



(12) 发明专利申请

(10) 申请公布号 CN 113344062 A

(43) 申请公布日 2021.09.03

(21) 申请号 202110601169.2

(22) 申请日 2021.05.31

(71) 申请人 北京达佳互联信息技术有限公司
地址 100085 北京市海淀区上地西路6号1
幢1层101D1-7

(72) 发明人 任豫峰 许杰浩 李旭 郁其雨

(74) 专利代理机构 广州华进联合专利商标代理
有限公司 44224

代理人 景怀宇

(51) Int. Cl.

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

权利要求书2页 说明书12页 附图3页

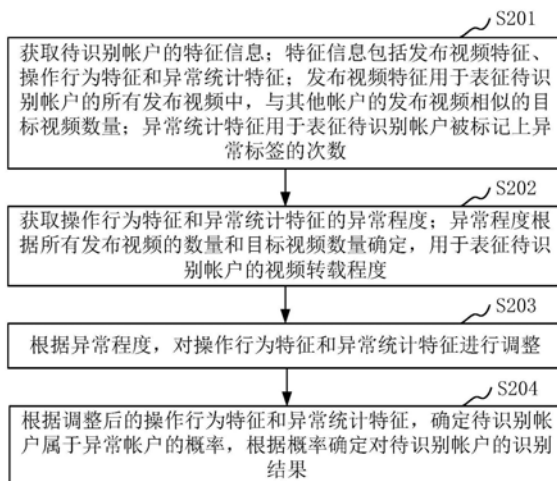
(54) 发明名称

异常帐户识别方法、装置、服务器及存储介
质

(57) 摘要

本公开关于一种异常帐户识别方法、装置、服务器及存储介质,该方法包括:获取待识别帐户的特征信息,包括发布视频特征、操作行为特征和异常统计特征;发布视频特征表征待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;异常统计特征表征待识别帐户被标记上异常标签的次数;获取操作行为特征和异常统计特征的异常程度;异常程度根据所有发布视频的数量和目标视频数量确定,用于表征待识别帐户的视频转载程度;根据异常程度,调整操作行为特征和异常统计特征;根据调整后的操作行为特征和异常统计特征,确定待识别帐户属于异常帐户的概率,根据概率确定对待识别帐户的识别结果。本公开可以提高异常帐户识别的准确性。

CN 113344062 A



1. 一种异常帐户识别方法,其特征在于,包括:

获取待识别帐户的特征信息;所述特征信息包括发布视频特征、操作行为特征和异常统计特征;所述发布视频特征用于表征所述待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;所述异常统计特征用于表征所述待识别帐户被标记上异常标签的次数;

获取所述操作行为特征和所述异常统计特征的异常程度;所述异常程度根据所述所有发布视频的数量和所述目标视频数量确定,用于表征所述待识别帐户的视频转载程度;

根据所述异常程度,调整所述操作行为特征和所述异常统计特征;

根据调整后的操作行为特征和异常统计特征,确定所述待识别帐户属于异常帐户的概率,根据所述概率确定对所述待识别帐户的识别结果。

2. 根据权利要求1所述的异常帐户识别方法,其特征在于,所述根据调整后的操作行为特征和异常统计特征,确定所述待识别帐户属于异常帐户的概率,包括:

拼接所述调整后的操作行为特征和异常统计特征,得到所述待识别帐户的目标特征;

将所述目标特征输入预先训练的异常帐户识别模型,所述异常帐户识别模型对输入的所述目标特征进行分类处理,输出包括类别标签和所述类别标签对应的预测概率;其中,所述类别标签中异常类别标签对应的预测概率为所述待识别帐户属于异常帐户的概率。

3. 根据权利要求2所述的异常帐户识别方法,其特征在于,所述预先训练的异常帐户识别模型通过以下方式获得,包括:

获取样本异常帐户的调整后的操作行为特征和异常统计特征;

将所述样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型,得到所述样本异常帐户属于异常帐户的预测概率;

根据所述样本异常帐户属于异常帐户的预测概率和实际概率之间的差值,得到损失值;

在所述损失值大于等于预设阈值时调整所述待训练的异常帐户识别模型的模型参数,得到调整后的异常帐户识别模型,并重复执行将所述样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型、根据所述样本异常帐户属于异常帐户的预测概率和实际概率之间的差值得到损失值以及在所述损失值大于等于预设阈值时调整所述待训练的异常帐户识别模型的模型参数的过程,直至损失值小于所述预设阈值,则将所述训练后的异常帐户识别模型作为所述预先训练的异常帐户识别模型。

