(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107357790 B
(45)授权公告日 2021.06.04

(21)申请号 201610302742.9

(22)申请日 2016.05.09

(65)同一申请的已公布的文献号
申请公布号 CN 107357790 A

(43)申请公布日 2017.11.17

(73)专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72)发明人 靳玉康 方亮 许涵斌

(74)专利代理机构 北京三友知识产权代理有限公司 11127
代理人 李辉

(51)Int.Cl.
G06F 16/9532(2019.01)
G06F 16/958(2019.01)

(56)对比文件
CN 105335354 A,2016.02.17
US 2015237061 A1,2015.08.20
Fei Tony Liu、Kai Ming Ting、Zhi-Hua Zhou.“Isolation-based Anomaly Detection”.《ACM Transactions on Knowledge Discovery from Data》.2012,

审查员 胡赢

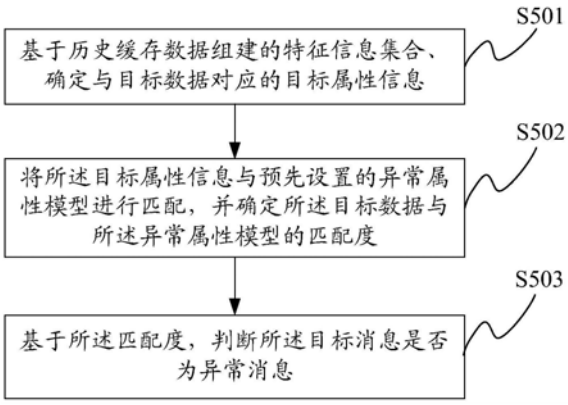
权利要求书5页 说明书19页 附图7页

(54)发明名称

一种异常消息检测方法、装置及系统

(57)摘要

本申请提供了一种异常消息检测方法、装置和系统，其中方法包括：基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息；其中，目标数据用于表示待检测的目标消息；将目标属性信息与预先设置的异常属性模型进行匹配，并确定目标数据与异常属性模型的匹配度；基于所述匹配度，判断所述目标消息是否为异常消息。由于特征信息集合较难发生改变，所以从特征信息集合中提取得到的属性信息也不会轻易改变。因此，本申请提供属性信息对比方式，可以准确确定异常消息，进而方便技术人员依据异常消息解决欺诈问题。



CN 107357790 B

1.一种异常消息检测方法,其特征在于,包括:

基于历史缓存数据组建的特征信息集合确定与目标数据对应的目标属性信息;其中,所述目标数据用于表示待检测的目标消息;所述特征信息集合包括共用特征信息以及目标发送方标识对应的目标私有特征信息;所述目标发送方标识根据用于表示目标消息发送方的目标发送方账户标识以及用于表示所述目标消息发送方使用的发送方机器的目标发送方机器标识确定;所述目标私有特征信息包括目标发送方账户利用目标发送方机器发送所有消息第一总数量、向卖家发送所有消息的第一卖家消息数量、向买家发送所有消息的第一买家消息数量、目标发送方机器标识被举报的第一举报次数、目标发送方账户注册时的第一地理位置标识、各个历史消息的发送次数;所述共用特征信息包括历史欺诈机器标识列表、各个历史举报信息以及各个接收方身份标识;所述目标属性信息包括第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度以及用于表示目标接收方账户为卖家或买家的目标接收方身份标识;相应的,所述确定与目标数据对应的目标属性信息,包括:将所述目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述第一买家消息数量、所述第一举报次数和所述第一地理位置标识,分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识;在所述目标私有特征信息的各个历史消息中存在与所述目标消息一致的历史目标消息的情况下,将所述历史目标消息对应的第一发送次数赋值于所述第二发送次数;若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识,则将表示欺诈机器的第一标识赋值于所述欺诈机器标识,否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识;计算所述共用特征信息中各个举报消息与所述目标消息的相似度,将各个相似度中的最高相似度赋值于所述目标属性信息中的所述相似度;在所述共用特征信息中各个接收方身份标识中,查找与所述目标接收方账户标识对应的身份标识,并将该身份标识赋予所述目标接收方身份标识;

将所述目标属性信息与预先设置的异常属性模型进行匹配,并确定所述目标数据与所述异常属性模型的匹配度;其中,所述异常属性模型基于异常森林模型构建;

基于所述匹配度,判断所述目标消息是否为异常消息。

2.如权利要求1所述的方法,其特征在于,所述基于所述匹配度判断所述目标消息是否为异常消息,包括:

对所述匹配度进行归一化处理;

若所述匹配度归一化处理后数据值大于预设数据值,则确定所述目标消息为异常消息。

3.如权利要求1所述的方法,其特征在于,所述异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成,每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定。

4.如权利要求3所述的方法,其特征在于,所述目标数据包括:

用于表示所述目标消息发送方的目标发送方账户标识、用于表示所述发送方使用的发送方机器的目标发送方机器标识、用于表示所述目标消息的接收方的目标接收方账户标识

和所述目标消息。

5.如权利要求4所述的方法,其特征在于,所述基于历史缓存数据组建的特征信息集合确定与目标数据对应的目标属性信息,包括:

利用所述目标数据中的目标发送方账户标识和目标发送方机器标识,计算目标发送方标识;

获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息,以利用所述目标私有特征信息和共用特征信息确定与所述目标数据对应的目标属性信息。

6.如权利要求1所述的方法,其特征在于,所述异常属性模型包括有多个异常树iTree组成的异常森林iForest,每个iTree包括异常数据集合中若干个异常数据的属性信息中的属性值;则所述将所述目标属性信息与预先设置的异常属性模型进行匹配,并确定所述目标数据与所述异常属性模型的匹配度,包括:

将所述目标属性信息中的属性值与每个itree中的属性值进行对比,确定所述目标属性信息与每个iTree的匹配度;

将所有iTree对应的匹配度的综合值,确定为所述目标数据与所述异常属性模型的匹配度。

7.如权利要求6所述的方法,其特征在于,所述iTree具有预设最大高度,所述iTree的每层对应一个属性,每个节点对应一个属性值;则所述将所述目标属性信息中的属性值与每个iTree中的属性值进行对比,确定所述目标属性信息与每个iTree的匹配度,包括:

从iTree的根节点开始由上至下遍历iTree;

获取iTree的一个节点对应属性以及第一属性值,以及所述目标属性信息中相同属性对应的第二属性值;

判断第一属性值与第二属性值是否一致;

若所述第一属性值与所述第二属性值的误差在预设范围内,则进入下一层节点;重新进入获取iTree的一个节点对应属性以及第一属性值,以及所述目标属性信息中相同属性对应的第二属性值的步骤;

若所述第一属性值与所述第二属性值的误差不在预设范围内,则遍历同层的其它节点,若第一属性值与同层其它节点的属性值均不一致,则停止遍历;

将当前层与根节点之间的层数,确定为所述目标属性信息与该iTree的匹配度。

8.如权利要求6所述的方法,其特征在于,所述每个iTree构建过程包括:

由根节点开始从上至下构建iTree的每个节点:

步骤1:随机选择一个执行属性,并在异常数据集合的剩余异常数据中随机选择一个执行异常数据,并将执行异常数据的属性信息中与执行属性对应执行属性值确定为一个节点;

步骤2:在剩余异常数据中排除执行异常数据;

步骤3:按执行属性的执行属性值对剩余异常数据进行分类;具体包括:将执行属性的属性值小于执行属性值的异常数据归属于左子树,将执行属性的属性值大于执行属性值的异常数据归属于右子树;

重复执行步骤1、步骤2和步骤3,递归的构造左子树和右子树,直到满足以下条件之一

则终止:条件1:用于构造iTree的剩余异常数据只有一个异常数据或者多个相同的异常数据;

条件2:iTree的高度达到预设高度。

9.如权利要求1所述的方法,其特征在于,还包括:

在确定所述目标数据为异常数据后,更新异常数据集合。

10.如权利要求4所述的方法,其特征在于,

所述目标发送方机器标识包括发送方机器的MAC地址和发送方机器的硬盘号码;

在所述目标消息为文本情况,所述目标数据中目标消息为文本内容,在所述目标消息为图片时,所述目标数据中的目标消息为图片的MD5值。

11.如权利要求1所述的方法,其特征在于,所述各个历史消息利用历史消息的MD5值进行存储。

12.一种异常消息检测装置,其特征在于,包括:

第一确定属性单元,用于基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息;其中,所述目标数据用于表示待检测的目标消息;所述特征信息集合包括共用特征信息以及目标发送方标识对应的目标私有特征信息;所述目标发送方标识根据用于表示目标消息发送方的目标发送方账户标识以及用于表示所述目标消息发送方使用的发送方机器的目标发送方机器标识确定;所述目标私有特征信息包括目标发送方账户利用目标发送方机器发送所有消息第一总数量、向卖家发送所有消息的第一卖家消息数量、向买家发送所有消息的第一买家消息数量、目标发送方机器标识被举报的第一举报次数、目标发送方账户注册时的第一地理位置标识、各个历史消息的发送次数;所述共用特征信息包括历史欺诈机器标识列表、各个历史举报信息以及各个接收方身份标识;所述目标属性信息包括第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度以及用于表示目标接收方账户为卖家或买家的目标接收方身份标识;相应的,所述确定与目标数据对应的目标属性信息,包括:将所述目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述第一买家消息数量、所述第一举报次数和所述第一地理位置标识,分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识;在所述目标私有特征信息的各个历史消息中存在与所述目标消息一致的历史目标消息的情况下,将所述历史目标消息对应的第一发送次数赋值于所述第二发送次数;若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识,则将表示欺诈机器的第一标识赋值于所述欺诈机器标识,否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识;计算所述共用特征信息中各个举报消息与所述目标消息的相似度,将各个相似度中的最高相似度赋值于所述目标属性信息中的所述相似度;在所述共用特征信息中各个接收方身份标识中,查找与所述目标接收方账户标识对应的身份标识,并将该身份标识赋予所述目标接收方身份标识;

匹配单元,用于将所述目标属性信息与预先设置的异常属性模型进行匹配,并确定所述目标数据与所述异常属性模型的匹配度;

判断单元,用于基于所述匹配度,判断所述目标消息是否为异常消息。

13.如权利要求12所述的装置，其特征在于，所述第一确定属性单元，包括：

计算单元，用于利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；

获取单元，用于获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息，以利用所述目标私有特征信息和共用特征信息确定与所述目标数据对应的目标属性信息。

14.如权利要求12所述的装置，其特征在于，所述异常属性模型包括有多个异常树iTree组成的异常森林iForest，每个iTree包括异常数据集合中若干个异常数据的属性信息中的属性值；则匹配单元，包括：

对比单元，用于将所述目标属性信息中的属性值与每个itree中的属性值进行对比，确定所述目标属性信息与每个iTree的匹配度；

确定匹配度单元，用于将所有iTree对应的匹配度的综合值，确定为所述目标数据与所述异常属性模型的匹配度。

15.一种异常消息检测系统，其特征在于，包括：处理设备和与所述处理设备相连的多个缓存服务器；

其中，所述多个缓存服务器，用于存储基于历史缓存数据组建的特征信息集合；每个缓存服务器中存储有共用特征信息以及与发送方标识对应的私有特征信息；所述发送方标识根据用于表示消息发送方的发送方账户标识以及用于表示消息发送方使用的发送方机器的发送方机器标识确定；所述私有特征信息包括发送方账户利用发送方机器发送所有消息第一总数量、向卖家发送所有消息的第一卖家消息数量、向买家发送所有消息的第一买家消息数量、发送方机器标识被举报的第一举报次数、发送方账户注册时的第一地理位置标识、各个历史消息的发送次数；所述共用特征信息包括历史欺诈机器标识列表、各个历史举报信息以及各个接收方身份标识；

