时不考虑被删除的目标属性值。针对删除后剩下的目标属性值，依次确定其中任意 2 个目标属性值的组合的支持度，任意 3 个目标属性值的组合的支持度，任意 4 个目标属性值的组合的支持度，直至确定完最大数量的目标属性值的组合的支持度，最大数量即删除后剩余的目标属性值的个数。

25 例如，确定 2 个目标属性值的组合的支持度时，目标属性值的组合的支持度=目标属性值的组合在异常交易中的出现频次/异常交易的数量。对于目标属性值的组合 ace 而言，在异常交易中出现了 6 次，异常交易的数量为 10，因此目标属性值的组合 ace 的支持度=6/10=60%，或者目标属性值的组合 ace 的支持度=6。其他多个目标属性值的组合的支持度以此类推确定，不再赘述。

确定了不同目标属性值的组合的支持度后，对于任一目标属性值的组合而言，基于该目标属性值的组合的支持度和阈值#3 确定该目标属性值的组合是否是候选属性值组合。比如，若目标属性值的组合的支持度小于或等于阈值#3，则该目标属性值的组合不是候选属性值组合，若目标属性值的组合的支持度大于阈值#3，则该目标属性值的组合是候选属性值组合。

下面介绍基于所述多个正常交易的交易信息，从所述 M 个候选属性值组合中确定所述 N 个属性值组合的具体方式，M 为大于或等于 N 的整数。

在一些实施例中，挖掘设备 500 基于所述多个正常交易的交易信息，从所述 M 个候选属性值组合中排除第一属性值组合得到所述 N 个属性值组合，其中，所述第一属性值组合与所述多个正常交易中的一正常交易的交易信息存在相同的属性值，并且相同的属性值的数量大于或等于阈值#5。

例如，M 个候选属性值组合为候选属性值组合 aceg、ace、acg、aeg、ac。当阈值#5 为候选属性值组合中属性值的数量时，若某一候选属性值组合在某一正常交易的交易信息中出现，则认为该候选属性值组合为第一属性值组合。针对候选属性值组合 ace，查找每个正常交易的交易信息中是否出现属性值组合 ace，若有一个正常交易的交易信息中包括属性值组合 ace，则认为属性值组合 ace 为第一属性值组合。

再例如，当阈值#5 是预设的百分数与候选属性值组合中属性值的数量的乘积时，若所述预设的百分数为 60%，某一候选属性值组合中的属性值的 60%以上出现在某一正常交易的交易信息中，则认为该候选属性值组合为第一属性值组合。针对候选属性值组合 ace，查找每个正常交易的交易信息中是否出现属性值组合 ace 中的两个属性值或者 3 个属性值，若某一正常交易的交易信息中包括属性值组合 ace 中的两个属性值或者 3 个属性值，则认为该属性值组合 ace 为第一属性值组合。

按照上述示例的方式依次确定候选属性值组合是否为第一属性值组合，并将 M 个候选属性值组合中的第一属性值组合排除，从而得到 N 个属性值组合。这种方式可以避免确定的 N 个属性值组合在正常交易的交易信息中出现，从而后续根据该 N 个属性值组合识别目标群体进行的交易时，可以避免影响正常群体的交易，提高识别的准确度。

方式 2

挖掘设备 500 基于所述多个异常交易的交易信息和所述多个正常交易的交易信息确定 N 个候选属性值组合。

在一些实施例中，如图 7 所示，挖掘设备 500 可以按照方式 1 中的去重方式对所述多个异常交易的交易信息中的属性值进行去重，得到多个目标属性值。挖掘设备 500 得到多个目标属性值后，确定每一个目标属性值的支持度，此处与方式 1 中确定支持度的方式不同。在方式 2 中，挖掘设备 500 确定目标属性值的支持度时，除考虑目标属性值在异常交易中出现的频次之外，还需要考虑目标属性值在正常交易中出现的频次。后续基于目标属性值的支持度，多个异常交易（图 7 示例的 $n1$ 个异常交易， $n1$ 大于 1）的交易信息，以及多个正常交易（图 7 示例的 $n2$ 个正常交易， $n2$ 大于 1）的交易信息确定 N 个属性值组合。

