



(12) 发明专利

(10) 授权公告号 CN 112966728 B

(45) 授权公告日 2024. 08. 20

(21) 申请号 202110216921.1

G06F 18/214 (2023.01)

(22) 申请日 2021.02.26

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 111861486 A, 2020.10.30

申请公布号 CN 112966728 A

CN 110163714 A, 2019.08.23

(43) 申请公布日 2021.06.15

审查员 杨亚杰

(73) 专利权人 中国银联股份有限公司

地址 200135 上海市浦东新区含笑路36号

(72) 发明人 郭琦 闵勇 葛鸣铭 杨旭恒

韩昊 朱青源

(74) 专利代理机构 北京同达信恒知识产权代理

有限公司 11291

专利代理师 王春波

(51) Int. Cl.

G06F 18/241 (2023.01)

G06Q 20/38 (2012.01)

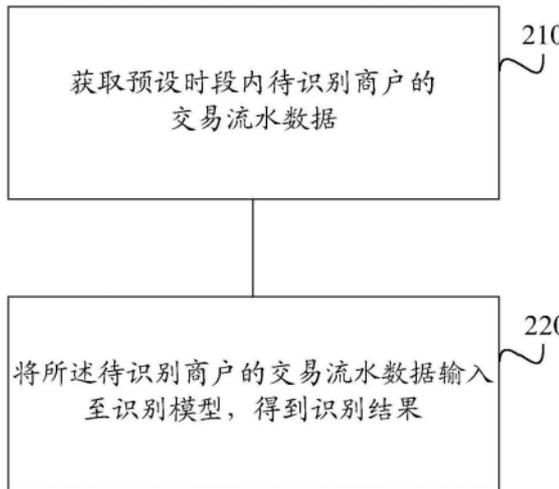
权利要求书2页 说明书11页 附图4页

(54) 发明名称

一种交易监测的方法及装置

(57) 摘要

本发明公开了一种交易监测的方法及装置,包括:获取预设时段内待识别商户的交易流水数据,然后将待识别商户的交易流水数据输入至识别模型,得到识别结果,其中,识别结果用于指示待识别商户是否为非法商户,识别模型是通过非法种子商户的交易流水数据和非法关联商户的交易流水数据进行训练得到的,以此增加了模型的泛化能力,提升了监测待识别商户的范围,非法关联商户是通过对非法种子商户进行关联扩散确定的,因此降低了所需样本的数据数量。



1. 一种交易监测的方法,其特征在于,包括:

获取预设时段内待识别商户的交易流水数据;

将所述待识别商户的交易流水数据输入至识别模型,得到识别结果;所述识别结果用于指示所述待识别商户是否为非法商户;

所述识别模型的训练过程包括:

分别对非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,得到多维度特征的各样本数据;其中,所述多维度特征包括预设时间内与商户存在交易的交易账户数、预设时间内同一交易账户交易金额最小值、预设时间内交易金额大于阈值的交易占比、预设时间内借记卡占比、预设时间内交易时间点占比、预设时间内发卡机构所在地高危地区占比、预设时间内交易金额为整数金额的占比;

根据非法种子商户的交易流水数据对应的样本数据、非法关联商户的交易流水数据对应的样本数据中的任一样本数据,根据所述样本数据的初始识别结果、所述样本数据的样本属性确定出第一结果差异值;通过预设超参数以及所述样本数据在关联扩散中的关联代值对所述第一结果差异值进行加权,得到第二结果差异值;根据各样本的第二结果差异值,确定损失函数值;所述初始识别结果是所述样本数据输入至初始识别模型得到的;根据所述损失函数值更新所述初始识别模型,直至得到所述识别模型;

所述非法关联商户是通过对所述非法种子商户进行多次关联扩散确定的;所述样本数据在关联扩散中的关联代值对应确定非法关联商户的次数。

2. 如权利要求1所述的方法,其特征在于,每个非法种子商户和每个非法关联商户均分别对应一个负样本属性;每个合法商户对一个正样本属性。

3. 如权利要求1所述的方法,其特征在于,根据下述公式(1)确定所述损失函数值;

$$L(y_i, f(x_i)) = \frac{1}{n} \sum_{i=1}^n \left[ (1-\gamma)^l (y_i - f(x_i))^2 \right] \dots \dots \dots (1);$$

其中, $L(y_i, f(x_i))$ 为所述损失函数值; $l$ 为第*i*个样本数据在关联扩散中的关联代值, $l$ 为自然数; $n$ 为样本数据的数量; $f(x_i)$ 为第*i*个样本数据的初始识别结果; $x_i$ 为第*i*个样本数据在所述初始识别模型中的输入值; $y_i$ 为第*i*个样本数据的样本属性; $\gamma$ 为预设超参数, $0 < \gamma < 1$ 。

4. 如权利要求1至3任一项所述的方法,其特征在于,所述非法关联商户是通过对所述非法种子商户进行关联扩散确定的,包括:

确定与所述非法种子商户存在交易的各关联账户;

根据账户关联商户的第一关联特征,从所述各关联账户中确定出疑似账户;所述账户关联商户的第一关联特征是指从账户的交易流水数据中确定出与商户关联的特征;

确定与所述疑似账户存在交易的关联商户;

根据商户关联账户的第二关联特征,从所述关联商户中确定出所述非法关联商户;所述商户关联账户的第二关联特征是指从商户的交易流水数据中确定出与账户关联的特征;

将所述非法关联商户更新为非法种子商户,返回确定与所述非法种子商户存在交易的各关联账户的步骤,直至满足设定条件。

5. 如权利要求4所述的方法,其特征在于,所述方法还包括:

将所述关联商户中除所述非法关联商户外的商户作为合法商户。

6. 一种交易监测的装置,其特征在於,包括:

获取模块,用于获取预设时段内待识别商户的交易流水数据;

处理模块,用于将所述待识别商户的交易流水数据输入至识别模型,得到识别结果;所述识别结果用于指示所述待识别商户是否为非法商户;

所述识别模型的训练过程包括:

分别对非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,得到多维度特征的各样本数据;其中,所述多维度特征包括预设时间内与商户存在交易的交易账户数、预设时间内同一交易账户交易金额最小值、预设时间内交易金额大于阈值的交易占比、预设时间内借记卡占比、预设时间内交易时间点占比、预设时间内发卡机构所在地高危地区占比、预设时间内交易金额为整数金额的占比;

根据非法种子商户的交易流水数据对应的样本数据、非法关联商户的交易流水数据对应的样本数据中的任一样本数据,根据所述样本数据的初始识别结果、所述样本数据的样本属性确定出第一结果差异值;通过预设超参数以及所述样本数据在关联扩散中的关联代值对所述第一结果差异值进行加权,得到第二结果差异值;根据各样本的第二结果差异值,确定损失函数值;所述初始识别结果是所述样本数据输入至初始识别模型得到的;

根据所述损失函数值更新所述初始识别模型,直至得到所述识别模型;

所述非法关联商户是通过对所述非法种子商户进行关联扩散确定的;所述样本数据在关联扩散中的关联代值对应确定非法关联商户的次数。

7. 一种计算设备,其特征在於,包括:

存储器,用于存储程序指令;

处理器,用于调用所述存储器中存储的程序指令,按照获得的程序执行权利要求1至5任一项所述的方法。

8. 一种计算机可读存储介质,其特征在於,所述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用于使计算机执行权利要求1至5任一项所述的方法。

## 一种交易监测的方法及装置

### 技术领域

[0001] 本发明涉及金融领域,尤其涉及一种交易监测的方法及装置。

### 背景技术

[0002] 近些年,网络技术越来越发达,传统的人与人的交互可以在网络中进行实现,如桌面游戏等,同样的,非法网络赌博也日益猖獗,具体的,非法分子会建立赌博网站,并依托非法网络支付平台注册的大量虚假商户向赌客提供二维码渠道或无卡渠道的线上赌资充值服务。其中,网络赌博的关键环节之一就是赌客赌资充值,赌资充值主要是通过通过在赌博网站提供的线上虚假商户消费实现的。

[0003] 目前,针对网络赌博的监测主要包括两种方法,一种为有监督算法的网络赌博监测,另一种为无监督算法的网络赌博监测。但是,有监督算法的网络赌博监测需要大量准确的标签样本数据,其泛化能力低,对标签样本数据质量和数量过高,而大量准确的标签样本数据往往是很难获得的。无监督算法的网络赌博监测无需标签样本数据,但其识别精度低,无法准确识别出网络赌博商户。

[0004] 因此,现需要一种网络赌博交易监测的方法,提升网络赌博监测的泛化能力和准确度,减低所需样本数据数量。

### 发明内容

[0005] 本发明实施例提供一种交易监测的方法及装置,用于增加监测的泛化能力。

[0006] 第一方面,本发明实施例提供一种交易监测的方法,包括:

[0007] 获取预设时段内待识别商户的交易流水数据;

[0008] 将所述待识别商户的交易流水数据输入至识别模型,得到识别结果;所述识别结果用于指示所述待识别商户是否为非法商户;所述识别模型是通过非法种子商户的交易流水数据和非法关联商户的交易流水数据进行训练得到的;所述非法关联商户是通过对所述非法种子商户进行关联扩散确定的。

[0009] 上述技术方案中,通过已经确定非法的非法种子商户得到非法关联商户,以此扩展模型训练所需的样本数据,具体的,非法关联商户是根据非法种子商户进行关联扩散得到的,因此降低了所需样本的数据数量。通过扩展的样本数据训练模型,以增加模型的泛化能力,提升了监测待识别商户的范围,且对于无监督的学习方法来说,扩展的样本数据是根据已经确定非法的非法种子商户得到的,因此增加了监测非法商户的准确性。

[0010] 可选的,所述识别模型是通过非法种子商户的交易流水数据和非法关联商户的交易流水数据进行训练得到的,包括:

[0011] 分别对所述非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,得到各样本数据;其中,每个非法种子商户和每个非法关联商户均分别对应一个负样本属性;每个合法商户对一个正样本属性;

[0012] 将各样本数据分别输入至初始识别模型,得到各样本数据的初始识别结果;

[0013] 根据各样本数据的初始识别结果、各样本数据的样本属性及各样本数据在关联扩散中的关联代值,确定损失函数值;

[0014] 根据所述损失函数值更新所述初始识别模型,直至得到所述识别模型。

[0015] 上述技术方案中,针对商户的交易流水数据进行特征提取,得到商户对应的样本数据,其中,商户又分为非法商户和合法商户,通过非法商户和合法商户的样本数据对初始识别模型进行训练,以提升识别模型识别的准确性。具体的,初始识别模型的训练是根据样本数据在关联扩散中的关联代值确定的损失函数值进行收敛的,以此增加模型识别的泛化能力。

[0016] 可选的,根据各样本数据的初始识别结果、各样本数据的样本属性及各样本数据在关联扩散中的关联代值,确定损失函数值,包括:

[0017] 针对任一样本数据,根据所述样本数据的初始识别结果和所述样本数据的样本属性确定出第一结果差异值;

[0018] 通过预设超参数和样本数据在关联扩散中的关联代值对所述第一结果差异值进行加权,得到第二结果差异值;

[0019] 根据各样本的第二结果差异值确定损失函数值。

[0020] 上述技术方案中,损失函数值是根据各样本数据的初始识别结果、各样本数据的样本属性、预设超参数及各样本数据在关联扩散中的关联代值确定的,其中,样本数据在关联扩散中的关联代值用于对第一结果差异值进行加权,相当于对非法关联商户的样本数据进行加权,以此增加识别模型的泛化能力,提升了识别模型监测待识别商户的范围。

[0021] 可选的,根据下述公式(1)确定所述损失函数值;

$$[0022] \quad L(y_i, f(x_i)) = \frac{1}{n} \sum_{i=1}^n \left[ (1-\gamma)^l (y_i - f(x_i))^2 \right] \dots \dots \dots (1);$$

[0023] 其中, $L(y_i, f(x_i))$ 为所述损失函数值; $l$ 为第*i*个样本数据在关联扩散中的关联代值, $l$ 为自然数; $n$ 为样本数据的数量; $f(x_i)$ 为第*i*个样本数据的初始识别结果; $x_i$ 为第*i*个样本数据在所述初始识别模型中的输入值; $y_i$ 为第*i*个样本数据的样本属性; $\gamma$ 为预设超参数, $0 < \gamma < 1$ 。