所述处理设备，用于从所述多个缓存服务器中确定与目标数据对应的目标属性信息；其中，所述目标数据用于表示待检测的目标消息；所述目标属性信息包括第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度以及用于表示目标接收方账户为卖家或买家的目标接收方身份标识；相应的，所述确定与目标数据对应的目标属性信息，包括：将目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述第一买家消息数量、所述第一举报次数和所述第一地理位置标识，分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识；在所述目标私有特征信息的各个历史消息中存在与所述目标消息一致的历史目标消息的情况下，将所述历史目标消息对应的第一发送次数赋值于所述第二发送次数；若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识，则将表示欺诈机器的第一标识赋值于所述欺诈机器标识，否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识；计算所述共用特征信息中各个举报消息与所述目标消息的相似度，将各个相似度中的最高相似度赋值于所述目标属性信息中的所述相似度；在所述共用特征信息中各个接收方身份标识中，查找与所述目标接收方账户标识对应的身份标识，并将该身份标识赋予所述

目标接收方身份标识；

所述处理设备，还用于将所述目标属性信息与预先设置的异常属性模型进行匹配，并确定所述目标数据与所述异常属性模型的匹配度；其中，所述异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成，每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定；若对所述匹配度进行归一化的数据值大于预设数据值，则确定所述目标消息为异常消息。

16.如权利要求15所述的系统，其特征在于，所述处理设备包括：第一服务器；所述第一服务器，具体用于利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；并根据预先存储的发送方标识与缓存服务器标识的对应关系，确定与所述目标发送方标识对应的目标缓存服务器，向所述目标缓存服务器发送目标发送方标识；基于共用特征信息和目标私有特征信息确定与目标数据对应的目标属性信息；

所述目标缓存服务器，用于获取共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；并将所述共用特征信息和目标私有特征信息发送至所述处理设备。

17.如权利要求15所述的系统，其特征在于，所述处理设备包括：第一服务器和与所述第一服务器相连的第二服务器，所述第二服务器与多个缓存服务器相连；

所述第二服务器，用于获取所述第一服务器发送的目标数据，利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；并根据预先存储的发送方标识与缓存服务器标识的对应关系，确定与所述目标发送方标识对应的目标缓存服务器，向所述目标缓存服务器发送目标发送方标识；基于共用特征信息和目标私有特征信息确定与目标数据对应的目标属性信息；并将所述目标属性信息发送至第一服务器；

则所述第一服务器，用于在获取目标数据之后，将目标数据发送至第二服务器，并获取所述目标属性信息；

所述目标缓存服务器，用于获取共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；并将所述共用特征信息和目标私有特征信息发送至所述处理设备。

一种异常消息检测方法、装置及系统

技术领域

[0001] 本申请涉及通信技术领域，尤其涉及一种异常消息检测方法、装置及系统。

背景技术

[0002] 伴随着网络技术的不断进步，即时通讯软件不断发展。即时聊天软件中人群复杂经常有欺诈情况出现。如，在阿里旺旺中存在不少欺诈的案例；比如：使用被盗账号冒充账号主人要求好友转钱、冲手机话费等。因此，衍生出一些解决欺诈问题的方法。

[0003] 为了解决欺诈问题，技术人员通常将欺诈聊天消息统称为异常消息。相对于正常消息而言，异常消息的消息量极小，加之正常消息内容多样，所以异常消息会淹没在正常消息中。因此，无法采用传统的分类方法，来区分正常消息和异常消息。

[0004] 目前，检测异常消息的方法主要为敏感词检测方式，即预先设置大量的敏感词。当聊天消息中出现敏感词时，可以认为聊天消息为异常消息。但是，敏感词可以采用变形词或者拼音等方式避过。因此，目前检测异常消息的方式无法准确检测异常消息。

[0005] 因此，现在需要一种新型方式来检测异常消息，以便准确检测异常消息，进而方便技术人员依据异常消息解决欺诈问题。

发明内容

[0006] 本申请提供了一种异常消息检测方法、装置及系统，本申请可以准确检测异常消息。

[0007] 为了实现上述目的，本申请提供以下技术手段：一种异常消息检测方法，包括：

[0008] 基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息；其中，所述目标数据用于表示待检测的目标消息；

[0009] 将所述目标属性信息与预先设置的异常属性模型进行匹配，并确定所述目标数据与所述异常属性模型的匹配度；

[0010] 基于所述匹配度，判断所述目标消息是否为异常消息。

[0011] 优选的，所述基于所述匹配度判断所述目标消息是否为异常消息，包括：

[0012] 对所述匹配度进行归一化处理；

[0013] 若所述匹配度归一化处理后数据值大于预设数据值，则确定所述目标消息为异常消息。

[0014] 优选的，所述预设异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成，每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定。

[0015] 优选的，所述目标数据包括：

[0016] 用于表示所述目标消息发送方的目标发送方账户标识、用于表示所述发送方使用的发送方机器的目标发送方机器标识、用于表示所述目标消息的接收方的目标接收方账户标识和所述目标消息。

[0017] 优选的，所述基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息，包括：

[0018] 利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；

[0019] 获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；其中，所述特征信息集合包括所有发送方共同使用的共用特征信息和多个与发送方标识对应的私有特征信息；

[0020] 利用所述目标私有特征信息和共用特征信息，确定与所述目标数据对应的目标属性信息。

[0021] 优选的，所述目标私有特征信息包括：目标发送方账户利用目标发送方机器发送所有消息第一总数量，向卖家发送所有消息的第一卖家消息数量，向买家发送所有消息的第一买家消息数量，目标发送方机器标识被举报的第一举报次数，目标发送方账户注册时的第一地理位置标识，各个历史消息的发送次数；

[0022] 所述共用特征信息包括：历史欺诈机器标识列表、各个历史举报信息和/或各个接收方身份标识；

[0023] 所述目标属性信息包括：第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度和/或用于表示目标接收方账户为卖家或买家的目标接收方身份标识；

[0024] 则利用所述目标私有特征信息，确定与所述目标数据对应的目标属性信息，包括：

[0025] 将所述目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述第一买家消息数量、所述第一举报次数和所述第一地理位置标识，分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识；

[0026] 对于所述目标属性信息中第二发送次数：在所述目标私有特征信息的各个历史消息判断是否有与所述目标消息一致的历史目标消息；若有，则将所述历史目标消息对应的第一发送次数，赋值于所述第二发送次数；

[0027] 对于所述目标属性信息中的欺诈机器标识：若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识，则将表示欺诈机器的第一标识赋值于所述欺诈机器标识，否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识；

[0028] 对于所述目标属性信息中的相似度：计算所述共用特征信息中各个举报消息与所述目标消息的相似度，将各个相似度中的最高相似度、赋值于所述相似度；

[0029] 对于所述目标属性信息中的目标接收方身份标识：在所述共用特征信息中各个接收方身份标识中，查找与所述目标接收方账户标识对应的身份标识，并将该身份标识赋予所述目标接收方身份标识。

[0030] 优选的，所述异常属性模型包括有多个异常树iTree组成的异常森林iForest，每个iTree包括所述异常数据集合中若干个异常数据的属性信息中的属性值；则所述将所述目标属性信息与预先设置的异常属性模型进行匹配，并确定所述目标数据与所述异常属性模型的匹配度，包括：

[0031] 将所述目标属性信息中的属性值与每个itree中的属性值进行对比，确定所述目标属性信息与每个iTree的匹配度；

[0032] 将所有iTree对应的匹配度的综合值，确定为所述目标数据与所述异常属性模型的匹配度。

[0033] 优选的，所述iTree具有预设最大高度，所述iTree的每层对应一个属性，每个节点对应一个属性值；则所述将所述目标属性信息中的属性值与每个iTree中的属性值进行对比，确定所述目标属性信息与每个iTree的匹配度，包括：

[0034] 从iTree的根节点开始由上至下遍历iTree；

[0035] 获取iTree的一个节点对应属性以及第一属性值，以及所述目标属性信息中相同属性对应的第二属性值；

[0036] 判断第一属性值与第二属性值是否一致；

[0037] 若所述第一属性值与所述第二属性值的误差在预设范围内，则进入下一层节点；重新进入获取iTree的一个节点对应属性以及第一属性值，以及所述目标属性信息中相同属性对应的第二属性值的步骤；

[0038] 若所述第一属性值与所述第二属性值的误差不在预设范围内，则遍历同层的其它节点，若第一属性值与同层其它节点的属性值均不一致，则停止遍历；

[0039] 将当前层与根节点之间的层数，确定为所述目标属性信息与该iTree的匹配度。

[0040] 优选的，所述每个iTree构建过程包括：

[0041] 由根节点开始从上至下构建iTree的每个节点：

[0042] 步骤1：随机选择一个执行属性，并在异常数据集合的剩余异常数据中随机选择一个执行异常数据，并将执行异常数据的属性信息中与执行属性对应执行属性值确定为一个节点；

[0043] 步骤2：在剩余异常数据中排除执行异常数据；

[0044] 步骤3：按执行属性的执行属性值对剩余异常数据进行分类；具体包括：将执行属性的属性值小于执行属性值的异常数据归属于左子树，将执行属性的属性值大于执行属性值的异常数据归属于右子树；

[0045] 重复执行步骤1、步骤2和步骤3，递归的构造左子树和右子树，直到满足以下条件之一则终止：条件1：用于构造iTree的剩余异常数据只有一个异常数据或者多个相同的异常数据；

[0046] 条件2：iTree的高度达到预设高度。

[0047] 优选的，还包括：

[0048] 在确定所述目标数据为异常数据后，更新所述异常数据集合。

[0049] 优选的，所述目标发送方机器标识包括发送方机器的MAC地址和发送方机器的硬盘号码；

[0050] 在所述目标消息为文本情况，所述目标数据中目标消息为文本内容，在所述目标消息为图片时，所述目标数据中的目标消息为图片的MD5值。

[0051] 优选的，所述各个历史消息利用历史消息的MD5值进行存储。

[0052] 一种异常消息检测装置，包括：

[0053] 第一确定属性单元，用于基于历史缓存数据组建的特征信息集合、确定与目标数

据对应的目标属性信息;其中,所述目标数据用于表示待检测的目标消息;

[0054] 匹配单元,用于将所述目标属性信息与预先设置的异常属性模型进行匹配,并确定所述目标数据与所述异常属性模型的匹配度;

[0055] 确定异常单元,用于基于所述匹配度,判断所述目标消息是否为异常消息。

[0056] 优选的,所述确定异常单元具体用于:对所述匹配度进行归一化处理;若所述匹配度归一化处理后数据值大于预设数据值,则确定所述目标消息为异常消息。

[0057] 优选的,所述预设异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成,每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定。

[0058] 优选的,所述目标数据包括:

[0059] 用于表示所述目标消息发送方的目标发送方账户标识、用于表示所述发送方使用的发送方机器的目标发送方机器标识、用于表示所述目标消息的接收方的目标接收方账户标识和所述目标消息。

[0060] 优选的,所述第一确定属性单元,包括:

[0061] 计算单元,用于利用所述目标数据中的目标发送方账户标识和目标发送方机器标识,计算目标发送方标识;

[0062] 获取单元,用于获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息;其中,所述特征信息集合包括所有发送方共同使用的共用特征信息和多个与发送方标识对应的私有特征信息;