例如，多个异常交易为上述 10 个异常交易，正常交易包括 2 个，分别为属性值 a、c、q、y、w 和属性值 a、c、h、u、v。对于目标属性值 a、b、c、d、e、f、g、i、j、l、m、n、o、p 中的每一个目标属性值，挖掘设备 500 获得所述目标属性值在所述多个异常交易中的第一出现频次，并获得所述目标属性值在所述多个正常交易中的第二出现频次，根据所述第一出现频次和所述第二出现频次确定所述目标属性值的支持度。

比如，对于每一个目标属性值，挖掘设备 500 可以确定第一出现频次与第二出现频次之间的差值，并基于所述差值确定该目标属性值的支持度。第一出现频次为所述目标属性值在异常交易中出现的次数，第二出现频次为所述目标属性值在正常交易中出现的次数。所述目标属性值的支持度 = （第一出现频次 - 第二出现频次）；或者所述目标属性值的支持度 = （第一出现频次 - 第二出现频次） / （正常交易的数量 + 异常交易的数量）；或者，所述目标属性值的支持度 = （第一出现频次 - 第二出现频次） / 异常交易的数量；或者，所述目标属性值的支持度 = （第一出现频次 / 异常交易的数量） - （第二出现频次 / 正常交易的数量）；或者所述目标属性值的支持度还有类似上述公式的确定方式，此处不一一列举。本说明书公式中出现的“/”表示除以。以目标属性值的支持度 = （第一出现频次 - 第二出现频次）为例说明，目标属性值 a 的第一出现频次为 8，第二出现频次为 2，因此目标属性值 a 的支持度 = $8 - 2 = 6$ ，其余目标属性值的支持度以此类推。

一种确定 N 个属性值组合的方式，挖掘设备 500 按照目标属性值的支持度大小建立项头表，并按照目标属性值的支持度大小对每一个异常交易中的目标属性值进行排序。进一步的，根据项头表和排序后的异常交易建立属性值树，属性值树建立完成后，从项头表的底部依次向上挖掘。在挖掘的过程中，在所述属性值树中找到项头表中每一项目标属性值对应的属性值子树，根据每一项目标属性值对应的属性值子树挖掘得到所述 M 个候选属性值组合。

例如，确定 a、b、c、d、e、f、g、i、j、l、m、n、o、p 中每一个目标属性值的支持度，将支持度小于阈值#4 的目标属性值删除。比如目标属性值 i、j、l、m、n、o、p 的支持度都小于阈值#4，因此 i、j、l、m、n、o、p 不会出现在项头表中，并根据剩余目标属性值 a、b、c、d、e、f、g 的支持度大小从高到低建立项头表，建立的项头表如图 8 所示。并按照图 8 所示，在每一个异常交易的交易信息中将支持度小于阈值#4 的目标属性值删除，针对每一个异常交易，将每一个异常交易中包括的目标属性值按照支持度大小排序，形成排序后的异常交易的交易信息。

后续根据项头表和排序后的异常交易的交易信息建立属性值树，以及根据属性值树从项头表的底部依次向上挖掘从而得到 N 个属性值组合的过程与方式 1 中挖掘得到 M 个候选属性值组合的过程类似，可参考上文中的描述，不再赘述。

另一种确定 N 个属性值组合的方式，挖掘设备 500 确定所有目标属性值的支持度后，将支持度小于阈值#4 的目标属性值删除，即后续确定候选属性值组合时不考虑被删除的目标属性值。针对删除后剩下的目标属性值，依次确定其中任意 2 个目标属性值的组合的支持度，任意 3 个目标属性值的组合的支持度，任意 4 个目标属性值的组合的支持度，直至确定完最大数量的目标属性值的组合的支持度，最大数量即删除后剩下的目标属性值的个数。

例如，目标属性值的组合的支持度=目标属性值的组合在异常交易中的出现频次；或者，目标属性值的组合的支持度=目标属性值的组合在异常交易中的出现频次/异常交易的数量；或者，目标属性值的组合的支持度=目标属性值的组合在异常交易中的出现频次/（异常交易的数量+正常交易的数量）。如果一个目标属性值的组合除了在异常交易中出现，还在正常交易中出现，则该目标属性值的组合的支持度=目标属性值的组合在异常交易中的出现频次-目标属性值的组合在正常交易中的出现频次；或者，该目标属性值的组合的支持度=（目标属性值的组合在异常交易中的出现频次-目标属性值的组合在正常交易中的出现频次）/异常交易的数量；或者，该目标属性值的组合的支持度=（目标属性值的组合在异常交易中的出现频次-目标属性值的组合在正常交易中的出现频次）/（异常交易的数量+正常交易的数量）。