[0024] 可选的,所述非法关联商户是通过所述非法种子商户进行关联扩散确定的,包括:

[0025] 确定与所述非法种子商户存在交易的各关联账户;

[0026] 根据账户关联商户的第一关联特征,从所述各关联账户中确定出疑似账户;

[0027] 确定与所述疑似账户存在交易的关联商户;

[0028] 根据商户关联账户的第二关联特征,从所述关联商户中确定出所述非法关联商户;

[0029] 将所述非法关联商户更新为非法种子商户,返回确定与所述非法种子商户存在交易的各关联账户的步骤,直至满足设定条件。

[0030] 上述技术方案中,通过在与非法种子商户存在交易的各关联账户中确定出疑似账户,又根据在与疑似账户存在交易的关联商户中确定出非法关联商户,即上述过程为关联扩散的过程,其中,设定条件可以是确定非法关联商户的数量或关联扩散的次数等,以此增加训练初始识别模型的数据样本。

[0031] 可选的,所述方法还包括:

[0032] 将所述关联商户中除所述非法关联商户外的商户作为合法商户。

[0033] 上述技术方案中,通过将关联商户分为非法商户和合法商户,以此使数据样本分为正样本和负样本,对于训练后的识别模型而言,提升了监测非法商户的准确性。

[0034] 可选的,所述账户关联商户的第一关联特征是指从账户的交易流水数据中确定出与商户关联的特征;

[0035] 所述商户关联账户的第二关联特征是指从商户的交易流水数据中确定出与账户关联的特征。

[0036] 上述技术方案中,因为第一关联特征是从账户的交易流水数据中确定出与商户关联的特征,进而提升了从关联账户中确定出疑似账户的准确性,因为第二关联特征是从商户的交易流水数据中确定出与账户关联的特征,进而提升了从关联商户中确定出非法关联商户的准确性。

[0037] 第二方面,本发明实施例提供一种交易监测的装置,包括:

[0038] 获取模块,用于获取预设时段内待识别商户的交易流水数据;

[0039] 处理模块,用于将所述待识别商户的交易流水数据输入至识别模型,得到识别结果;所述识别结果用于指示所述待识别商户是否为非法商户;所述识别模型是通过非法种子商户的交易流水数据和非法关联商户的交易流水数据进行训练得到的;所述非法关联商户是通过对所述非法种子商户进行关联扩散确定的。

[0040] 可选的,所述处理模块具体用于:

[0041] 分别对所述非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,得到各样本数据;其中,每个非法种子商户和每个非法关联商户均分别对应一个负样本属性;每个合法商户对一个正样本属性;

[0042] 将各样本数据分别输入至初始识别模型,得到各样本数据的初始识别结果;

[0043] 根据各样本数据的初始识别结果、各样本数据的样本属性及各样本数据在关联扩散中的关联代值,确定损失函数值;

[0044] 根据所述损失函数值更新所述初始识别模型,直至得到所述识别模型。

[0045] 可选的,所述处理模块具体用于:

[0046] 针对任一样本数据,根据所述样本数据的初始识别结果和所述样本数据的样本属性确定出第一结果差异值;

[0047] 通过预设超参数和样本数据在关联扩散中的关联代值对所述第一结果差异值进行加权,得到第二结果差异值;

[0048] 根据各样本的第二结果差异值确定损失函数值。

[0049] 可选的,根据下述公式(1)确定所述损失函数值;

$$[0050] \quad L(y_i, f(x_i)) = \frac{1}{n} \sum_{i=1}^n \left[ (1-\gamma)^l (y_i - f(x_i))^2 \right] \dots \dots \dots (1);$$

[0051] 其中, $L(y_i, f(x_i))$ 为所述损失函数值; $l$ 为第 $i$ 个样本数据在关联扩散中的关联代值, $l$ 为自然数; $n$ 为样本数据的数量; $f(x_i)$ 为第 $i$ 个样本数据的初始识别结果; $x_i$ 为第 $i$ 个样本数据在所述初始识别模型中的输入值; $y_i$ 为第 $i$ 个样本数据的样本属性; $\gamma$ 为预设超参数, $0 < \gamma < 1$ 。

- [0052] 可选的,所述处理模块具体用于:
- [0053] 确定与所述非法种子商户存在交易的各关联账户;
- [0054] 根据账户关联商户的第一关联特征,从所述各关联账户中确定出疑似账户;
- [0055] 确定与所述疑似账户存在交易的关联商户;
- [0056] 根据商户关联账户的第二关联特征,从所述关联商户中确定出所述非法关联商户;
- [0057] 将所述非法关联商户更新为非法种子商户,返回确定与所述非法种子商户存在交易的各关联账户的步骤,直至满足设定条件。
- [0058] 可选的,所述处理模块还用于:
- [0059] 将所述关联商户中除所述非法关联商户外的商户作为合法商户。
- [0060] 可选的,所述账户关联商户的第一关联特征是指从账户的交易流水数据中确定出与商户关联的特征;
- [0061] 所述商户关联账户的第二关联特征是指从商户的交易流水数据中确定出与账户关联的特征。
- [0062] 第三方面,本发明实施例还提供一种计算设备,包括:
- [0063] 存储器,用于存储程序指令;
- [0064] 处理器,用于调用所述存储器中存储的程序指令,按照获得的程序执行上述交易监测的方法。
- [0065] 第四方面,本发明实施例还提供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计算机可执行指令,所述计算机可执行指令用于使计算机执行上述交易监测的方法。

### 附图说明

[0066] 为了更清楚地说明本发明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要介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的普通技术人员来讲,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。

- [0067] 图1为本发明实施例提供的一种系统架构示意图;
- [0068] 图2为本发明实施例提供的一种交易监测的方法的流程示意图;
- [0069] 图3为本发明实施例提供的一种交易监测的方法的流程示意图;
- [0070] 图4为本发明实施例提供的一种构建初始识别模型的流程示意图;
- [0071] 图5为本发明实施例提供的一种交易监测的装置的机构示意图。