[0063] 第二确定属性单元,用于利用所述目标私有特征信息和共用特征信息,确定与所述目标数据对应的目标属性信息。

[0064] 优选的,所述目标私有特征信息包括:目标发送方账户利用目标发送方机器发送所有消息第一总数量,向卖家发送所有消息的第一卖家消息数量,向买家发送所有消息的第一买家消息数量,目标发送方机器标识被举报的第一举报次数,目标发送方账户注册时的第一地理位置标识,各个历史消息的发送次数;

[0065] 所述共用特征信息包括:历史欺诈机器标识列表、各个历史举报信息和/或各个接收方身份标识;

[0066] 所述目标属性信息包括:第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度和/或用于表示目标接收方账户为卖家或买家的目标接收方身份标识;

[0067] 则第二确定属性单元,包括:

[0068] 第一赋值单元,用于将所述目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述第一买家消息数量、所述第一举报次数和所述第一地理位置标识,分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识;

[0069] 第二赋值单元,用于对于所述目标属性信息中第二发送次数:在所述目标私有特征信息的各个历史消息判断是否有与所述目标消息一致的历史目标消息;若有,则将所述历史目标消息对应的第一发送次数,赋值于所述第二发送次数;

[0070] 第三赋值单元，用于对于所述目标属性信息中的欺诈机器标识：若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识，则将表示欺诈机器的第一标识赋值于所述欺诈机器标识，否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识；

[0071] 第四赋值单元，用于对于所述目标属性信息中的相似度：计算所述共用特征信息中各个举报消息与所述目标消息的相似度，将各个相似度中的最高相似度、赋值于所述相似度；

[0072] 第五赋值单元，用于对于所述目标属性信息中的目标接收方身份标识：在所述共用特征信息中各个接收方身份标识中，查找与所述目标接收方账户标识对应的身份标识，并将该身份标识赋予所述目标接收方身份标识。

[0073] 优选的，所述异常属性模型包括有多个异常树iTree组成的异常森林iForest，每个iTree包括所述异常数据集合中若干个异常数据的属性信息中的属性值；则匹配单元，包括：

[0074] 对比单元，用于将所述目标属性信息中的属性值与每个itree中的属性值进行对比，确定所述目标属性信息与每个iTree的匹配度；

[0075] 确定匹配度单元，用于将所有iTree对应的匹配度的综合值，确定为所述目标数据与所述异常属性模型的匹配度。

[0076] 优选的，所述iTree具有预设最大高度，所述iTree的每层对应一个属性，每个节点对应一个属性值；

[0077] 则对比单元，具体用于从iTree的根节点开始由上至下遍历iTree；获取iTree的一个节点对应属性以及第一属性值，以及所述目标属性信息中相同属性对应的第二属性值；判断第一属性值与第二属性值是否一致；若所述第一属性值与所述第二属性值的误差在预设范围内，则进入下一层节点；重新进入获取iTree的一个节点对应属性以及第一属性值，以及所述目标属性信息中相同属性对应的第二属性值的步骤；若所述第一属性值与所述第二属性值的误差不在预设范围内，则遍历同层的其它节点，若第一属性值与同层其它节点的属性值均不一致，则停止遍历；将当前层与根节点之间的层数，确定为所述目标属性信息与该iTree的匹配度。

[0078] 优选的，还包括：

[0079] 更新单元，用于在确定所述目标数据为异常数据后，更新所述异常数据集合。

[0080] 一种异常消息检测系统，包括：处理设备和与所述处理设备相连的多个缓存服务器；

[0081] 其中，所述多个缓存服务器，用于存储基于历史缓存数据组建的特征信息集合；

[0082] 所述处理设备，用于从所述多个缓存服务器中确定与目标数据对应的目标属性信息；其中，所述目标数据用于表示待检测的目标消息；将所述目标属性信息与预先设置的异常属性模型进行匹配，并确定所述目标数据与所述异常属性模型的匹配度；其中，所述预设异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成，每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定；若对所述匹配度进行归一化的数据值大于预设数据值，则确定所述目标消息为异常消息。

[0083] 优选的，所述处理设备包括：第一服务器；每个缓存服务器中存储有共用特征信息

以及与发送方标识对应的私有特征信息；

[0084] 则所述处理设备基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息，具体包括：

[0085] 所述第一服务器，具体用于利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；并根据预先存储的发送方标识与缓存服务器标识的对应关系，确定与所述目标发送方标识对应的目标缓存服务器，向所述目标缓存服务器发送目标发送方标识；基于共用特征信息和目标私有特征信息确定与目标数据对应的目标属性信息；

[0086] 所述目标缓存服务器，用于获取共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；并将所述共用特征信息和目标私有特征信息发送至所述处理设备。

[0087] 优选的，所述处理设备包括：第一服务器和与所述第一服务器相连的第二服务器，所述第二服务器与多个缓存服务器相连；每个缓存服务器中存储有共用特征信息以及与发送方标识对应的私有特征信息；

[0088] 则所述处理设备基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息，具体包括：

[0089] 所述第二服务器，用于获取所述第一服务器发送的目标数据，利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；并根据预先存储的发送方标识与缓存服务器标识的对应关系，确定与所述目标发送方标识对应的目标缓存服务器，向所述目标缓存服务器发送目标发送方标识；基于共用特征信息和目标私有特征信息确定与目标数据对应的目标属性信息；并将所述目标属性信息发送至第一服务器；

[0090] 则所述第一服务器，用于在获取目标数据之后，将目标数据发送至第二服务器，并获取所述目标属性信息；

[0091] 所述目标缓存服务器，用于获取共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；并将所述共用特征信息和目标私有特征信息发送至所述处理设备。

[0092] 从以上技术手段可以看出本申请具有以下有益效果：

[0093] 本申请提供的一种异常消息检测方法，本申请基于历史缓存数据中提取出的特征信息集合确定异常数据的属性信息，并利用异常数据的属性信息构建异常属性模型。然后，基于历史缓存数据中提取出的特征信息集合确定目标数据的目标属性信息，计算目标属性信息与异常属性模型的匹配度。若对匹配度归一化后的数据值大于预设数据值，则确定目标数据为异常数据，目标消息为异常消息。

[0094] 本申请中，由于特征信息集合较难发生改变，所以从特征信息集合中提取得到的属性信息也不会轻易改变。因此，本申请提供属性信息对比方式，可以准确确定异常消息，进而方便技术人员依据异常消息解决欺诈问题。

附图说明

[0095] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术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申请的一些实施例，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。

[0096] 图1为本申请提供的一种异常消息检测系统的结构示意图；

[0097] 图2为本申请提供的一种异常消息检测方法的流程图；

[0098] 图3a-3b为本申请提供的又一种iTree的示意图；

[0099] 图4为本申请提供的又一种异常消息检测方法的流程图；

[0100] 图5为本申请提供的又一种异常消息检测方法的流程图；

[0101] 图6为本申请提供的又一种异常消息检测方法的流程图；

[0102] 图7为本申请提供的又一种异常消息检测方法的流程图；

[0103] 图8为本申请提供的又一种异常消息检测方法的流程图；

[0104] 图9为本申请提供的一种异常消息检测装置的结构示意图；

[0105] 图10为本申请提供的又一种异常消息检测装置的结构示意图；

[0106] 图11为本申请提供的又一种异常消息检测装置的结构示意图；

[0107] 图12为本申请提供的又一种异常消息检测系统的结构示意图；

[0108] 图13为本申请提供的又一种异常消息检测系统的结构示意图。

具体实施方式

[0109] 下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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他实施例，都属于本申请保护的范围。

[0110] 为了使本领域技术人员清楚了解本申请中技术术语，下面对技术术语进行解释说明：

[0111] 消息：在即时通讯软件中通讯双方相互发送的内容。

[0112] 正常消息：在消息内容中不具有欺诈内容的消息。

[0113] 异常消息：在消息内容中具有欺诈内容的消息。

[0114] 数据：在本申请用于唯一表示一个消息，数据可以包括消息的发送方账户标识、发送方机器标识、接收方账户标识和消息自身，此外还可以包括其它内容。

[0115] 历史缓存数据：本申请缓存用户之前发送的消息所对应的数据，将所有缓存数据的统称为历史缓存数据。

[0116] 异常数据集合：历史缓存数据中所有异常数据的集合。

[0117] 特征信息：在本申请中表示历史缓存数据中一类数据集合所具有的各个特征，特征信息包括私有特征信息和共用特征信息。私有特征信息可以包括(1)第一总数量，即表示发送方账户通过发送方机器标识发送所有消息的总数量；(2)第一卖家消息数量，即在第一总数量中向卖家发送的消息数量。(3)第一买家消息数量，即在第一总数量中向买家发送的消息数量。(4)第一举报次数，即发送方机器标识被举报的第一举报次数。(5)第一地理位置标识，即每个发送方账户注册时的地理位置标识。(6)各个历史消息的发送次数。共用特征信息可以包括(1)历史欺诈机器标识列表；(2)各个历史举报信息；(3)各个接收方身份标识。

[0118] 匹配度：一般指某物质与另一物质比较的相似度的数据化衡量。

[0119] 归一化：为了数据处理方便所提出来的、将数据映射到0～1范围之内的过程。

[0120] 异常属性模型:在本申请含义是为了判断目标数据是否异常而构建的软件模型。例如,异常森林模型(Isolation Forest,iForest)。

[0121] iTree:一种随机二叉树,每个节点有两个内部节点作为子女或者只有叶子节点,不存在只有一个内部节点的情况。

[0122] 下面介绍本申请的详细执行过程:

[0123] 本申请的发明人在研究过程中发现:虽然消息中的关键词较容易被改变,但是与消息有关的一部分内容是不容易发生改变。例如,用于表示消息发送方的发送方账户标识、用于表示发送方使用的发送方机器的发送方机器标识以及用于表示消息接收方的接收方账户标识等等。所以,可以利用这些不容易发生改变的内容来表示一个消息;并且,为了区分各个消息,可以增加消息本身。

[0124] 即,一个消息在本申请中采用发送方账户标识、发送方机器标识、接收方账户标识和消息来表示。为了便于称呼防止混淆,后续将发送方账户标识、发送方机器标识、接收方账户标识和消息的组合称为数据。当然,为了完善一个消息对应的数据,数据还可以包括:发送时间和接收方机器标识。

[0125] 其中,发送方机器标识包括发送方的MAC地址和发送方的硬盘号码。当消息为文本时,数据中的消息即为文本内容;当消息为图片时,数据中的消息内容为图片的MD5码。

[0126] 为了确定一个消息是否异常,可以确定该消息对应的数据是否异常。因此,本申请发明人设想获取大量的历史缓存数据,并从大量的历史缓存数据中提取出每个异常数据的属性信息。为了确定每个异常数据的属性信息,需要基于大量历史缓存数据确定特征信息集合。为此,本申请可以获取大量的历史缓存数据。

[0127] 历史缓存数据中有很多发送方发送的数据,不同发送方发送对应的特征信息不尽相同。因此,本申请按发送方作为主键对大量的历史缓存数据进行分类,并确定出每类数据集合的特征信息。可以理解的是,不同的发送方对应的发送方账户标识不同,理论上可以采用发送方账户标识来区分历史缓存数据。

[0128] 但是,由于相同发送方账户标识可以在不同的发送方机器上登录,在不同的发送方机器上登录时可能对应的用户不同。例如,发送方账户的正常用户使用发送方机器A发送数据,而发送方账户被盗用后,会使用发送方机器B登录。因此,为了精准区分不同用户,可以采用发送方账户标识和发送方机器标识一并作为主键,来对历史缓存数据进行分类。