确定了不同目标属性值的组合的支持度后，对于任一目标属性值的组合而言，基于该目标属性值的组合的支持度和阈值#6 确定该目标属性值的组合是否属于所述 N 个属性值组合之一。比如，若目标属性值的组合的支持度小于或等于阈值#6，则该目标属性值的组合不属于所述 N 个属性值组合之一，若目标属性值的组合的支持度大于阈

值#6, 则该目标属性值的组合属于所述 N 个属性值组合之一。阈值#6 可以和阈值#3 相同, 也可以不同。

5 综上, 方式 2 在确定目标属性值的支持度的过程中, 考虑了目标属性值在正常交易中出现的频次, 因此该方式确定的 N 个属性值组合在正常交易中出现的概率较低。后续根据所述 N 个属性值组合识别目标群体进行的交易时, 在一定程度上也能够提高识别的准确度。

S130: 基于所述 N 个属性值组合, 生成所述目标群体对应的群体特征。

基于 S110 中的两种情况, 生成所述目标群体对应的群体特征也分为以下两种情况。

情况 1

10 情况 1 针对 S110 中的第一种情况, 即挖掘设备 500 在 S110 中获取的多个异常交易都是由目标群体执行的。在该情况下, 可以直接根据所述 N 个属性值组合生成目标群体对应的群体特征, 即目标群体的群体特征包括所述 N 个属性值组合。

情况 2

15 情况 2 针对 S110 中的第二种情况, 即挖掘设备 500 在 S110 中获取的多个异常交易并非都是由目标群体执行的。在此情况下, 需要将所述多个异常交易中由目标群体执行的交易从所述多个异常交易中筛选出来。

20 在一些实施例中, 异常交易的交易信息中不包括异常标签, 因此需要获得所述多个异常交易中包含所述 N 个属性值组合中的任一属性值组合的异常交易的异常标签。若所述异常交易的交易信息中包括异常标签, 则可以直接使用异常标签, 不再重复获取。

25 比如, 针对所述 N 个属性值组合中的每一个属性值组合, 挖掘设备 500 查找包含所述属性值组合的异常交易, 并根据该异常交易对应的异常标签确定是否要将所述属性值组合从所述 N 个属性值组合中排除。在排除一些属性值组合后, 挖掘设备 500 可以根据 N 个属性值组合中排除剩下的目标属性值组合生成所述目标群体对应的群体特征。

30 具体而言, 针对所述 N 个属性值组合中的属性值组合 1: 在所述异常事件中确定至少一个目标异常事件, 所述目标异常事件的事件信息中包含属性值组合 1。进一步根据至少一个目标异常事件对应的异常标签, 确定属性值组合 1 对应的异常概率。例如, 包含属性值组合 1 的目标异常事件有 n1 个, 根据这 n1 个目标异常事件对应的异常标签能够确定 n1 个目标异常事件中由目标群体执行的目标异常事件的数量 n2, n1 大于 n2,

n_2 为大于或等于 1 的整数，属性值组合 1 对应的异常概率 $=n_2/n_1$ 。若属性值组合 1 对应的异常概率小于预设的阈值 θ ，则将属性值组合 1 从所述 N 个属性值组合中排除。

以此类推，挖掘设备 500 确定 N 个属性值组合中的每一个属性值组合对应的异常概率，将异常概率小于阈值 θ 的属性值组合从所述 N 个属性值组合中排除，剩余的属性值组合即为目标属性值组合。挖掘设备 500 可以直接根据目标属性值组合生成目标群体对应的群体特征，即目标群体的群体特征包括所述目标属性值组合。可以理解，上文中的目标群体可以是一个或多个群体。

综上所述，本说明书提供的群体特征的挖掘方法，可以根据获得的正常事件和异常事件的事件信息确定属性值组合，后续根据该属性值组合生成表征目标群体的群体特征，该属性值组合可以高度概括异常事件，即属性值组合在异常事件的事件信息中频繁出现，且在正常事件的事件信息中不出现或者少出现。可见，该方案在生成目标群体的群体特征的过程中考虑了正常事件的事件信息，尽量避免确定的群体特征在正常事件的事件信息中出现过，以便准确的防控目标群体，且不影响正常群体。并且，本方案在挖掘目标群体的群体特征时，可以同时处理多个正常事件以及多个异常事件的事件信息，多个异常事件的事件信息可以由多个目标群体执行生成的，因此采用本方案可以提高群体特征的挖掘效率，使得对目标群体的防控更加及时。另外，针对某一交易平台挖掘到的目标群体的群体特征可以上传到另一交易平台使用，具有可迁移性。