### 具体实施方式

[0072] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0073] 近几年,网络非法行为日益猖獗,例如,不法分子建立赌博网站,以正常商户作为

幌子,实际用于网络赌博充值、贩卖非法物品(如武器弹药,毒品等)等非法行为。具体的,以网络赌博为例,不法分子依托非法网络支付平台注册非法赌博商户的二维码,以使网络赌博客户通过二维码在网络上充值赌资。

[0074] 现有技术中,为了更好的监测出非法商户,一般是通过机器学习算法构建模型,通过构建的模型对商户进行识别,确定出非法商户。在构建模型时,需要使用训练样本对模型进行训练,而训练样本目前一般包括两种,无标签训练样本和有标签训练样本,即通过无标签训练样本训练模型为无监督学习算法,通过有标签训练样本训练模型为有监督学习算法,例如以下示例。

[0075] 1、基于无监督学习算法的监测:通过网络理论与网络非法交易特征相结合,根据商户与用户间交易关系构成网络,然后对商户进行社团划分后对可疑非法商户进一步验证。其中,无监督学习算法可以为均值聚类算法、层次聚类算法和密度聚类算法等。例如,通过均值聚类算法训练模型,使训练后的模型可以将训练样本划分为多个社团,然后使用该模型确定出待识别商户所在社团。

[0076] 但是,上述方法中,一般是对商户和用户间的关系进行训练模型,然后基于出入度、中心度、连通等学习算法将商户和用户间的关系图划分为多个社团子图,再根据商户和用户间的关联关系识别可疑社团,但因为样本数据为无标签的样本,因此较难区分多个风险场景的不同类型可疑社团,且识别精度较低。

[0077] 2、基于有监督学习算法的监测:根据已确定非法的种子非法商户进行二度关联,得到可疑关联商户和可疑关联用户,利用相似度比较算法(如链路预测森林算法、协同过滤算法等)比较种子网络赌博商户和可疑关联商户的相似度,以此确定可疑关联商户是否为网络赌博商户。

[0078] 但是,上述方法中,基于相似度算法的监测模型计算种子非法商户和可疑关联商户的相似度从而给可疑关联商户的可疑度打分,从而确定可疑关联商户是否为网络赌博商户。但是通过该方法能识别的非法商户范围较小,存在局限性,只能识别与种子非法商户直接关联的商户,无法对全量商户进行识别。

[0079] 3、基于图特征的有监督学习监测:对有标签的非法样本数据构建二部图,提取非法商户的图特征作为模型的输入,训练出网络非法商户监测模型,通过该模型确定商户是否为非法商户。

[0080] 但是,上述方法中,构建监测模型时,需要大量带标签样本数据,且对样本质量要求很高,而在很多非法种类中,大量的质量很高的带标签样本数据是难以得到的。

[0081] 因此,现需要一种非法交易监测的方法,通过少量的带标签样本数据,提升网络赌博监测的泛化能力,且相对于无监督学习的监测方法,提升准确度,相对于有监督学习的监测方法,降低所需带标签样本数据的数量。

[0082] 图1示例性的示出了本发明实施例所适用的一种系统架构,该系统架构包括非法种子商户采集模块110、关联拓展模块120、数据准备模块130、特征计算模块140、模型训练模块150、模型识别模块160、数据获取模块170和模型输出模块180。

[0083] 其中,非法种子商户采集模块110,用于通过信息共享、舆论感知、实际测试、数据爬取、外部报送等方式收集非法种子商户。

[0084] 关联拓展模块120,用于通过获取的非法种子商户的交易流水数据关联出非法关

联商户和合法商户。

[0085] 数据准备模块130,用于获取非法种子商户的交易流水数据、非法关联商户的交易流水数据段和合法商户的交易流水数据。

[0086] 特征计算模块140,用于根据预设维度特征,根据非法种子商户的交易流水数据、非法关联商户的交易流水数据段和合法商户的交易流水数据确定出各样本数据。

[0087] 模型训练模块150,用于根据各样本数据训练模型,确定出识别模型。

[0088] 模型识别模块160,用于根据待识别商户的交易流水数据确定待识别商户是否为非法商户。

[0089] 数据获取模块170,用于获取预设时段内待识别商户的交易流水数据,并输入至模型识别模块160。

[0090] 模型输出模块180,用于输出数据获取模块170获取的待识别商户的识别结果。

[0091] 需要说明的是,上述图1所示的结构仅是一种示例,本发明实施例对此不做限定。

[0092] 基于上述描述,图2示例性的示出了本发明实施例提供的一种交易监测的方法的流程示意图,该流程可由交易监测的装置执行。

[0093] 如图2所示,该流程具体包括:

[0094] 步骤210,获取预设时段内待识别商户的交易流水数据。

[0095] 本发明实施例中,利用网络获取数据的手段,在预设时段内获取待识别商户的交易流水数据,例如,通过信息共享、数据爬取、外部报送等方式获取当前时间之前12小时内待识别商户的交易流水数据。

[0096] 步骤220,将所述待识别商户的交易流水数据输入至识别模型,得到识别结果。

[0097] 本发明实施例中,识别结果用于指示待识别商户是否为非法商户,例如,识别结果为0或1,“1”指示待识别商户是非法商户,“0”指示待识别商户是合法商户。或者,识别结果直接输出待识别商户为非法商户或合法商户,例如,识别模型根据待识别商户的交易流水数据计算出非法值,在非法值大于非法阈值时,将待识别商户确定为非法商户,反之,在非法值不大于非法阈值时,将待识别商户确定为合法商户。

[0098] 需要说明的是,上述识别模型是通过非法种子商户的交易流水数据和非法关联商户的交易流水数据进行训练得到的。

[0099] 在本发明实施例中,根据非法种子商户的交易流水数据和非法关联商户的交易流水数据确定出训练模型所需的训练样本,进而通过训练样本对模型进行训练,得到识别模型。

[0100] 进一步地,分别对非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,得到各样本数据,其中,每个非法种子商户和每个非法关联商户均分别对应一个负样本属性,每个合法商户对一个正样本属性,然后将各样本数据分别输入至初始识别模型,得到各样本数据的初始识别结果,再根据各样本数据的初始识别结果、各样本数据的样本属性及各样本数据在关联扩散中的关联代值,确定损失函数值,最后根据损失函数值更新初始识别模型,直至得到识别模型。