[0129] 在确定每类数据集合的特征信息时,需要使用异常数据集合。因此,可以在历史缓存数据中确定出异常数据结合。确定异常数据结合的过程可以为:从历史缓存数据中确定出被用户举报的所有举报数据,以及通过人工筛选方式筛选出所有欺诈数据,将所有举报数据和所有欺诈数据确定为异常数据集合。

[0130] 可以理解的是,本申请对每类数据集合的处理过程均是一致的。因此,下面以一类数据集合为例,对一类数据集合确定特征信息的过程进行详细描述。由前述定义部分可知,特征信息可以包括9个特征,下面一一对每个特征的确定过程进行描述:

[0131] (1)针对第一总数量:将该类数据集合中数据个数确定为第一总数量。

[0132] 由于每类数据集合均是按照发送方账户标识和发送方机器标识为主键进行分类的,所以,每类数据集合中每个数据均是发送方账户通过发送方机器标识发送的数据。因此,统计该类数据集合中数据个数,便可以确定出第一总数量。

[0133] (2)针对第一卖家消息数量:将该类数据集合中接收方身份标识为卖家的数据个数,确定为第一卖家消息数量。

[0134] 在有买卖交易的即时通讯软件中(例如,阿里旺旺)有卖家和买家角色,因此,在有买卖交易的即时通讯中可以具有该特征。在其它即时通讯软件并没有卖家和买家角色,因此,可以不采用该特征。

[0135] 在每个数据中具有接收方账户标识,并且,本申请预先存储有各个账户标识与身份标识的对应关系。因此,可以通过接收方账户标识确定出接收方的身份标识(卖家或买家)。然后在该类数据集合中统计出接收方身份标识为卖家的数据个数,并将该个数作为第一卖家消息数量。

[0136] (3)针对第一买家消息数量:将该类数据集合中接收方身份标识为卖家的数据个数,确定为第一卖家消息数量。

[0137] (4)针对第一举报次数:将该类数据集合中发送方机器标识在举报数据中出现次数,确定为第一举报次数。。

[0138] 在异常数据集合包括举报数据,每个举报数据均包含发送方机器标识。然后,可以统计出该类数据集合中发送方机器标识在所有举报数据中的出现次数,并将出现次数确定为第一举报次数。

[0139] (5)针对第一地理位置标识:依据发送方账户标识与地理位置标识的对应关系,确定该类数据集合中发送账户标识对应的第一地址位置标识。

[0140] 本申请包含每个发送方账户标识在注册时的地理位置标识,例如,杭州采用1、南京采用2、北京采用3等等。因此,可以通过预先存储的发送方账户标识与地理位置标识的对应关系,可以确定出该类数据集合中发送账户标识对应的第一地址位置标识。

[0141] (6)针对各个历史消息的发送次数:统计该类数据集合中不同的历史消息,以及每个历史消息的发送次数。

[0142] 在该类数据集合中包括多个数据,每个数据中均有消息本身。可以理解的是,发送方可以发送相同的历史消息。因此,可以在该类数据集合中统计出发送方采用发送方机器发送的各个历史消息,以及各个历史消息的发送次数。

[0143] 为了减少存储空间,各个历史消息可以采用MD5值进行表示。

[0144] (7)针对历史欺诈机器标识列表:将异常数据集合中各个数据的发送方机器标识的集合,确定为历史欺诈机器标识列表。

[0145] 在异常数据集合每个异常数据确定出发送方机器标识,这些发送方机器标识经常用于发送欺诈消息,因此,可以将其作为欺诈机器标识。然后,将所有欺诈机器标识组成历史欺诈机器标识列表,也即黑名单。

[0146] (8)针对各个历史举报信息:将异常数据集合中提取出来的历史消息,确定为历史举报信息。

[0147] 从异常数据集合中的各个举报数据中确定出各个历史举报信息。可以理解的是,若一个数据中消息与历史举报信息相同,则很大程度上表示该数据为异常数据。

[0148] (9)针对各个接收方身份标识:在机器标识与身份标识的对应关系中,确定与各个接收方机器标识对应的接收方身份标识。在上述9个特征中前6个特征是根据发送方账户标识和发送方机器标识不同而不同的特征,因此,前6个特征可以作为私有特征信息。后面三

个特征，是所有发送方共同使用的特征信息，因此，可以将后面三个特征作为共用特征信息。当然，本申请仅提供了以上9个特征，还可以采用其它特征，在此不再一一列举。

[0149] 上述确定每类数据集合的特征信息的过程可以由后续执行异常消息检测方法的核心设备来执行，或者由核心设备之外的其它设备来执行，本申请并不限定其上述提取特征信息的执行设备。

[0150] 按上述过程可以确定每类数据的特征信息，由于后续需要基于每类数据的特征信息确定异常数据的属性信息，因此需要对每类数据的特征信息进行存储。由于基于历史数据发送方账户数量巨大，因此，经过上述过程确定特征信息的数量巨大。为此，本申请提供了分布式缓存方案。即，提供多个缓存服务器，在每个缓存服务器上存储一部分特征信息。

[0151] 为了方便后续使用，可以计算每类数据集合的发送方标识，由于每类数据由发送方标识和发送方机器标识进行分类，因此，可以利用发送方标识和发送方机器标识一并确定发送标识。然后，确定存储该类数据集合对应特征信息的缓存服务器标识。最后，构建发送方标识与缓存服务器标识的对应关系。

[0152] 为了使本领域技术人员清楚了解本申请的应用场景，首先介绍一种异常消息检测系统。参见图1，异常消息检测系统包括：处理设备100与所述处理设备100相连的多个缓存服务器200。

[0153] 为了防止多个缓存服务器崩溃或者出现故障时丢失特征信息，因此，本申请提供的异常消息检测系统还可以包括第二级缓存服务器，作为特征信息的冗余存储。

[0154] 其中，处理设备100上发送方标识与缓存服务器标识的对应关系，存储有每个缓存服务器200上存储有共用特征信息以及多个与发送方标识对应的私有特征信息。

[0155] 在基于历史缓存数据确定每类数据集合的特征信息后，将各个特征信息集合称为特征信息集合。然后可以基于特征信息集合确定异常数据的属性信息。属性信息可以包括以下9个属性，其中一些属性与特征一致，一些属性与特征不一致。下面以一个执行异常数据为例，对确定执行异常数据的属性信息的过程进行详细论述。如图2所示，确定执行异常数据的属性信息的过程具体包括：

[0156] 步骤S201：利用所述执行数据中的执行发送方账户标识和执行发送方机器标识，计算执行发送方标识。

[0157] 执行异常数据包括执行发送方账户标识、执行发送方机器标识、执行接收方账户标识和执行消息。为了确定异常数据的属性信息，需要在缓存服务器中获取与执行发送方账户标识和执行发送方机器标识对应的特征信息。

[0158] 为此，首先利用执行数据中的执行发送方账户标识和执行发送方机器标识，计算执行发送方标识。

[0159] 步骤S202：获取所述特征信息集合中的共用特征信息以及与所述执行发送方标识对应的执行私有特征信息；其中，所述特征信息集合包括所有发送方共同使用的共用特征信息和多个与发送方标识对应的私有特征信息。

[0160] 在处理设备预先存储的发送方标识与缓存服务器标识的对应关系中，确定与执行发送方标识对应的执行缓存服务器标识。继而在执行缓存服务器标识对应的执行缓存服务器中，获取与执行发送方标识对应的执行特征信息以及共用特征信息。

[0161] 步骤S203：利用所述执行私有特征信息和共用特征信息，确定与所述执行数据对

应的执行属性信息。

[0162] 执行私有特征信息包括:第一总数量、第一卖家消息数量、第一买家消息数量、第一举报次数、第一地理位置标识、各个历史消息的发送次数;共用特征信息包括:历史欺诈机器标识列表、各个历史举报信息和/或各个接收方身份标识。

[0163] 执行属性信息包括:第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、执行消息的第二发送次数、用于标识执行发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、执行消息与举报信息的相似度和/或用于表示执行接收方账户为卖家或买家的执行接收方身份标识。

[0164] 下面详细说明本步骤的具体执行过程:

[0165] (1)对于第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数和第二地理位置标识这5个属性,其含义与特征信息中的含义相同,因此,可以直接进行赋值。

[0166] 即,将执行私有特征信息中的第一总数量、第一卖家消息数量、第一买家消息数量、第一举报次数和第一地理位置标识,分别赋值于执行属性信息中的第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数和第二地理位置标识;

[0167] (2)对于执行属性信息中第二发送次数:

[0168] 在执行私有特征信息的各个历史消息判断是否有与执行消息一致的历史执行消息;若有,则将历史执行消息对应的第一发送次数,赋值于第二发送次数。若无,则确定第二发送次数为零。

[0169] (3)对于执行属性信息中的欺诈机器标识:

[0170] 判断共用特征信息中的历史欺诈机器标识列表是否包含执行发送方机器标识,若是,则将表示欺诈机器的第一标识赋值于欺诈机器标识;否则将表示非欺诈机器的第二标识赋值于欺诈机器标识。

[0171] (4)对于执行属性信息中的相似度:

[0172] 计算共用特征信息中各个举报消息与执行消息的相似度,将各个相似度中的最高相似度、赋值于相似度。

[0173] (5)对于执行属性信息中的执行接收方身份标识:在共用特征信息中各个接收方身份标识中,查找与执行接收方账户标识对应的身份标识,并将该身份标识赋予执行接收方身份标识。

[0174] 按(1)、(2)、(3)、(4)和(5)的过程,确定属性信息中各个属性的属性值。

[0175] 在按图2所示的过程,确定每个异常数据的属性信息后,可以将各个异常数据的属性信息组成异常属性模型。以便后续判断待检测消息是否为异常消息。可以理解的是,异常属性模型可以有多种形式。异常属性模型可以包括多个异常树(Isolation Tree,iTree)组成的异常森林(Isolation Forest,iForest)。iTree是一种随机二叉树。如图3所示,图3a为iTree,图3b因为根节点只有一个内部节点,所以不是iTree。

[0176] 由于iForest由多个iTree组成,因此确定每个iTree之后,便可以获得iForest。由于每个iTree的构建过程类似,所以本实施例后续着重描述一个iTree的构建过程。

[0177] 由于一个iTree由若干个异常数据的属性值组成,因此,构建一个iTree的过程为在异常数据集合中选择若干个异常数据的属性值的过程。如图4所示,构建iTree的过程具体包括以下步骤:

[0178] 由根节点开始从上至下构建iTree的每个节点：

[0179] 步骤1:随机选择一个执行属性,并在异常数据集合的剩余异常数据中随机选择一个执行异常数据,并将执行异常数据的属性信息中与执行属性对应执行属性值确定为一个节点。

[0180] 为了清楚表述iTree的处理过程,下面以一个实例来对每个步骤进行详细说明。假设异常数据集合中剩余异常数据为异常数据1、异常数据2、异常数据3和异常数据4四个异常数据,每个异常数据包括A、B、C、D和E共5个属性。

[0181] 详细属性值见表1

[0182] 表1

[0183]

项目	属性 A	属性 B	属性 C	属性 D	属性 E
异常数据 1	2	0.1	4	3	0.1

[0184]

异常数据 2	3	0.2	3	2	0.3
异常数据 3	1	0.5	6	1	0.7
异常数据 4	2	0.4	3	5	0.4

[0185] 首先确定根节点,则假设随机选择的执行属性为属性A,并在剩余异常数据中随机选择异常数据4。则将异常数据4的属性A对应的属性值2确定为根节点。

[0186] 步骤2:在剩余异常数据中排除执行异常数据。

[0187] 由于异常数据4已经被放置于根节点,因此,可以在剩余异常数据中排除异常数据4。

[0188] 步骤3:按执行属性的执行属性值对剩余异常数据进行分类;具体包括:将执行属性的属性值小于执行属性值的异常数据归属于左子树,将执行属性的属性值大于执行属性值的异常数据归属于右子树。