上文描述了挖掘设备 500 进行群体特征的挖掘过程。下面结合图 9 对服务器 300 或者终端设备 200 或 700 基于群体特征进行事件识别的方案进行详细说明。

图 9 示出了根据本说明书的实施例提供的一种事件识别方法 P200 的流程图。如前，服务器 300 或者终端设备 200 或 700 可以执行本说明书的事件识别方法 P200。具体地，服务器 300 或者终端设备 200 或 700 中的处理器 820 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，执行本说明书的事件识别方法 P200，后文以服务器 300 执行方法 P200 为例说明。方法 P200 可应用于交易领域的事件识别、网络安全领域的事件识别、金融领域的事件识别等。为清楚、完整的描述方法 P200，后文以识别购物平台的交易事件为例描述，方法 P200 在其他领域的应用可以此类推。

如图 9 所示，事件识别方法 P200 可以包括：

S210：响应于接收到事件识别请求，获得待识别的目标事件的事件信息，所述事

件信息包括描述所述目标事件的多个属性值。

目标事件可以是指网络中已发生或者即将发生的事件。例如，某一用户或群体在购物平台上正在执行的交易事件，该交易事件正在进行中，还未完成。比如，某用户或群体在购物平台上下了订单，还未付款，此时购物平台的客户端可以触发事件识别请求的发送。服务器 300 接收到所述事件识别请求后，获取所述订单的交易信息，交易信息中包括的用于描述订单的属性值可参考方法 P100 中属性值的举例，不再赘述。

S220：获得目标群体对应的群体特征，其中，所述群体特征是采用方法 P100 挖掘得到的，且用于描述所述目标群体所执行的事件满足的属性特征。

例如，挖掘设备 500 根据方法 P100 生成目标群体的群体特征后，将所述群体特征发送到服务器 300，服务器 300 可以直接从本地获取所述群体特征。

S230：将所述目标事件的事件信息与所述群体特征进行匹配，以确定所述目标事件是否是由所述目标群体执行的事件。

在一些实施例，如果所述目标事件的事件信息中包括所述群体特征，则确定所述目标事件是由所述目标群体执行的事件。比如，群体特征包括多个属性值组合，若目标事件的事件信息中包含其中一个属性值组合，则认为所述目标事件是由目标群体执行的事件。

确定所述目标事件是由所述目标群体执行的事件后，服务器 300 可以向购物平台的客户端发送验证要求，用来要求所述目标群体在购物平台的客户端采用刷脸或者输入信息等方式进行验证。如果验证不通过，比如目标群体输入的信息与该账号原始在购物平台注册的信息不一致，则阻止该交易的进行。或者，服务器 300 确定所述目标事件是由所述目标群体执行的事件后，直接阻止或拦截该订单的支付，使其支付不成功，交易不能顺利完成。

综上所述，本说明书提供的事件识别方法，能够利用所述挖掘方法中生成的群体特征识别目标事件，在识别到目标事件是由目标群体执行的事件后，可以通过手段阻碍目标事件的执行，从而达到防控目标群体的目的。所述事件识别的方案可以提高防控的及时性和准确性。

需要说明的是，上述描述方法 P100 和 P200 的过程中，以购物平台的应用场景举例，因此具体描述中采用的是该购物场景中会涉及的术语，例如交易，属性值的举例等。如果将方法 P100 和 P200 应用到前文提及的其他领域，所涉及的术语可以适应性的调整，此处不一一枚举。

本说明书另一方面提供一种非暂时性存储介质，存储有至少一组用来进行挖掘群体特征的可执行指令或者至少一组用来进行事件识别的可执行指令。当所述可执行指令被处理器执行时，所述可执行指令指导所述处理器实施本说明书所述的群体特征的挖掘方法 P100 或者事件识别方法 P200 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。当所述程序产品在挖掘设备 500 上或在服务器 300 或者终端设备 200 或 700 上运行时，所述程序代码用于使这些设备执行本说明书描述的群体特征的挖掘方法 P100 或事件识别方法 P200 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，并可以在挖掘设备 500 上或在服务器 300 或者终端设备 200 或 700 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。