[0101] 需要说明的是,每个非法种子商户和每个非法关联商户均被认为是非法商户,以此其分别对应一个负样本属性,负样本属性可以为一个值,以指示每个非法种子商户和每个非法关联商户是非法商户的准确率,例如每个非法种子商户的负样本属性为100%,即非

法种子商户一定为非法商户,某一非法关联商户的负样本属性为90%,即该非法种子商户高疑似为非法商户,同理,若某一商户对应的正样本属性为10%,即该商户高疑似为合法商户。

[0102] 在模型收敛时,一般是根据模型对应机器学习算法的函数损失值确定,在不同的机器学习算法中,其对应的函数损失值的确定方法亦不相同,在本发明实施例中,损失函数值可以根据各样本数据的初始识别结果、各样本数据的样本属性、预设超参数及各样本数据在关联扩散中的关联代值确定。

[0103] 本发明实施例中,非法关联商户是通过对非法种子商户进行关联扩散确定的,因此,各样本数据在关联扩散中的关联代值是根据非法关联商户与非法种子商户之间的关联扩散确定的。

[0104] 具体的,确定与非法种子商户存在交易的各关联账户,再根据账户关联商户的第一关联特征,从各关联账户中确定出疑似账户,然后确定与疑似账户存在交易的关联商户,再根据商户关联账户的第二关联特征,从关联商户中确定出非法关联商户,最后将非法关联商户更新为非法种子商户,返回确定与非法种子商户存在交易的各关联账户的步骤,直至满足设定条件。

[0105] 其中,设定条件可以是非法种子商户的数量或确定出非法关联商户的次数等,相应的,样本数据在关联扩散中的关联代值可以是对应确定非法关联商户的次数。为了更好的解释上述技术方案,下面再具体实例中进行阐述。

[0106] 实例1

[0107] 假设与非法种子商户A存在交易的关联账户包括a1、a2和a3,根据第一关联特征在a1、a2和a3中确定出疑似账户a2和a3,然后再确定与疑似账户a2和a3存在交易的关联商户,其中,与疑似账户a2存在交易的关联商户为B、C和D,与疑似账户a3存在交易的关联商户为B、C和F,再根据第二关联特征从关联商户B、C、D和F中确定出非法关联商户B和C,因为商户B和C是第一次确定出的非法关联商户,所以,若将非法种子商户A的样本数据在关联扩散中的关联代值确定为0,则非法关联商户B和C的样本数据在关联扩散中的关联代值可以确定为1。

[0108] 在确定出非法关联商户B和C之后,将非法关联商户B和C更新为非法种子商户,若设定条件为非法种子商户的数量阈值为5,则根据上述方法,确定出非法种子商户B和C的非法关联商户X、Y和Z,因为商户X、Y和Z相对于非法种子商户A是第二次确定出的非法关联商户,因此,非法关联商户X、Y和Z的样本数据在关联扩散中的关联代值可以确定为2,然后将非法关联商户X、Y和Z更新为非法种子商户,此时,非法种子商户的数量为6,大于设定条件,因此可以结束上述方法流程。

[0109] 需要说明的是,在将关联商户中除非法关联商户外的商户作为合法商户,例如上述实例1中,商户D和F为合法商户。

[0110] 本发明实施例中,账户关联商户的第一关联特征是指从账户的交易流水数据中确定出与商户关联的特征。例如,各交易账户在第一历史时段内,与同一商户的交易次数最大值、与所述第一交易账户存在交易的商户数、交易金额以及交易金额的交易频率。

[0111] 根据上述实例1进行举例,例如,网络赌博中,一般会提供用户固定的充值金额,如30元,50元等金额选项,因此,在与非法种子商户A存在交易的关联账户a1、a2和a3中,a2和

a3的交易金额为预设的固定金额,且以每天为单位的交易频率大于频率阈值时,将账户a2和a3确定为疑似账户。

[0112] 商户关联账户的第二关联特征是指从商户的交易流水数据中确定出与账户关联的特征。例如,各商户第二历史时段内,商户中疑似账户的平均交易天数、平均交易重复率、疑似账户的交易金额以及交易金额的交易频率,其中平均交易重复率是各预设单位时间内的交易重复率的平均值,交易重复率是在预设单位时间内与商户存在交易的疑似账户的数量和疑似账户总数量的比值。

[0113] 根据上述实例1进行举例,例如,在关联商户B、C、D和F中,因疑似账户a2和a3均与商户B和C存在交易,且交易金额满足条件,交易频率满足条件等,则将商户B和C确定为非法关联商户。需要说明的是,在确定非法关联商户时,还可以根据商户自身特征进行确认,提升确定非法商户的准确性,例如,商户B和C满足确定为非法关联商户的条件,但其中商户B的自身特征为公知合法商户,地铁,公交等,则商户B即使满足确定为非法关联商户的条件,但其仍为合法商户。

[0114] 本发明实施例中,在确定出非法种子商户、非法关联商户和合法商户之后,对非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取时,可以根据预设维度的特征进行提取,以使各样本数据的是同维度的数据,而维度可以根据交易流水数据进行预先设置。例如,交易流水数据是针对赌博的交易,因此,可以将维度特征设置为卡维度特征、商户维度特征、交易维度特征、时间维度特征、地理维度特征等,以此得到该交易流水数据对应的样本数据。

[0115] 具体的,维度特征可以包括预设时间内与商户存在交易的交易账户数、预设时间内同一交易账户交易金额最小值、预设时间内交易金额大于阈值的交易占比、预设时间内借记卡占比、预设时间内交易时间点占比、预设时间内发卡机构所在地高危地区占比、预设时间内交易金额为整数金额的占比等。其中,预设时间可以相同,可以不同,在此不做具体限定。

[0116] 本发明实施例中,根据各样本数据来确定模型的损失函数值,进一步地,针对任一样本数据,根据样本数据的初始识别结果和样本数据的样本属性确定出第一结果差异值,通过预设超参数和样本数据在关联扩散中的关联代值对第一结果差异值进行加权,得到第二结果差异值,根据各样本的第二结果差异值确定损失函数值。