[0189] 在排除异常数据4之后,剩余的为异常数据1、异常数据2和异常数据3,为了进一步构建下一层节点,并便于后续的比较过程。可以对异常数据1、异常数据2和异常数据3进行分类。

[0190] 根节点对应的属性值2,异常数据1属性A的属性值为2,与根节点的属性值2相等,因此将异常数据1归属于右子树。异常数据2属性A的属性值为3,大于与根节点的属性值2,因此将异常数据2归属于右子树。异常数据3属性A的属性值为1,小于与根节点的属性值2,因此将异常数据3归属于左子树。

[0191] 步骤4:重复执行步骤1、步骤2和步骤3,递归的构造左子树和右子树,直到满足以下条件之一则终止:条件1:用于构造iTree的剩余异常数据只有一个异常数据或者多个相同的异常数据;条件2:iTree的高度达到预设高度。

[0192] 然后,在左子树对应的数据中重复执行步骤1、步骤2和步骤3构建根节点下的子节点,以及,在左子树对应的数据中重复执行步骤1、步骤2和步骤3构建根节点下的子节点。直到剩余异常数据满足条件1:只有一个异常数据或者多个相同的异常数据;或者条件2:iTree的高度达到预设高度。

[0193] 可以理解的是,在剩余异常数据较多时,一般在满足条件2时构建iTree的过程结

束。随着一个又一个iTree的构建，由于不断排除异常数据，所以剩余异常数据越来越少，因此最后一个iTree的终止条件应该为条件1。

[0194] 构建iTree数量的多少与iTree的预设高度和异常数据集合中异常数据数量有关。即，在异常数据集合中异常数据数量一定时，iTree的预设高度越大，构建得到的iTree数量越少，iTree的预设高度越小，构建得到的iTree数量越多。

[0195] 但是，本申请发明人在实际研究过程中发现，iTree的预设高度并不是越大越好，也不是越少越好。经过实验确定在实际应用中，iTree的预设高度可以为7，即iTree最高为7层。这样的实验效果较高，即可以较为准确地确定待检测消息是否为异常消息。

[0196] 以上基于历史缓存数据确定特征信息集合，基于特征信息集合确定异常数据的属性信息，并利用异常数据的属性信息构建异常属性模型的过程，均为本申请的预先准备过程。在上述准备过程中执行完毕之后，可以确定一个待检测的消息是否为异常消息，由于本申请对每个消息的执行过程均是一致的。因此，本申请仅以目标消息为例，对确定目标消息的执行过程进行详细描述。

[0197] 本申请提供了一种异常消息检测方法，应用于图1所示的处理设备。如图5所示，具体包括以下步骤：

[0198] 步骤S501：基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息；其中，所述目标数据用于表示待检测的目标消息。

[0199] 如图6所示，本步骤具体包括以下步骤：

[0200] 步骤S601：利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识。

[0201] 步骤S602：获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；其中，所述特征信息集合包括所有发送方共同使用的共用特征信息和多个与发送方标识对应的私有特征信息。

[0202] 步骤S603：利用所述目标私有特征信息和共用特征信息，确定与所述目标数据对应的目标属性信息。

[0203] 目标私有特征信息包括：目标发送方账户利用目标发送方机器发送所有消息第一总数量，向卖家发送所有消息的第一卖家消息数量，向买家发送所有消息的第一买家消息数量，目标发送方机器标识被举报的第一举报次数，目标发送方账户注册时的第一地理位置标识，各个历史消息的发送次数。

[0204] 所述共用特征信息包括：历史欺诈机器标识列表、各个历史举报信息和/或各个接收方身份标识。

[0205] 所述目标属性信息包括：第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度和/或用于表示目标接收方账户为卖家或买家的目标接收方身份标识。

[0206] 那么步骤S603的具体执行过程如下：

[0207] (1)对于第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数和第二地理位置标识这5个属性，其含义与特征信息中的含义相同，因此，可以直接进行赋值。

[0208] 即，将所述目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述

第一买家消息数量、所述第一举报次数和所述第一地理位置标识，分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识。

[0209] (2)对于所述目标属性信息中第二发送次数：

[0210] 在所述目标私有特征信息的各个历史消息判断是否有与所述目标消息一致的历史目标消息；若有，则将所述历史目标消息对应的第一发送次数，赋值于所述第二发送次数。

[0211] (3)对于所述目标属性信息中的欺诈机器标识：若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识，则将表示欺诈机器的第一标识赋值于所述欺诈机器标识，否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识。

[0212] (4)对于所述目标属性信息中的相似度：计算所述共用特征信息中各个举报消息与所述目标消息的相似度，将各个相似度中的最高相似度、赋值于所述相似度。

[0213] (5)对于所述目标属性信息中的目标接收方身份标识：在所述共用特征信息中各个接收方身份标识中，查找与所述目标接收方账户标识对应的身份标识，并将该身份标识赋予所述目标接收方身份标识。

[0214] 接着返回图5，进入步骤S502：将所述目标属性信息与预先设置的异常属性模型进行匹配，并确定所述目标数据与所述异常属性模型的匹配度。其中，所述预设异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成，每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定。

[0215] 步骤S503：基于所述匹配度，判断所述目标消息是否为异常消息。

[0216] 在确定目标数据的目标属性信息后，将目标属性信息与异常属性模型进行匹配，从而确定目标数据与异常属性模型的匹配度。由于异常属性模型均由异常数据的特征信息组成。因此若匹配度越高，则表示目标数据越趋向于异常数据，若匹配度越低，则表示目标数据越趋向于正常数据。

[0217] 由于异常数据模型中的参数不同，会对匹配度有不同的影响。因此，可以对匹配对进行归一化处理，并以便确定出适应于不同参数下利用异常数据模型确定异常数据的过程。若对所述匹配度进行归一化的数据值大于预设数据值，则确定所述目标消息为异常消息。否则确定所述目标消息为正常消息。其中，预设数据值为区分正常数据和异常数据的界限。

[0218] 由上述实施例可以看出本申请具有以下有益效果：

[0219] 本申请提供的一种异常消息检测方法，本申请基于历史缓存数据中提取出的特征信息集合确定异常数据的属性信息，并利用异常数据的属性信息构建异常属性模型。然后，基于历史缓存数据中提取出的特征信息集合确定目标数据的目标属性信息，计算目标属性信息与异常属性模型的匹配度。若对匹配度归一化后的数据值大于预设数据值，则确定目标数据为异常数据，目标消息为异常消息。

[0220] 本申请中，由于特征信息集合较难发生改变，所以从特征信息集合中提取得到的属性信息也不会轻易改变。因此，本申请提供属性信息对比方式，可以准确确定异常消息，进而方便技术人员依据异常消息解决欺诈问题。

[0221] 下面详细介绍图5所示的步骤S502中的匹配过程：

[0222] 以所述异常属性模型包括有多个iTree(Isolation Tree)组成的iForest(Isolation Forest)模型为例,每个iTree包括若干个异常数据的属性信息中的属性值。如图7所示,本步骤包括:

[0223] 步骤S701:将所述目标属性信息中的属性值与每个itree中的属性值进行对比,确定所述目标属性信息与每个iTree的匹配度。本步骤将在后续实施例中进行详细描述。

[0224] 步骤S702:将所有iTree对应的匹配度的综合值,确定为所述目标数据与所述异常属性模型的匹配度。

[0225] 在确定目标属性信息与各个iTree的匹配度之后,可以将所有iTree对应的匹配度综合值,确定目标数据与所述异常属性模型的匹配度。

[0226] 所有iTree对应匹配度的综合值,可以为所有匹配度的综合值。或者,设定各个匹配度的权重,将各个匹配度与对应的权重的乘积的和值,确定为综合值。当然还可以采用其它综合值的计算方式,再次做限定。

[0227] 下面对步骤S701确定所述目标属性信息与每个iTree的匹配度的过程,进行详细描述。如图8所示,具体包括以下步骤:

[0228] 从iTree的根节点开始由上至下遍历iTree,可以理解的是,iTree有多个层。由上至下遍历iTree,以判断目标属性信息中的属性值与iTree中属性值的匹配程度。

[0229] 步骤S801:获取iTree的一个节点对应属性以及第一属性值,以及所述目标属性信息中相同属性对应的第二属性值。

[0230] 获取iTree的一个节点对应属性的属性值,并在目标属性信息中获取相同属性的属性值,以便判断该节点与目标属性信息是否一致。

[0231] 步骤S802:判断第一属性值与第二属性值误差在预设范围内。若是,则进入步骤S803,否则进入步骤S804。

[0232] 预先设定一个预设范围,并判断第一数据值与第二属性值的误差是否在预设范围内。若第一属性值与第二属性值的误差在预设范围内,则确定第一属性值与第二属性值一致;否则确定第一属性值与第二属性值不一致。

[0233] 步骤S803:若所述第一属性值与所述第二属性值的误差在预设范围内,则进入下一层节点;反馈步骤S801。

[0234] 若所述第一属性值与所述第二属性值的误差在预设范围内,则确定第一属性值与第二属性值一致,因此,可以确定该节点与目标属性信息中的节点一致。然后进入下一层节点,并在下一层节点中选择一个节点重新执行步骤S801和步骤S802。这样做的目的在于,确定目标属性信息可以与iTree上属性值的匹配程度。

[0235] 步骤S804:若所述第一属性值与所述第二属性值的误差不在预设范围内,则判断是否有同层其它节点;若是,则进入步骤S805;否则进入步骤S806。

[0236] iTree一层对应一个属性,一层节点可以有多个节点。若其中iTree的一层节点中一个节点不一致,则可以判断目标属性信息中的第一属性值是否与该层其它节点的属性值一致。

[0237] 步骤S805:确定同层其它节点,返回步骤S801。

[0238] 若同层还有其它未匹配过的节点,则重新确定同层其它节点。然后进入步骤S801,以便判断目标属性信息中的第一属性值与同层其它节点的属性值是否一致。

[0239] 步骤S806:若第一属性值与同层其它节点的属性值均不一致,则停止遍历。

[0240] 若目标属性信息中第一属性值与同层其它节点的属性值均不一致,则说明第一属性值无法在iTree中找到匹配的节点,因此,可以结束对iTree的遍历。

[0241] 步骤S807:将当前层与根节点之间的层数,确定为所述目标属性信息与该iTree的匹配度。

[0242] 若停止遍历的当前层与根节点之间的层数,确定为目标属性信息与iTree的匹配度。例如,匹配到第三层时停止遍历,则将当前层3与根节点之间的层数2确定目标属性信息与该iTree的匹配度。

[0243] 按图8所示的过程,可以确定目标属性信息与各个iTree的匹配度。

[0244] 可以理解的是,在图5所示的实施例之后,不论目标数据是正常数据还是异常数据,均可以将目标数据更新至历史缓存数据中,并更新与目标发送方账户标识对应的私有特征信息,以便进行后续的待检测的消息的判断过程。

[0245] 可以理解的是,在确定目标数据确定为异常数据后,将目标数据添加至异常数据集合中。然后,可以根据更新后的异常数据集合重新确定异常属性模型,以便实时更新异常属性模型。

[0246] 如图9所示,本申请还提供了一种异常消息检测装置,包括:

[0247] 第一确定属性单元91,用于基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息;其中,所述目标数据用于表示待检测的目标消息。

[0248] 所述目标数据包括:用于表示所述目标消息发送方的目标发送方账户标识、用于表示所述发送方使用的发送方机器的目标发送方机器标识、用于表示所述目标消息的接收方的目标接收方账户标识和所述目标消息。所述目标发送方机器标识包括发送方机器的MAC地址和发送方机器的硬盘号码;在所述目标消息为文本情况,所述目标数据中目标消息为文本内容,在所述目标消息为图片时,所述目标数据中的目标消息为图片的MD5值。

[0249] 匹配单元92,用于将所述目标属性信息与预先设置的异常属性模型进行匹配,并确定所述目标数据与所述异常属性模型的匹配度;其中,所述预设异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成,每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定。

[0250] 判断单元93,用于基于所述匹配度,判断所述目标消息是否为异常消息。

[0251] 判断单元93具体可以用于对所述匹配度进行归一化处理;若所述匹配度归一化处理后数据值大于预设数据值,则确定所述目标消息为异常消息。

[0252] 此外,还可以包括更新单元94,用于在确定所述目标数据为异常数据后,更新所述异常数据集合。

[0253] 其中,如图10所示,所述第一确定属性单元91,包括:

[0254] 计算单元101,用于利用所述目标数据中的目标发送方账户标识和目标发送方机器标识,计算目标发送方标识。

[0255] 获取单元102,用于获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息;其中,所述特征信息集合包括所有发送方共同使用的共用特征信息和多个与发送方标识对应的私有特征信息。

[0256] 第二确定属性单元103,用于利用所述目标私有特征信息和共用特征信息,确定与

所述目标数据对应的目标属性信息。

[0257] 其中，所述目标私有特征信息包括：目标发送方账户利用目标发送方机器发送所有消息第一总数量，向卖家发送所有消息的第一卖家消息数量，向买家发送所有消息的第一买家消息数量，目标发送方机器标识被举报的第一举报次数，目标发送方账户注册时的第一地理位置标识，各个历史消息的发送次数。

[0258] 所述共用特征信息包括：历史欺诈机器标识列表、各个历史举报信息和/或各个接收方身份标识。

[0259] 所述目标属性信息包括：第二总数量、第二卖家消息数量、第二买家消息数量、第二举报次数、第二地理位置标识、目标消息的第二发送次数、用于标识目标发送方机器标识是否处于历史欺诈机器标识列表的欺诈机器标识、目标消息与举报信息的相似度和/或用于表示目标接收方账户为卖家或买家的目标接收方身份标识；

[0260] 如图10所示，则第二确定属性单元103，包括：

[0261] 第一赋值单元1031，用于将所述目标私有特征信息中的所述第一总数量、所述第一卖家消息数量、所述第一买家消息数量、所述第一举报次数和所述第一地理位置标识，分别赋值于所述目标属性信息中的所述第二总数量、所述第二卖家消息数量、所述第二买家消息数量、所述第二举报次数和所述第二地理位置标识。

[0262] 第二赋值单元1032，用于对于所述目标属性信息中第二发送次数：在所述目标私有特征信息的各个历史消息判断是否有与所述目标消息一致的历史目标消息；若有，则将所述历史目标消息对应的第一发送次数，赋值于所述第二发送次数。其中，各个历史消息利用历史消息的MD5值进行存储。

[0263] 第三赋值单元1033，用于对于所述目标属性信息中的欺诈机器标识：若所述共用特征信息中的历史欺诈机器标识列表包含所述目标发送方机器标识，则将表示欺诈机器的第一标识赋值于所述欺诈机器标识，否则将表示非欺诈机器的第二标识赋值于所述欺诈机器标识。

[0264] 第四赋值单元1034，用于对于所述目标属性信息中的相似度：计算所述共用特征信息中各个举报消息与所述目标消息的相似度，将各个相似度中的最高相似度、赋值于所述相似度。

[0265] 第五赋值单元1035，用于对于所述目标属性信息中的目标接收方身份标识：在所述共用特征信息中各个接收方身份标识中，查找与所述目标接收方账户标识对应的身份标识，并将该身份标识赋予所述目标接收方身份标识。

[0266] 其中，所述异常属性模型包括有多个异常树iTree组成的异常森林iForest，每个iTree包括所述异常数据集合中若干个异常数据的属性信息中的属性值。如图11所示，则匹配单元92，包括：

[0267] 对比单元111，用于将所述目标属性信息中的属性值与每个itree中的属性值进行对比，确定所述目标属性信息与每个iTree的匹配度。

[0268] 其中，所述iTree具有预设最大高度，所述iTree的每层对应一个属性，每个节点对应一个属性值；

[0269] 则对比单元，具体用于从iTree的根节点开始由上至下遍历iTree；获取iTree的一个节点对应属性以及第一属性值，以及所述目标属性信息中相同属性对应的第二属性值；

判断第一属性值与第二属性值是否一致;若所述第一属性值与所述第二属性值的误差在预设范围内,则进入下一层节点;重新进入获取iTree的一个节点对应属性以及第一属性值,以及所述目标属性信息中相同属性对应的第二属性值的步骤;若所述第一属性值与所述第二属性值的误差不在预设范围内,则遍历同层的其它节点,若第一属性值与同层其它节点的属性值均不一致,则停止遍历;将当前层与根节点之间的层数,确定为所述目标属性信息与该iTree的匹配度。

[0270] 确定匹配度单元112,用于将所有iTree对应的匹配度的综合值,确定为所述目标数据与所述异常属性模型的匹配度。

[0271] 其中,在iForest中每个iTree构建过程包括:由根节点开始从上至下构建iTree的每个节点。

[0272] 步骤1:随机选择一个执行属性,并在异常数据集合的剩余异常数据中随机选择一个执行异常数据,并将执行异常数据的属性信息中与执行属性对应执行属性值确定为一个节点。

[0273] 步骤2:在剩余异常数据中排除执行异常数据。

[0274] 步骤3:按执行属性的执行属性值对剩余异常数据进行分类;具体包括:将执行属性的属性值小于执行属性值的异常数据归属于左子树,将执行属性的属性值大于执行属性值的异常数据归属于右子树;

[0275] 步骤4:重复执行步骤1、步骤2和步骤3,递归的构造左子树和右子树,直到满足以下条件之一则终止:条件1:用于构造iTree的剩余异常数据只有一个异常数据或者多个相同的异常数据;条件2:iTree的高度达到预设高度。

[0276] 如图1所示,本申请提供了一种异常消息检测系统。其中包括处理设备和与所述处理设备100相连的多个缓存服务器200。

[0277] 所述多个缓存服务器200,用于存储基于历史缓存数据组建的特征信息集合。

[0278] 所述处理设备100,用于从所述多个缓存服务器中确定与目标数据对应的目标属性信息;其中,所述目标数据用于表示待检测的目标消息;将所述目标属性信息与预先设置的异常属性模型进行匹配,并确定所述目标数据与所述异常属性模型的匹配度;其中,所述预设异常属性模型由所述历史缓存数据中的异常数据集合中异常数据的属性信息组成,每个异常数据的属性信息均基于所述历史缓存数据组建的特征信息集合确定;若对所述匹配度进行归一化的数据值大于预设数据值,则确定所述目标消息为异常消息。

[0279] 本申请提供的处理设备100具有两种实现方式:

[0280] 第一种实现方式:处理设备100包括一个执行服务器。

[0281] 如图12所示,处理设备100包括第一服务器。

[0282] 其中,每个缓存服务器中存储有共用特征信息以及与发送方标识对应的私有特征信息;那么,处理设备100基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息,具体包括:

[0283] 所述第一服务器101,具体用于利用所述目标数据中的目标发送方账户标识和目标发送方机器标识,计算目标发送方标识;并根据预先存储的发送方标识与缓存服务器标识的对应关系,确定与所述目标发送方标识对应的目标缓存服务器,向所述目标缓存服务器发送目标发送方标识;基于共用特征信息和目标私有特征信息确定与目标数据对应的目

标属性信息。

[0284] 所述目标缓存服务器200，用于获取共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；并将所述共用特征信息和目标私有特征信息发送至所述处理设备。

[0285] 处理设备包括一个执行服务器可以本申请的技术方案，但是执行服务器既用于执行异常消息检测过程，又用于执行确定目标数据的目标属性信息的过程。这样会降低执行服务器的处理效率。

[0286] 第二种实现方式：处理设备100包括两个执行服务器。

[0287] 如图13所示，处理设备100包括第一服务器101和第二服务器102。第二服务器102与多个缓存服务器200相连。

[0288] 则所述处理设备基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息，具体包括：

[0289] 所述第二服务器102，用于获取所述第一服务器101发送的目标数据，利用所述目标数据中的目标发送方账户标识和目标发送方机器标识，计算目标发送方标识；并根据预先存储的发送方标识与缓存服务器标识的对应关系，确定与所述目标发送方标识对应的目标缓存服务器，向所述目标缓存服务器发送目标发送方标识；基于共用特征信息和目标私有特征信息确定与目标数据对应的目标属性信息；并将所述目标属性信息发送至第一服务器101。

[0290] 则所述第一服务器101，用于在获取目标数据之后，将目标数据发送至第二服务器102，并获取所述第二服务器102发送的目标属性信息；

[0291] 所述目标缓存服务器200，用于获取共用特征信息以及与所述目标发送方标识对应的目标私有特征信息；并将所述共用特征信息和目标私有特征信息发送至所述处理设备。

[0292] 本实施例方法所述的功能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设备可读取存储介质中。基于这样的理解，本申请实施例对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一台计算设备(可以是个人计算机，服务器，移动计算设备或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM，Read-Only Memory)、随机存取存储器(RAM，Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0293] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。

[0294] 对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下，在其它实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

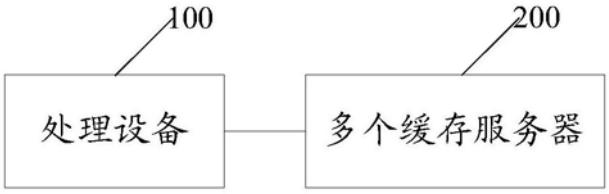

100
200
处理设备
多个缓存服务器

图1

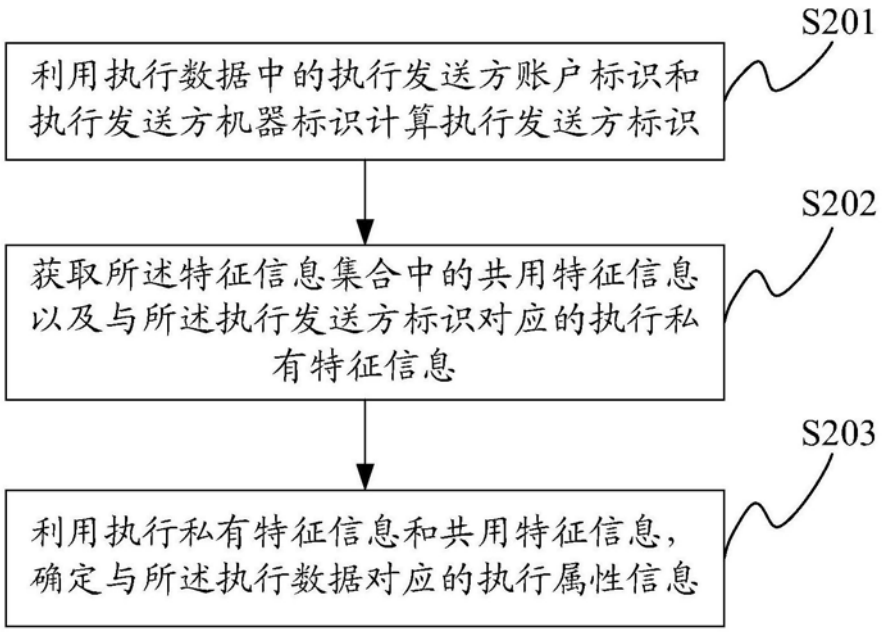

S201
利用执行数据中的执行发送方账户标识和执行发送方机器标识计算执行发送方标识
S202
获取所述特征信息集合中的共用特征信息以及与所述执行发送方标识对应的执行私有特征信息
S203
利用执行私有特征信息和共用特征信息，确定与所述执行数据对应的执行属性信息