4. 根据权利要求1所述的异常帐户识别方法,其特征在于,所述获取所述操作行为特征和所述异常统计特征的异常程度,包括:

获取所述目标视频数量与所述所有发布视频的数量之间的比值;

获取与所述比值对应的权重系数,所述权重系数作为所述调整后的操作行为特征和异常统计特征的异常程度。

5. 根据权利要求4所述的异常帐户识别方法,其特征在于,所述根据所述异常程度,对所述操作行为特征和所述异常统计特征进行调整,包括:

获取所述权重系数与所述操作行为特征的第一乘积,以及所述权重系数与所述异常统计特征的第二乘积;

分别对所述第一乘积和所述第二乘积进行标准化处理,得到标准化处理后的第一乘积

和第二乘积,所述标准化处理后的第一乘积作为所述调整后的操作行为特征,所述标准化处理后的第二乘积作为所述调整后的异常统计特征。

6. 根据权利要求1至5任一项所述的异常帐户识别方法,其特征在于,所述根据所述概率确定对所述待识别帐户的识别结果,包括:

若所述概率大于预设概率,则确认所述待识别帐户为异常帐户。

7. 一种异常帐户识别装置,其特征在于,包括:

信息获取单元,被配置为执行获取待识别帐户的特征信息;所述特征信息包括发布视频特征、操作行为特征和异常统计特征;所述发布视频特征用于表征所述待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;所述异常统计特征用于表征所述待识别帐户被标记上异常标签的次数;

系数获取单元,被配置为执行获取所述操作行为特征和所述异常统计特征的异常程度;所述异常程度根据所述所有发布视频的数量和所述目标视频数量确定,用于表征所述待识别帐户的视频转载程度;

特征调整单元,被配置为执行根据所述异常程度,调整所述操作行为特征和所述异常统计特征;

结果确定单元,被配置为执行根据调整后的操作行为特征和异常统计特征,确定所述待识别帐户属于异常帐户的概率,根据所述概率确定对所述待识别帐户的识别结果。

8. 一种服务器,其特征在于,包括:

处理器;

用于存储所述处理器可执行指令的存储器;

其中,所述处理器被配置为执行所述指令,以实现如权利要求1至6中任一项所述的异常帐户识别方法。

9. 一种计算机可读存储介质,其特征在于,当所述计算机可读存储介质中的指令由服务器的处理器执行时,使得所述服务器能够执行如权利要求1至6中任一项所述的异常帐户识别方法。

10. 一种计算机程序产品,包括计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1-6任一项所述的异常帐户识别方法。

异常帐户识别方法、装置、服务器及存储介质

技术领域

[0001] 本公开涉及互联网技术领域,尤其涉及一种异常帐户识别方法、装置、服务器、存储介质及计算机程序产品。

背景技术

[0002] 随着互联网技术的发展,人们的信息获取方式已变得多样化,每天都有大量帐户在平台上发布优质内容;然而,有些帐户为了快速吸引大众的关注,会通过程序脚本获取其他帐户已公开发布的内容并再次进行发布。这种行为会侵占平台的流量,打压用户发布新内容的积极性,因此,需要对该类帐户进行识别并处理。

[0003] 相关技术中,通常利用相似度算法比较新发布内容与已发布内容之间的相似度;当一个帐户频繁发布的内容与平台现有内容高度相似时,识别该帐户为发布非原创内容的异常帐户。但是,现有的相似度算法对非原创内容的识别效果较差,从而造成对发布非原创内容的帐户进行识别的准确率较低的问题。

发明内容

[0004] 本公开提供一种异常帐户识别方法、装置、服务器、存储介质及计算机程序产品,以至少解决相关技术中对异常帐户进行识别的准确率较低的问题。本公开的技术方案如下:

[0005] 根据本公开实施例的第一方面,提供一种异常帐户识别方法,包括:

[0006] 获取待识别帐户的特征信息;所述特征信息包括发布视频特征、操作行为特征和异常统计特征;所述发布视频特征用于表征所述待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;所述异常统计特征用于表征所述待识别帐户被标记上异常标签的次数;

[0007] 获取所述操作行为特征和所述异常统计特征的异常程度;所述异常程度根据所述所有发布视频的数量和所述目标视频数量确定,用于表征所述待识别帐户的视频转载程度;