[0117] 具体的,根据下述公式(1)确定所述损失函数值。

$$[0118] \quad L(y_i, f(x_i)) = \frac{1}{n} \sum_{i=1}^n \left[ (1-\gamma)^l (y_i - f(x_i))^2 \right] \dots \dots \dots (1);$$

[0119] 其中,L(y<sub>i</sub>, f(x<sub>i</sub>))为损失函数值;l为第i个样本数据在关联扩散中的关联代值,l为自然数;n为样本数据的数量;f(x<sub>i</sub>)为第i个样本数据的初始识别结果;x<sub>i</sub>为第i个样本数据在所述初始识别模型中的输入值;y<sub>i</sub>为第i个样本数据的样本属性;γ为预设超参数,0<γ<1。

[0120] 本发明实施例中,通过已经确定非法的非法种子商户得到非法关联商户,因此降低了所需样本的数据数量,扩展了模型训练所需的样本数据,具体的,针对商户的交易流水数据进行特征提取,得到商户对应的样本数据,其中,商户又分为非法商户和合法商户,通过非法商户和合法商户的样本数据对初始识别模型进行训练,以提升识别模型识别的准确

性。进一步地的,初始识别模型的训练是根据样本数据在关联扩散中的关联代值确定的损失函数值进行收敛的,以此增加模型识别的泛化能力,提升了监测待识别商户的范围。

[0121] 为了更好的解释上述技术方案,下面再具体实例中阐述上述技术方案。

[0122] 实例2

[0123] 图3示例性的示出了一种交易监测的方法的流程示意图,如图3所示,具体流程包括:

[0124] 步骤310,获取赌博种子商户。

[0125] 本实例以网络赌博为例,通过信息共享、舆论感知、实际测试、数据爬取、外部报送等方式收集非法网络支付平台为赌博网站提供的赌博商户的二维码,通过二维码获取赌博商户信息,如商户编号、商户名称、网络赌博商户主要交易方式等,并将其确定为赌博种子商户。

[0126] 步骤320,拓展赌博关联商户。

[0127] 在与非法种子商户存在交易的各关联账户中确定出疑似赌博账户,在与疑似赌博账户存在交易的关联商户中确定出赌博关联商户以及合法商户。

[0128] 步骤330,确定样本数据。

[0129] 分别对赌博种子商户的交易流水数据、赌博关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,其中,交易流水数据包括但不限于交易账户、交易金额、交易日期、交易时间、交易类型、交易商户、交易账户所属机构、商户收单机构等字段,进而得到各样本数据,其中,样本数据为预设的多维度特征的样本数据。

[0130] 步骤340,生成识别模型。

[0131] 根据各样本数据对初始识别模型进行训练,根据其损失函数值生成识别模型,其中,初始识别模型可以是根据多种机器学习算法(如决策树算法、最小二乘法和逻辑回归算法等)构建的,示例性的,本实例中根据决策树算法构建初始识别模型。图4示例性的示出了一种构建初始识别模型的流程示意图,如图4所示。

[0132] 步骤410,获取样本数据。

[0133] 获取上述步骤330中确定的各样本数据,其中,非法种子商户和非法关联商户的样本数据可以为负样本数据,合法商户的样本数据可以为正样本数据。

[0134] 步骤420,构建直方图。

[0135] 通过直方图算法将各样本数据中连续的浮点特征离散为 $k$ 个离散值,构建出宽度为 $k$ 的直方图, $k$ 为常数。

[0136] 步骤430,选取分割点。

[0137] 根据带有深度限制的按叶子生长(laef-wise)算法,选取出最优分割点。

[0138] 步骤440,生成识别模型。

[0139] 根据选取的分割点,进行重复迭代,直至生成一个 $N$ 层的提升树,进而生成识别模型,其中, $N$ 为自然数。

[0140] 步骤350,确定待识别商户的识别结果。

[0141] 获取当前时间的前6个小时至12个小时内的商户作为待识别商户,并得到待识别商户的交易流水数据,再将待识别商户的交易流水数据输入至识别模型,确定出各待识别商户的风险值,根据风险值和风险阈值确定识别结果,如待识别商户 $S$ 的风险值大于风险阈

值,则待识别商户S的识别结果定为非法商户。

[0142] 基于相同的技术构思,图5示例性的示出了本发明实施例提供的一种交易监测的装置的机构示意图,该装置可以执行上述交易监测的方法的流程。

[0143] 如图5所示,该装置具体包括:

[0144] 获取模块510,用于获取预设时段内待识别商户的交易流水数据;

[0145] 处理模块520,用于将所述待识别商户的交易流水数据输入至识别模型,得到识别结果;所述识别结果用于指示所述待识别商户是否为非法商户;所述识别模型是通过非法种子商户的交易流水数据和非法关联商户的交易流水数据进行训练得到的;所述非法关联商户是通过对所述非法种子商户进行关联扩散确定的。

[0146] 可选的,所述处理模块520具体用于:

[0147] 分别对所述非法种子商户的交易流水数据、非法关联商户的交易流水数据及合法商户的交易流水数据进行特征提取,得到各样本数据;其中,每个非法种子商户和每个非法关联商户均分别对应一个负样本属性;每个合法商户对一个正样本属性;

[0148] 将各样本数据分别输入至初始识别模型,得到各样本数据的初始识别结果;

[0149] 根据各样本数据的初始识别结果、各样本数据的样本属性及各样本数据在关联扩散中的关联代值,确定损失函数值;

[0150] 根据所述损失函数值更新所述初始识别模型,直至得到所述识别模型。

[0151] 可选的,所述处理模块520具体用于:

[0152] 针对任一样本数据,根据所述样本数据的初始识别结果和所述样本数据的样本属性确定出第一结果差异值;

[0153] 通过预设超参数和样本数据在关联扩散中的关联代值对所述第一结果差异值进行加权,得到第二结果差异值;

[0154] 根据各样本的第二结果差异值确定损失函数值。

[0155] 可选的,根据下述公式(1)确定所述损失函数值;

$$[0156] \quad L(y_i, f(x_i)) = \frac{1}{n} \sum_{i=1}^n \left[ (1-\gamma)^l (y_i - f(x_i))^2 \right] \dots \dots \dots (1);$$