图2

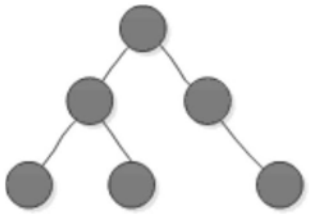


图3a

图3b

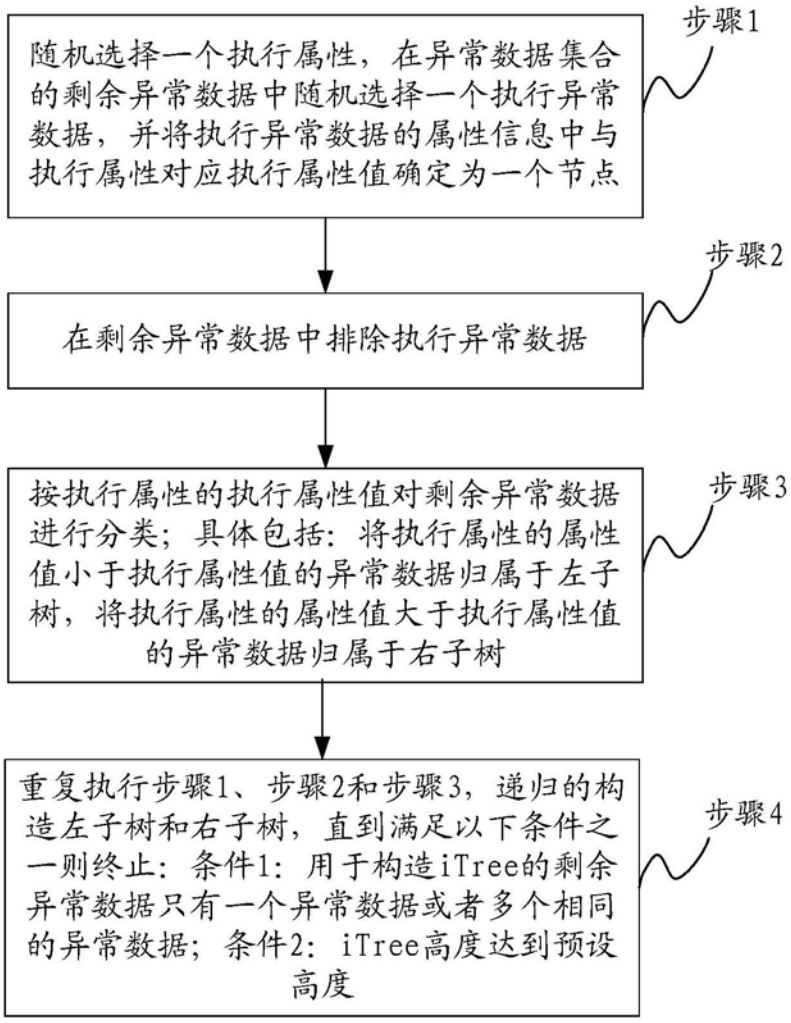

步骤1
随机选择一个执行属性，在异常数据集合的剩余异常数据中随机选择一个执行异常数据，并将执行异常数据的属性信息中与执行属性对应执行属性值确定为一个节点
步骤2
在剩余异常数据中排除执行异常数据
步骤3
按执行属性的执行属性值对剩余异常数据进行分类；具体包括：将执行属性的属性值小于执行属性值的异常数据归属于左子树，将执行属性的属性值大于执行属性值的异常数据归属于右子树
步骤4
重复执行步骤1、步骤2和步骤3，递归的构造左子树和右子树，直到满足以下条件之一则终止：条件1：用于构造iTree的剩余异常数据只有一个异常数据或者多个相同的异常数据；条件2：iTree高度达到预设高度

图4

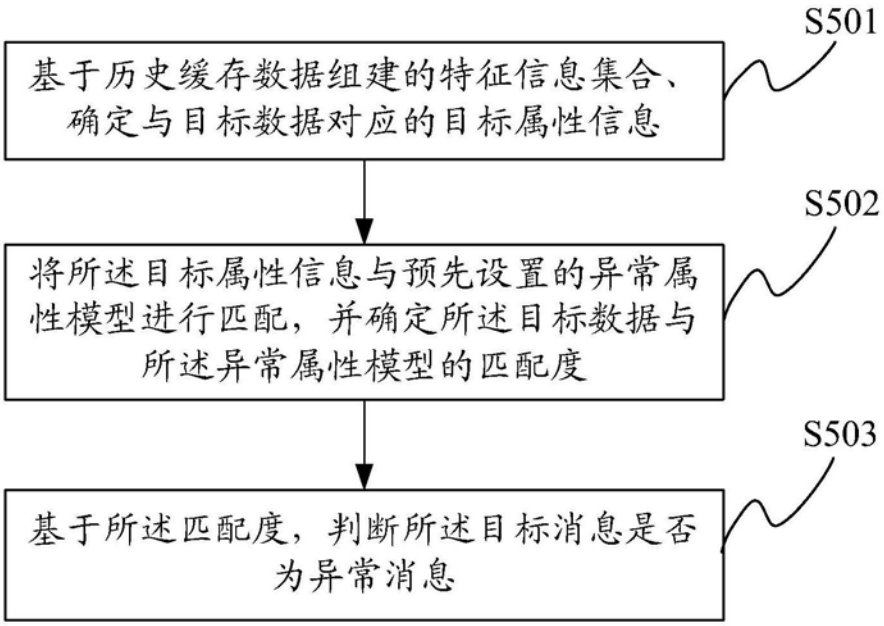

S501
基于历史缓存数据组建的特征信息集合、确定与目标数据对应的目标属性信息
S502
将所述目标属性信息与预先设置的异常属性模型进行匹配，并确定所述目标数据与所述异常属性模型的匹配度
S503
基于所述匹配度，判断所述目标消息是否为异常消息

图5

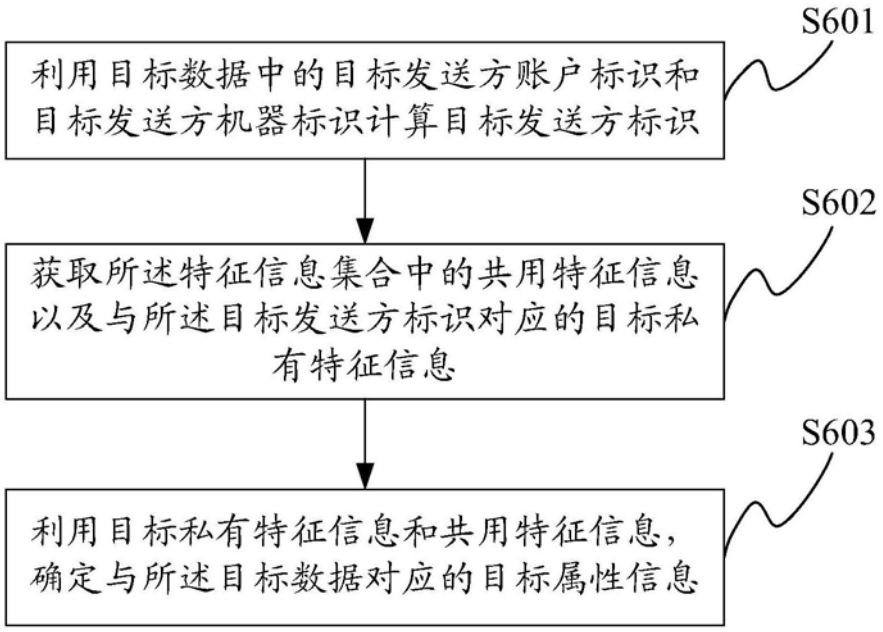

S601
利用目标数据中的目标发送方账户标识和目标发送方机器标识计算目标发送方标识
S602
获取所述特征信息集合中的共用特征信息以及与所述目标发送方标识对应的目标私有特征信息
S603
利用目标私有特征信息和共用特征信息，确定与所述目标数据对应的目标属性信息

图6

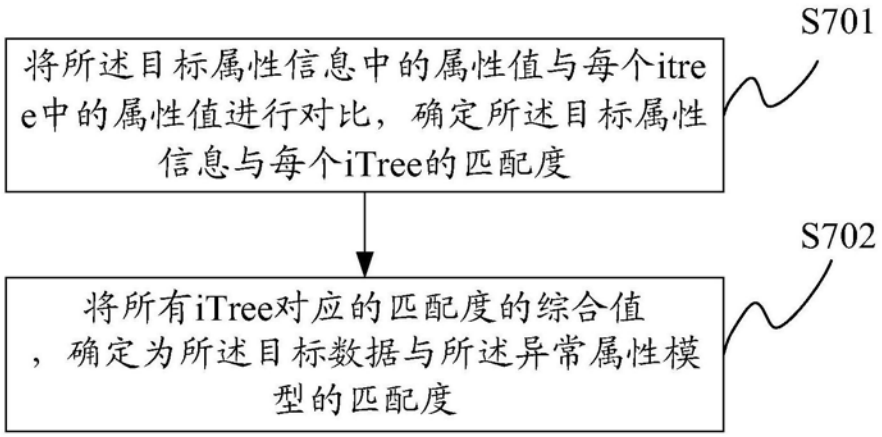

S701
将所述目标属性信息中的属性值与每个itree中的属性值进行对比，确定所述目标属性信息与每个iTree的匹配度
S702
将所有iTree对应的匹配度的综合值，确定为所述目标数据与所述异常属性模型的匹配度