[0008] 根据所述异常程度,调整所述操作行为特征和所述异常统计特征;

[0009] 根据调整后的操作行为特征和异常统计特征,确定所述待识别帐户属于异常帐户的概率,根据所述概率确定对所述待识别帐户的识别结果。

[0010] 在一示例性实施例中,所述根据调整后的操作行为特征和异常统计特征,确定所述待识别帐户属于异常帐户的概率,包括:

[0011] 拼接所述调整后的操作行为特征和异常统计特征,得到所述待识别帐户的目标特征;

[0012] 将所述目标特征输入预先训练的异常帐户识别模型,所述异常帐户识别模型对输入的所述目标特征进行分类处理,输出包括类别标签和所述类别标签对应的预测概率;其中,所述类别标签中异常类别标签对应的预测概率为所述待识别帐户属于异常帐户的概

率。

[0013] 在一示例性实施例中,所述预先训练的异常帐户识别模型通过以下方式获得,包括:

[0014] 获取样本异常帐户的调整后的操作行为特征和异常统计特征;

[0015] 将所述样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型,得到所述样本异常帐户属于异常帐户的预测概率;

[0016] 根据所述样本异常帐户属于异常帐户的预测概率和实际概率之间的差值,得到损失值;

[0017] 在所述损失值大于等于预设阈值时调整所述待训练的异常帐户识别模型的模型参数,得到调整后的异常帐户识别模型,并重复执行将所述样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型、根据所述样本异常帐户属于异常帐户的预测概率和实际概率之间的差值得到损失值以及在所述损失值大于等于预设阈值时调整所述待训练的异常帐户识别模型的模型参数的过程,直至损失值小于所述预设阈值,则将所述训练后的异常帐户识别模型,作为所述预先训练的异常帐户识别模型。

[0018] 在一示例性实施例中,所述获取所述操作行为特征和所述异常统计特征的异常程度,包括:

[0019] 获取所述目标视频数量与所述所有发布视频的数量之间的比值;

[0020] 获取与所述比值对应的权重系数,所述权重系数作为所述调整后的操作行为特征和异常统计特征的异常程度。

[0021] 在一示例性实施例中,所述根据所述异常程度,对所述操作行为特征和所述异常统计特征进行调整,包括:

[0022] 获取所述权重系数与所述操作行为特征的第一乘积,以及所述权重系数与所述异常统计特征的第二乘积;

[0023] 分别对所述第一乘积和所述第二乘积进行标准化处理,得到标准化处理后的第一乘积和第二乘积,所述标准化处理后的第一乘积作为所述调整后的操作行为特征,所述标准化处理后的第二乘积作为所述调整后的异常统计特征。

[0024] 在一示例性实施例中,所述根据所述概率确定对所述待识别帐户的识别结果,包括:

[0025] 若所述概率大于预设概率,则确认所述待识别帐户为异常帐户。

[0026] 在一示例性实施例中,在确认所述待识别帐户为异常帐户之后,还包括:

[0027] 标记所述待识别帐户为异常帐户;

[0028] 在预设时长范围内,拒绝响应所述异常帐户的视频发布请求。

[0029] 根据本公开实施例的第二方面,提供一种异常帐户识别装置,包括:

[0030] 信息获取单元,被配置为执行获取待识别帐户的特征信息;所述特征信息包括发布视频特征、操作行为特征和异常统计特征;所述发布视频特征用于表征所述待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;所述异常统计特征用于表征所述待识别帐户被标记上异常标签的次数;

[0031] 系数获取单元,被配置为执行获取所述操作行为特征和所述异常统计特征的异常程度;所述异常程度根据所述所有发布视频的数量和所述目标视频数量确定,用于表征所

述待识别帐户的视频转载程度；

[0032] 特征调整单元,被配置为执行根据所述异常程度,调整所述操作行为特征和所述异常统计特征；

[0033] 结果确定单元,被配置为执行根据调整后的操作行为特征和异常统计特征,确定所述待识别帐户属于异常帐户的概率,根据所述概率确定对所述待识别帐户的识别结果。