[0157] 其中, $L(y_i, f(x_i))$ 为所述损失函数值; $l$ 为第*i*个样本数据在关联扩散中的关联代值, $l$ 为自然数; $n$ 为样本数据的数量; $f(x_i)$ 为第*i*个样本数据的初始识别结果; $x_i$ 为第*i*个样本数据在所述初始识别模型中的输入值; $y_i$ 为第*i*个样本数据的样本属性; $\gamma$ 为预设超参数, $0 < \gamma < 1$ 。

[0158] 可选的,所述处理模块520具体用于:

[0159] 确定与所述非法种子商户存在交易的各关联账户;

[0160] 根据账户关联商户的第一关联特征,从所述各关联账户中确定出疑似账户;

[0161] 确定与所述疑似账户存在交易的关联商户;

[0162] 根据商户关联账户的第二关联特征,从所述关联商户中确定出所述非法关联商户;

[0163] 将所述非法关联商户更新为非法种子商户,返回确定与所述非法种子商户存在交易的各关联账户的步骤,直至满足设定条件。

[0164] 可选的,所述处理模块520还用于:

[0165] 将所述关联商户中除所述非法关联商户外的商户作为合法商户。

[0166] 可选的,所述账户关联商户的第一关联特征是指从账户的交易流水数据中确定出与商户关联的特征;

[0167] 所述商户关联账户的第二关联特征是指从商户的交易流水数据中确定出与账户关联的特征。

[0168] 基于相同的技术构思,本发明实施例还提供一种计算设备,包括:

[0169] 存储器,用于存储程序指令;

[0170] 处理器,用于调用所述存储器中存储的程序指令,按照获得的程序执行上述交易监测的方法。

[0171] 基于相同的技术构思,本发明实施例还提供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计算机可执行指令,所述计算机可执行指令用于使计算机执行上述交易监测的方法。

[0172] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0173] 本申请是参照根据本申请的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0174] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0175] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0176] 显然,本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样,倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内,则本申请也意图包含这些改动和变型在内。