图7

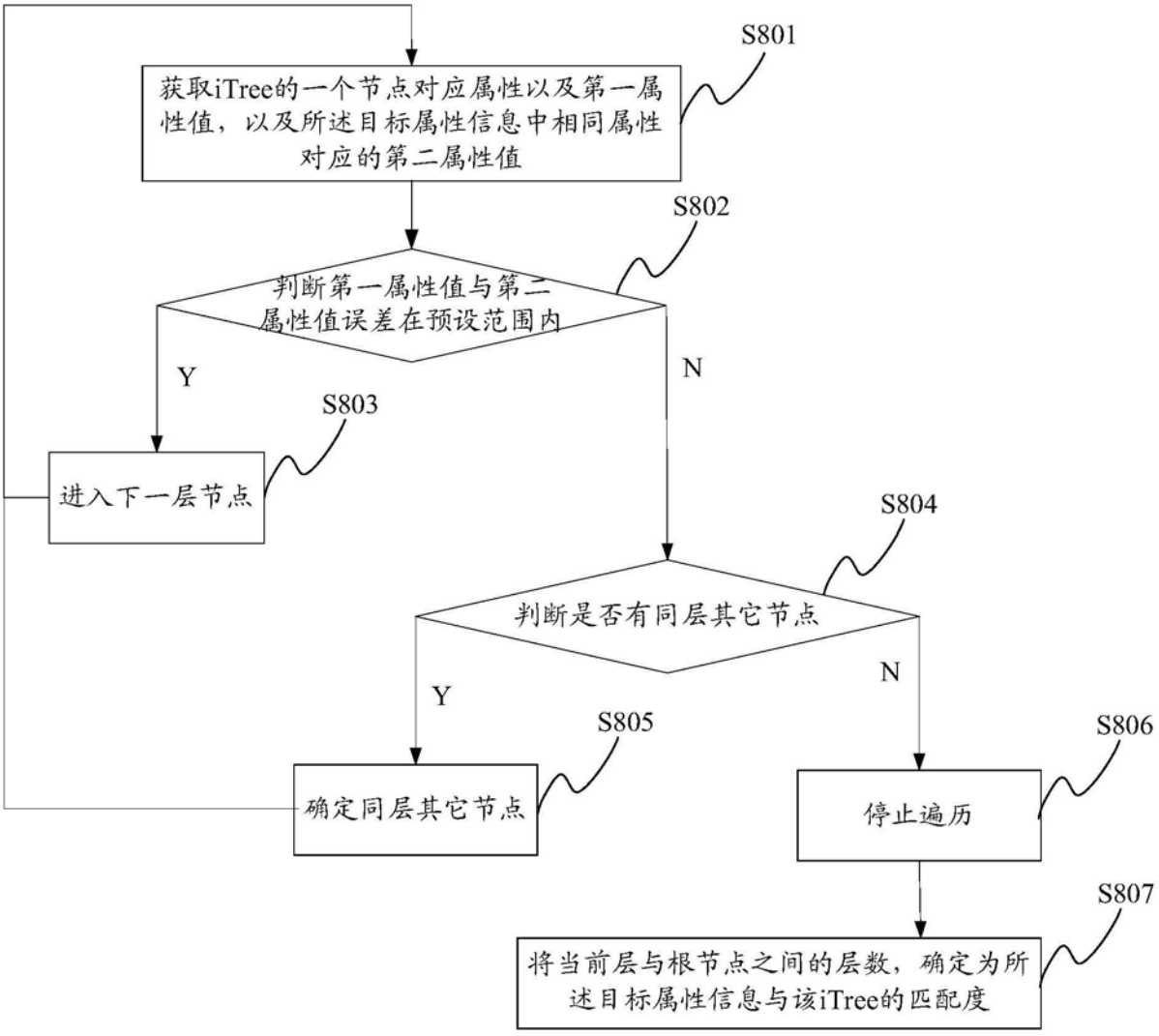

S801
获取iTree的一个节点对应属性以及第一属性值，以及所述目标属性信息中相同属性对应的第二属性值
S802
判断第一属性值与第二属性值误差在预设范围内
Y
N
S803
进入下一层节点
S804
判断是否有同层其它节点
Y
N
S805
确定同层其它节点
S806
停止遍历
S807
将当前层与根节点之间的层数，确定为所述目标属性信息与该iTree的匹配度

图8

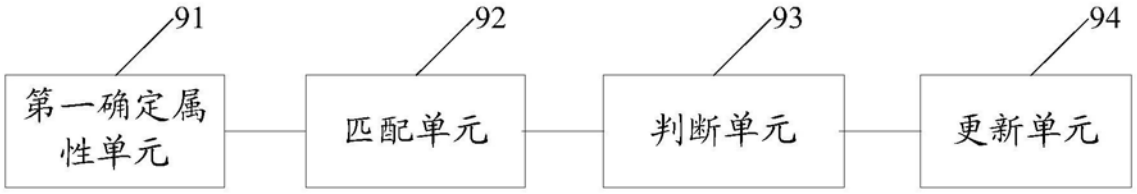

91
92
93
94
第一确定属性单元
匹配单元
判断单元
更新单元

图9

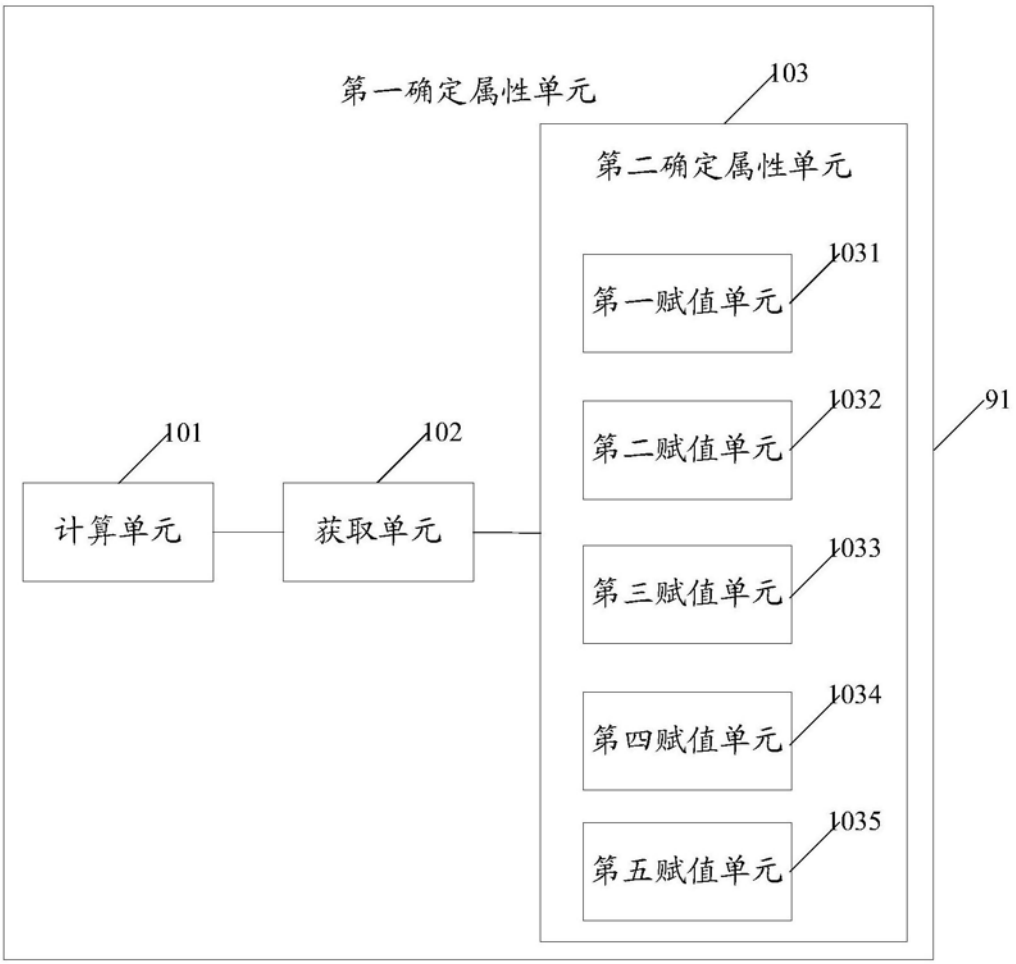

第一确定属性单元
103
第二确定属性单元
1031
第一赋值单元
1032
第二赋值单元
1033
第三赋值单元
1034
第四赋值单元
1035
第五赋值单元
91
101
计算单元
102
获取单元

图10

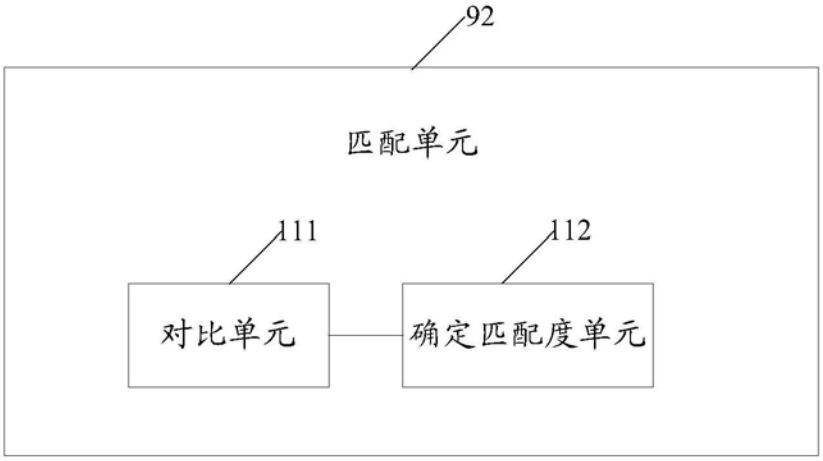


图11

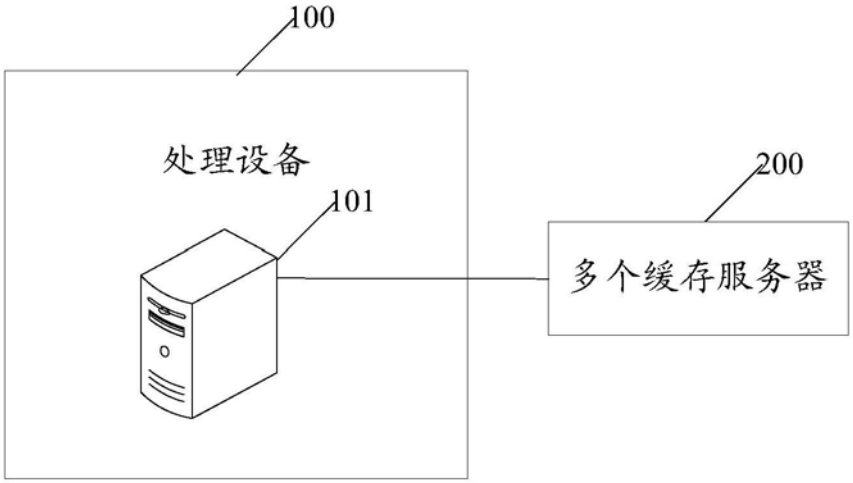


图12

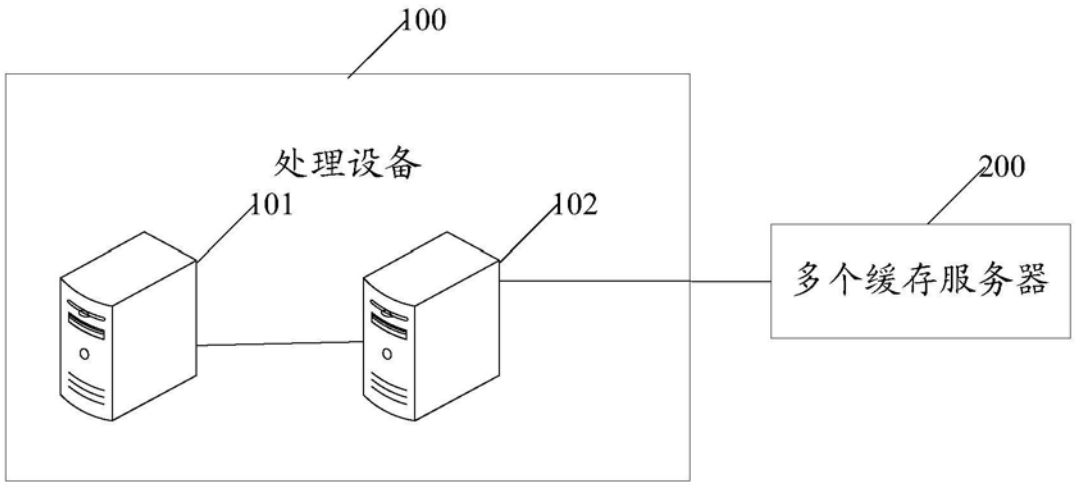


图13