[0034] 在一示例性实施例中,所述结果确定单元,进一步被配置为执行拼接所述调整后的操作行为特征和异常统计特征,得到所述待识别帐户的目标特征;将所述目标特征输入预先训练的异常帐户识别模型,所述异常帐户识别模型对输入的所述目标特征进行分类处理,输出包括类别标签和所述类别标签对应的预测概率;其中,所述类别标签中异常类别标签对应的预测概率为所述待识别帐户属于异常帐户的概率。

[0035] 在一示例性实施例中,所述异常帐户识别装置,还包括:模型训练单元,被配置为执行获取样本异常帐户的调整后的操作行为特征和异常统计特征;将所述样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型,得到所述样本异常帐户属于异常帐户的预测概率;根据所述样本异常帐户属于异常帐户的预测概率和实际概率之间的差值,得到损失值;在所述损失值大于等于预设阈值时调整所述待训练的异常帐户识别模型的模型参数,得到调整后的异常帐户识别模型,并重复执行将所述样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型、根据所述样本异常帐户属于异常帐户的预测概率和实际概率之间的差值得到损失值以及在所述损失值大于等于预设阈值时调整所述待训练的异常帐户识别模型的模型参数的过程,直至损失值小于所述预设阈值,则将所述训练后的异常帐户识别模型,作为所述预先训练的异常帐户识别模型。

[0036] 在一示例性实施例中,所述系数获取单元,进一步被配置为执行获取所述目标视频数量与所述所有发布视频的数量之间的比值;获取与所述比值对应的权重系数,作为所述调整后的操作行为特征和异常统计特征的异常程度。

[0037] 在一示例性实施例中,所述特征调整单元,进一步被配置为执行获取所述权重系数与所述操作行为特征的第一乘积,以及所述权重系数与所述异常统计特征的第二乘积;分别对所述第一乘积和所述第二乘积进行标准化处理,得到标准化处理后的第一乘积和第二乘积,所述标准化处理后的第一乘积作为所述调整后的操作行为特征,所述标准化处理后的第二乘积作为所述调整后的异常统计特征。

[0038] 在一示例性实施例中,所述结果确定单元,进一步被配置为执行若所述概率大于预设概率,则确认所述待识别帐户为异常帐户。

[0039] 在一示例性实施例中,所述结果确定单元,进一步被配置为执行标记所述待识别帐户为异常帐户;在预设时长范围内,拒绝响应所述异常帐户的视频发布请求。

[0040] 根据本公开实施例的第三方面,提供一种服务器,包括:处理器;用于存储所述处理器可执行指令的存储器;其中,所述处理器被配置为执行所述指令,以实现如第一方面的任一项实施例中所述的异常帐户识别方法。

[0041] 根据本公开实施例的第四方面,提供一种计算机可读存储介质,当计算机可读存储介质中的指令由服务器的处理器执行时,使得服务器能够执行上述第一方面任一项实施例中所述的异常帐户识别方法。

[0042] 根据本公开实施例的第五方面,提供一种计算机程序产品,所述计算机程序产品包括计算机程序,计算机程序被处理器执行时实现第一方面的任一项实施例中所述的异常帐户识别方法。

[0043] 本公开的实施例提供的技术方案至少带来以下有益效果:

[0044] 获取待识别帐户的特征信息;特征信息包括发布视频特征、操作行为特征和异常统计特征;发布视频特征用于表征待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;异常统计特征用于表征待识别帐户被标记上异常标签的次数;获取操作行为特征和异常统计特征的异常程度;异常程度根据所有发布视频的数量和目标视频数量确定,用于表征待识别帐户的视频转载程度;根据异常程度,调整操作行为特征和异常统计特征;根据调整后的操作行为特征和异常统计特征,确定待识别帐户属于异常帐户的概率,根据概率确定对待识别帐户的识别结果。本公开提供的异常帐户识别方法可以基于待识别帐户的发布视频特征、操作行为特征和异常统计特征,全方面地识别待识别帐户是否为异常帐户;综合考虑了发布视频特征、操作行为特征和异常统计特征,使得对异常帐户进行识别的准确性更高;避免了仅通过帐户所发布视频进行识别而导致的对异常帐户识别的准确性不高的问题,从而提高了异常帐户识别的准确性。

[0045] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0046] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并与说明书一起用于解释本公开的原理,并不构成对本公开的不当限定。