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。

- 5 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。

权 利 要 求 书

1. 一种群体特征的挖掘方法，包括：

获得多个正常事件的事件信息、以及多个异常事件的事件信息，每个事件的事件信息包括描述所述事件的多个属性值，所述多个异常事件的操作者包括目标群体；

基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，所述 N 为大于或等于 1 的整数，其中，每个属性值组合中包括至少一个属性值，每个属性值组合在所述多个异常事件中的出现频次高于第一预设阈值，且在所述多个正常事件中的出现频次为零；以及

基于所述 N 个属性值组合，生成所述目标群体对应的群体特征。

2. 根据权利要求 1 所述的方法，其中，所述基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，包括：

基于所述多个异常事件的事件信息确定 M 个候选属性值组合，所述候选属性值组合在所述多个异常事件中的出现频次高于所述第一预设阈值，所述 M 为大于或等于 N 的整数；以及

基于所述多个正常事件的事件信息，从所述 M 个候选属性值组合中确定所述 N 个属性值组合。

3. 根据权利要求 2 所述的方法，其中，所述基于所述多个异常事件的事件信息确定 M 个候选属性值组合，包括：

对所述多个异常事件的事件信息中的属性值进行去重，得到多个目标属性值；

针对所述多个目标属性值中的每一目标属性值，基于所述目标属性值在所述多个异常事件中的出现频次，确定所述目标属性值的支持度；以及

基于所述多个异常事件的事件信息、以及所述多个目标属性值的支持度，确定所述 M 个候选属性值组合。

4. 根据权利要求 2 所述的方法，其中，所述基于所述多个正常事件的事件信息，从所述 M 个候选属性值组合中确定所述 N 个属性值组合，包括：

基于所述多个正常事件的事件信息，从所述 M 个候选属性值组合中排除第一属性值组合得到所述 N 个属性值组合，

其中，所述第一属性值组合与所述多个正常事件中的一正常事件的事件信息存在相同的属性值，并且相同的属性值的数量大于或等于第二预设阈值。

5.根据权利要求 1 所述的方法，其中，所述基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，包括：

对所述多个异常事件的事件信息中的属性值进行去重，得到多个目标属性值；

针对所述多个目标属性值中的每一目标属性值，基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定所述目标属性值的支持度；以及

基于所述多个异常事件的事件信息、以及所述多个目标属性值的支持度，确定所述 N 个属性值组合。

6. 根据权利要求 5 所述的方法，其中，所述基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定所述目标属性值的支持度，包括：

获得所述目标属性值在所述多个异常事件中的第一出现频次，并获得所述目标属性值在所述多个正常事件中的第二出现频次；以及

基于所述第一出现频次和所述第二出现频次，确定所述目标属性值的支持度。

7.根据权利要求 6 所述的方法，其中，所述基于所述第一出现频次和所述第二出现频次，确定所述目标属性值的支持度，包括：

确定所述第一出现频次与所述第二出现频次之间的差值；以及

基于所述差值，确定所述目标属性值的支持度。

8.根据权利要求 1 所述的方法，其中，所述基于所述 N 个属性值组合，生成所述目标群体对应的群体特征，包括：

获得所述多个异常事件中包含所述 N 个属性值组合的异常事件对应的异常标签，所述异常标签表征所述异常事件的异常程度；

基于所述异常事件对应的异常标签，从所述 N 个属性值组合中排除部分属性值组合得到至少一个目标属性值组合；以及

基于所述至少一个目标属性值组合，生成所述目标群体对应的群体特征。

9. 根据权利要求 8 所述的方法，其中，所述基于所述异常事件对应的异常标签，从所述 N 个属性值组合中排除部分属性值组合得到至少一个目标属性值组合，包括：

针对所述 N 个属性值组合中的每个属性值组合：

在所述异常事件中确定至少一个目标异常事件，所述目标异常事件的事件信息中包含所述属性值组合；

基于所述至少一个目标异常事件对应的异常标签，确定所述属性值组合对应的异常概率；以及

若所述异常概率小于第三预设阈值，则将所述属性值组合从所述 N 个属性值组合中排除。

10. 一种事件识别方法，包括：

响应于接收到事件识别请求，获得待识别的目标事件的事件信息，所述事件信息包括描述所述目标事件的多个属性值；

获得目标群体对应的群体特征，其中，所述群体特征是采用如权利要求 1 至 9 中任一项所述的方法挖掘得到的，且用于描述所述目标群体所执行的事件满足的属性特征；以及

将所述目标事件的事件信息与所述群体特征进行匹配，以确定所述目标事件是否是由所述目标群体执行的事件。

11. 根据权利要求 10 所述的方法，其中，所述将所述目标事件的事件信息与所述群体特征进行匹配，以确定所述目标事件是否是由所述目标群体执行的事件，包括：

如果所述目标事件的事件信息中包含所述群体特征，则确定所述目标事件是由所述目标群体执行的事件。

12. 一种群体特征的挖掘设备，包括：

至少一个存储介质，存储有至少一个指令集，用于进行群体特征的挖掘；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，并且根据所述至少一个指令集的指示：