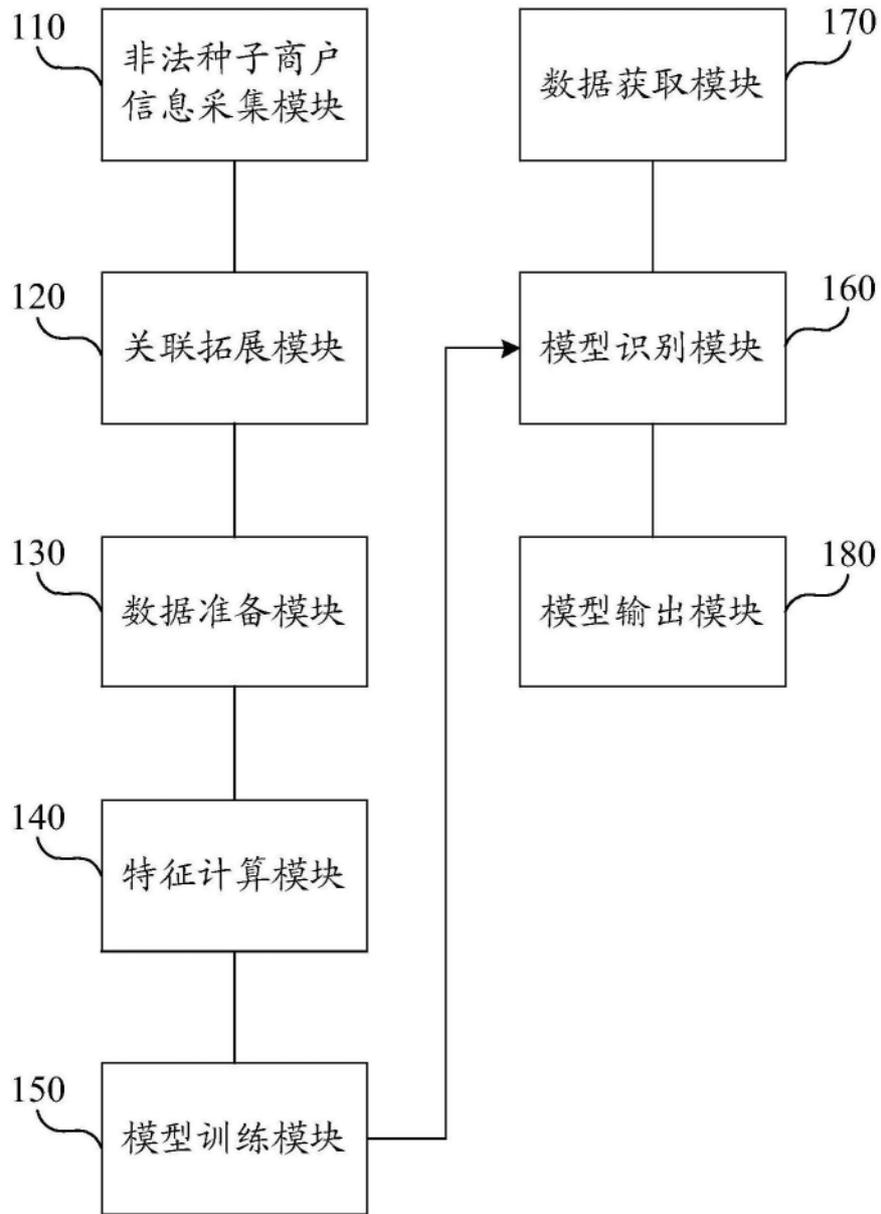


图1

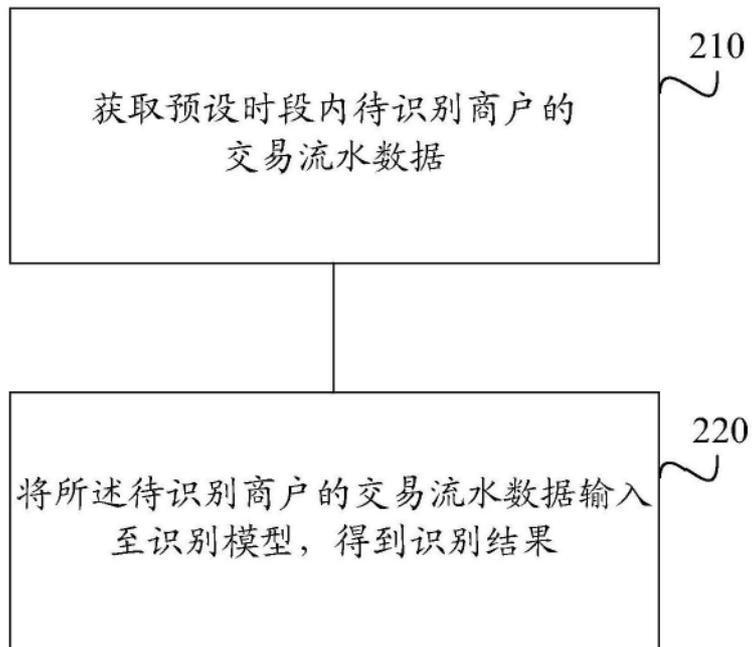


图2

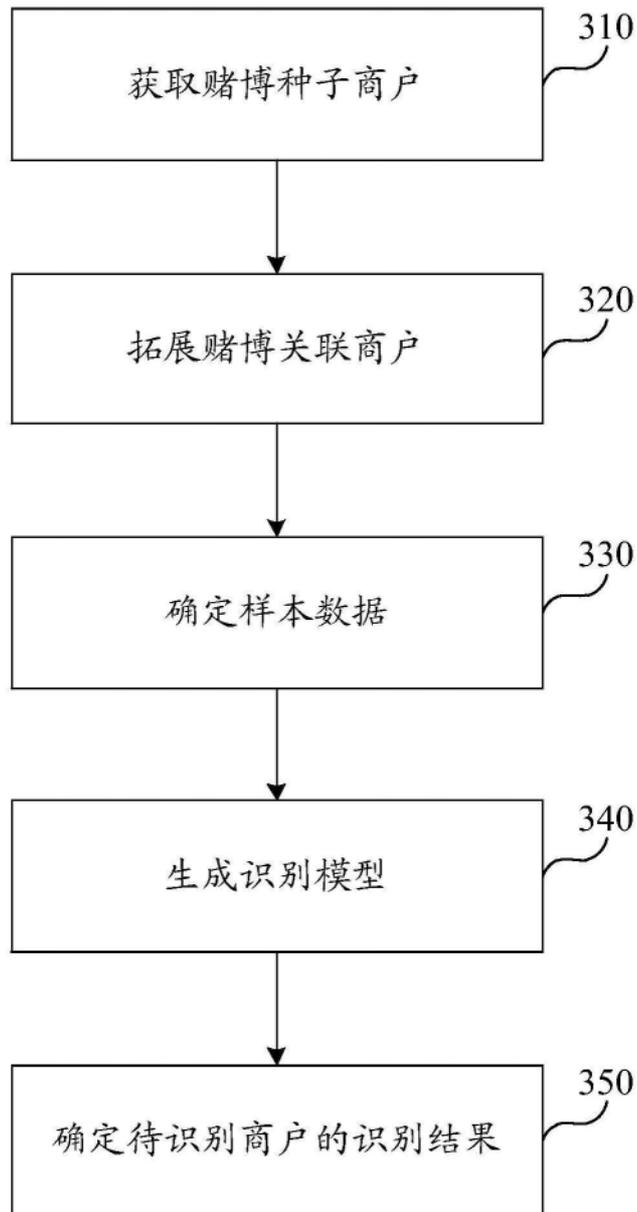


图3

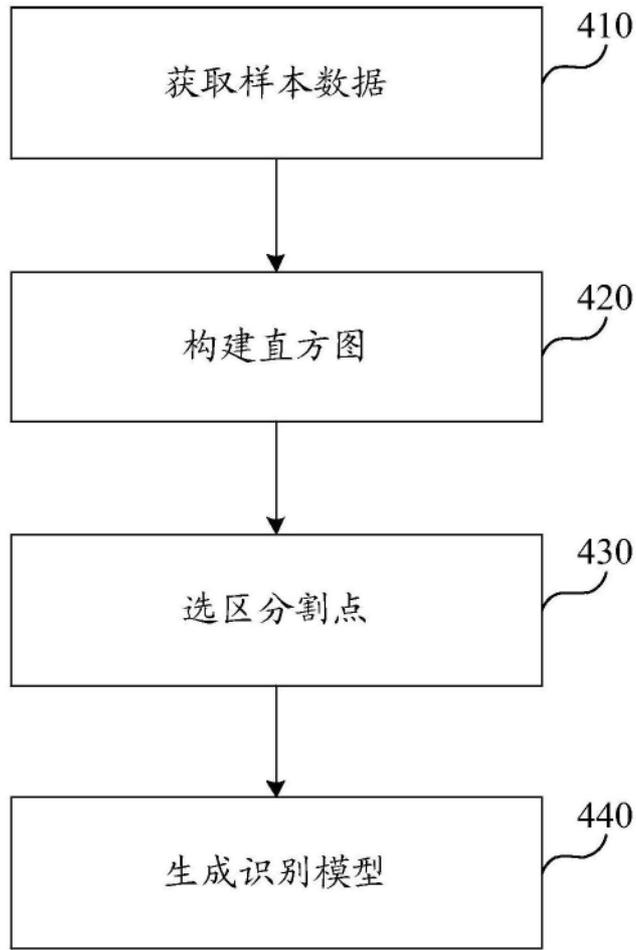


图4

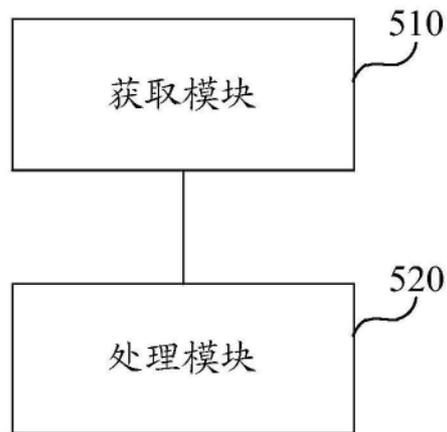


图5