[0047] 图1是根据一示例性实施例示出的一种异常帐户识别方法的应用环境图。

[0048] 图2是根据一示例性实施例示出的一种异常帐户识别方法的流程图。

[0049] 图3是根据一示例性实施例示出的一种预先训练的异常帐户识别模型的获取步骤的流程图。

[0050] 图4是根据一示例性实施例示出的一种获取操作行为特征和异常统计特征的异常程度步骤的流程图。

[0051] 图5是根据一示例性实施例示出的一种异常帐户识别装置的框图。

[0052] 图6是根据一示例性实施例示出的一种服务器的内部结构图。

具体实施方式

[0053] 为了使本领域普通人员更好地理解本公开的技术方案,下面将结合附图,对本公开实施例中的技术方案进行清楚、完整地描述。

[0054] 需要说明的是,本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本公开的实施例能够以除了在这里图示或描述的那些以外的顺序实施。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0055] 本公开所提供的异常帐户识别方法,可以应用于如图1所示的应用环境中。其中,终端110通过网络与服务器120进行交互,服务器120获取终端110中登录的待识别帐户的特征信息;特征信息包括发布视频特征、操作行为特征和异常统计特征;发布视频特征用于表征待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;异常统计特征用于表征待识别帐户被标记上异常标签的次数;服务器120获取操作行为特征和异常统计特征的异常程度;异常程度根据所有发布视频的数量和目标视频数量确定,用于表征待识别帐户的视频转载程度;根据异常程度,调整操作行为特征和异常统计特征;服务器120根据调整后的操作行为特征和异常统计特征,确定待识别帐户属于异常帐户的概率,根据概率确定对待识别帐户的识别结果。其中,终端110可以是但不限于各种智能手机、平板电脑或笔记本电脑等,服务器120可以为独立的服务器或者是多个服务器组成的服务器集群。

[0056] 图2是根据一示例性实施例示出的一种异常帐户识别方法的流程图,如图2所示,异常帐户识别方法用于如图1所示的服务器120中,至少包括以下步骤S201-S204。

[0057] 在步骤S201中,获取待识别帐户的特征信息;特征信息包括发布视频特征、操作行为特征和异常统计特征;发布视频特征用于表征待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;异常统计特征用于表征待识别帐户被标记上异常标签的次数。

[0058] 其中,待识别帐户可以是正在进行视频发布操作的帐户,也可以是已经发布视频的帐户,还可以是任意已注册的视频发布帐户。

[0059] 特征信息,是指待识别帐户的操作行为在预设评价维度上的体现,并且包括但不限于发布视频特征、操作行为特征和异常统计特征。

[0060] 目标视频,是指待识别帐户发布的与其他帐户所发布视频相似程度超过预设数值的视频,预设数值可以设置为70%、99%等,或者根据实际需要进行调整;例如a视频与b视频的视频相似程度达到90%时,将a视频作为目标视频。

[0061] 一个实施例中,视频的相似程度可以通过视频相似度算法进行检测得到,常用的视频相似度算法例如最小均方值 (Mean Square Error) 算法,能够通过计算每帧画面的相似度得到视频的相似度。目标视频可以是与现有视频完全一样的视频,也可以是对现有视频进行了剪切、速度调整、视频效果调整等处理后得到的视频文件;目标视频还可以是被服务器检测出存在抄袭的视频内容,例如画面完全一样但经过重新配音的视频,或者其他符合拼凑、重制等行为而产生的视频。

[0062] 发布视频特征包括待识别帐户发布的视频与其他帐户所发布视频相似的视频数量、待识别帐户在进行发布视频行为时的行为数据特征;待识别帐户发布的视频与其他帐户所发布视频相似的视频数量可以是视频发布数量、非原创视频数量、转载视频数量等;待识别帐户在进行发布视频行为时的行为数据特征可以是视频发布时间、视频长度等。

[0063] 操作行为特征是指能够体现出该帐户行为的特征数据,例如帐户观看视频时长、登录登出次数、帐户使用设备类型等。

[0064] 异常统计特征是指待识别帐户被标记上异常标签的次数,异常标签可以是服务器根据待识别帐户的行为进行标记,也可以是其他帐户中记录的对待识别帐户作出的标记。例如,服务器检测到待识别帐户的关注对象都是通过搜索特定的帐户名称而添加的,不符