获得多个正常事件的事件信息、以及多个异常事件的事件信息，每个事件的事件信息包括描述所述事件的多个属性值，所述多个异常事件的操作者包括目标群体，

基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定 N 个属性值组合，所述 N 为大于或等于 1 的整数，其中，每个属性值组合中包括至少一个属性值，每个属性值组合在所述多个异常事件中的出现频次高于第一预设阈值，且在所述多个正常事件中的出现频次为零，以及

基于所述 N 个属性值组合，生成所述目标群体对应的群体特征。

13. 根据权利要求 12 所述的设备，其中，为了确定所述 N 个属性值组合，所述至少一个处理器：

基于所述多个异常事件的事件信息确定 M 个候选属性值组合，所述候选属性值组合在所述多个异常事件中的出现频次高于所述第一预设阈值，所述 M 为大于或等于 N 的整数；以及

基于所述多个正常事件的事件信息，从所述 M 个候选属性值组合中确定所述 N 个属性值组合。

14. 根据权利要求 13 所述的设备，其中，为了确定所述 M 个候选属性值组合，所述至少一个处理器：

对所述多个异常事件的事件信息中的属性值进行去重，得到多个目标属性值；

针对所述多个目标属性值中的每一目标属性值，基于所述目标属性值在所述多个异常事件中的出现频次，确定所述目标属性值的支持度；以及

基于所述多个异常事件的事件信息、以及所述多个目标属性值的支持度，确定所述 M 个候选属性值组合。

15. 根据权利要求 13 所述的设备，其中，为了确定所述 N 个属性值组合，所述至少一个处理器：

基于所述多个正常事件的事件信息，从所述 M 个候选属性值组合中排除第一属性值组合得到所述 N 个属性值组合，

其中，所述第一属性值组合与所述多个正常事件中的一正常事件的事件信息存在相同的属性值，并且相同的属性值的数量大于或等于第二预设阈值。

16. 根据权利要求 12 所述的设备，其中，为了确定所述 N 个属性值组合，所述至少一个处理器：

对所述多个异常事件的事件信息中的属性值进行去重，得到多个目标属性值；

针对所述多个目标属性值中的每一目标属性值，基于所述多个正常事件的事件信息、以及所述多个异常事件的事件信息，确定所述目标属性值的支持度；以及

基于所述多个异常事件的事件信息、以及所述多个目标属性值的支持度，确定所述 N 个属性值组合。

17. 根据权利要求 16 所述的设备，其中，为了确定所述目标属性值的支持度，所述至少一个处理器：

获得所述目标属性值在所述多个异常事件中的第一出现频次，并获得所述目标属性值在所述多个正常事件中的第二出现频次；以及

基于所述第一出现频次和所述第二出现频次，确定所述目标属性值的支持度。

18. 根据权利要求 17 所述的设备，其中，为了确定所述目标属性值的支持度，所述至少一个处理器：

确定所述第一出现频次与所述第二出现频次之间的差值；以及

基于所述差值，确定所述目标属性值的支持度。

19. 一种事件识别设备，包括：

至少一个存储介质，存储有至少一个指令集，用于进行事件识别；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，并且根据所述至少一个指令集的指示：

响应于接收到事件识别请求，获得待识别的目标事件的事件信息，所述事件信息包括描述所述目标事件的多个属性值，

获得目标群体对应的群体特征，其中，所述群体特征是采用如权利要求 1 至 9 中任一项所述的方法挖掘得到的，且用于描述所述目标群体所执行的事件满足的属性特征，以及

将所述目标事件的事件信息与所述群体特征进行匹配，以确定所述目标事件是否是由所述目标群体执行的事件。

20. 一种计算机可读非暂时性存储介质，其中，所述计算机可读非暂时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 1-9 中任一项所述的方法，或者，实现如权利要求 10 或 11 所述的方法。