合普通用户的操作习惯,即待识别帐户可能由计算机程序操控,而并非真实用户。再例如,待识别帐户被举报,也会为待识别帐户作出异常标签的标记。

[0065] 进一步地,服务器可以将操作行为特征视作帐户自身的数据特征信息,而异常统计特征可以视为其他帐户的操作行为与待识别帐户产生关联的数据特征信息;例如观看视频时间、登陆登出次数是待识别帐户自身的,不受其他帐户影响的操作行为特征;而被其他帐户举报、被其他帐户放入黑名单等行为,是涉及到两个或更多帐户信息的数据特征,例如A帐户举报B帐户未经许可转载A帐户已发布的a视频,则A帐户的举报记录中会携带有B帐户的特征信息,即A帐户与B帐户之间产生了相关联的操作行为。

[0066] 本步骤服务器通过获取待识别帐户的发布视频特征、操作行为特征和异常统计特征,能够通过用户操作的行为特征,以及来自其他帐户对待识别帐户异常特征的识别,综合地反映出待识别帐户的异常情况,提高了异常帐户的识别效率和效果。

[0067] 在步骤S202中,获取操作行为特征和异常统计特征的异常程度;异常程度根据所有发布视频的数量和目标视频数量确定,用于表征待识别帐户的视频转载程度。

[0068] 其中,转载,是指将其他帐户已经发布的视频以自身帐号的为发布帐号进行发布;因此该转载的视频与其他帐户所发布视频的相似程度则超过一定数值,即转载的视频会被识别为目标视频。视频转载程度,是指待识别帐户中目标视频占全部视频的比例;例如,A帐户共发布视频10个,其中被识别为目标视频的视频数量为10个,则待识别帐户的视频转载程度为100%,即待识别帐号从未发布原创视频,而是全部转载其他帐户已发布的视频;因此,视频转载程度越高,越不符合正常帐户的特点,即异常程度越高。

[0069] 具体地,服务器在确定异常程度时可以先利用视频相似度算法,确定出待识别帐户中各个视频与其他帐户发布的视频的相似度,筛选出待识别帐户中相似度超过一定阈值的视频作为目标视频。之后,确定目标视频的数量与待识别帐户所有发布视频的数量之间的比例,得到用于调节操作行为特征及异常统计特征的异常程度。

[0070] 本步骤服务器通过待识别帐户的发布视频特征,确定得到异常程度,能够对待识别帐户的操作行为特征和异常统计特征进行调整;异常程度表征待识别帐户的视频转载程度,后续通过异常程度调整实现了对操作行为特征和异常统计特征的修正,避免了仅通过帐户所发布视频进行识别而导致的帐户异常识别准确性不高的问题,进而提高了异常帐户识别的准确性。

[0071] 在步骤S203中,根据异常程度,对操作行为特征和异常统计特征进行调整。

[0072] 具体地,异常程度是一种经标准化处理后得到的权重系数,标准化(normalization)是将数据按比例缩放,使之落入一个特定区间;常用的标准化处理方法包括min-max标准化、log函数转换、atan函数转换、归一化方法等,即利用与标准化处理方法对应的函数对数据进行运算,得到的运算结果即为标准化处理后的数据。操作行为特征和异常统计特征也需要进行标准化处理,过程为:利用异常程度所表征的权重系数分别乘以操作行为特征和异常统计特征,得到操作行为特征的第一计算结果和异常统计特征的第一计算结果;根据选取的标准化处理方法的函数,对操作行为特征的第一计算结果和异常统计特征的第一计算结果进行运算,得到操作行为特征的第二计算结果和异常统计特征的第二计算结果,操作行为特征的第二计算结果作为调整后的操作行为特征,异常统计特征的第二计算结果作为调整后的异常统计特征。在步骤S204中,根据调整后的操作行为特征和

异常统计特征,确定待识别帐户属于异常帐户的概率,根据概率确定对待识别帐户的识别结果。

[0073] 其中,异常帐户是指存在未经其他帐户允许,直接将多个其他帐户已发布的视频再次发布同时未注明来源行为的帐户,同时异常帐户较少或从不发布原创视频,与正常帐户的使用习惯存在较大差异。

[0074] 具体地,服务器将操作行为特征和异常统计特征利用异常程度进行调整后,可以用作对待识别帐户进行识别的特征数据;输入预先训练好的异常帐户识别模型后,得到该模型输出的待识别帐户属于异常帐户的概率;最后通过概率值大小确定出待识别帐户是否为异常帐户。

[0075] 本步骤服务器通过调整后的操作行为特征和异常统计特征能够确定出待识别帐户属于异常帐户的概率,并进一步得到识别结果;操作行为特征和异常统计特征能够更直观地体现出异常帐户与正常使用帐户在使用过程中的差异特征,因此具有很强的识别效果,同时具有较高的识别准确度。

[0076] 上述异常帐户识别方法中,通过终端获取待识别帐户的特征信息;特征信息包括发布视频特征、操作行为特征和异常统计特征;发布视频特征用于表征待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;异常统计特征用于表征待识别帐户被标记上异常标签的次数;获取操作行为特征和异常统计特征的异常程度;异常程度根据所有发布视频的数量和目标视频数量确定,用于表征待识别帐户的视频转载程度;根据异常程度,调整操作行为特征和异常统计特征;根据调整后的操作行为特征和异常统计特征,确定待识别帐户属于异常帐户的概率,根据概率确定对待识别帐户的识别结果。本公开提供的异常帐户识别方法综合考虑了发布视频特征、操作行为特征和异常统计特征,使得对异常帐户进行识别的准确性更高;弥补了传统技术避免了仅通过帐户所发布视频进行识别而导致的对异常帐户识别的准确性不高的问题,从而提高了异常帐户识别的准确性。

[0077] 在一示例性实施例中,在步骤S204中,可以进一步包括:拼接调整后的操作行为特征和异常统计特征,得到待识别帐户的目标特征;将目标特征输入预先训练的异常帐户识别模型,异常帐户识别模型对输入的目标特征进行分类处理,输出包括类别标签和类别标签对应的预测概率;其中,类别标签中异常类别标签对应的预测概率为待识别帐户属于异常帐户的概率。

[0078] 预先训练的异常帐户识别模型可以采用传统二分类模型xgboost (eXtreme Gradient Boosting,优化的分布式梯度增强库,可在Gradient Boosting框架下实现机器学习算法),也可以采用其他能够对目标特征进行识别训练的机器学习模型。预先训练的异常帐户识别模型能够对输入的目标特征进行卷积处理,得到包括类别标签和预测概率的分类结果。类别标签,包括正常类别标签和异常类别标签,正常类别标签表示待识别帐户属于正常帐户,异常类别标签表示待识别帐户属于异常帐户;正常类别标签对应的预测概率,表示待识别帐户属于正常帐户的概率;异常类别标签对应的预测概率,表示待识别帐户属于异常帐户的概率。

[0079] 具体地,服务器获取经过异常程度调整后的操作行为特征和异常统计特征,通过拼接完成对目标特征的构建;将目标特征输入预先训练的异常帐户识别模型,得到该模型输出的识别结果,识别结果通常是一个数值,表示输入的目标特征所对应的待识别帐户属

于异常帐户类别标签的概率;根据概率值的大小,可以确定待识别帐户是否为异常帐户。

[0080] 上述实施例,服务器通过预先训练的异常帐户识别模型对输入的目标特征进行处理,通过识别结果可以确定出待识别帐户是否为异常帐户的概率,并进一步确定是否可将待识别帐户确定为异常帐户;相较于传统的帐户类型识别方法,本公开的目标特征中包含有能够体现帐户操作行为的操作行为特征,以及自身和其他帐户在使用过程中反馈的异常标签信息,并且结合了发布视频特征的调整,使得预先训练的异常帐户识别模型能够从多个角度前面的对待识别帐户进行识别,具有较高的准确性。

[0081] 图3是根据一示例性实施例示出的一种异常帐户识别模型的训练方法的流程图,如图3所示,至少包括以下步骤S301-S304。

[0082] 在步骤S301中,获取样本异常帐户的调整后的操作行为特征和异常统计特征。

[0083] 在步骤S302中,将样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型,得到样本异常帐户属于异常帐户的预测概率。

[0084] 在步骤S303中,根据样本异常帐户属于异常帐户的预测概率和实际概率之间的差值,得到损失值。

[0085] 在步骤S304中,在损失值大于等于预设阈值时调整待训练的异常帐户识别模型的模型参数,得到调整后的异常帐户识别模型。

[0086] 在得到调整后的异常帐户识别模型之后,重复执行将样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型、根据样本异常帐户属于异常帐户的预测概率和实际概率之间的差值得到损失值以及在损失值大于等于预设阈值时调整待训练的异常帐户识别模型的模型参数,即重复上述S302-304的过程,直至损失值小于预设阈值,则将训练后的异常帐户识别模型,作为预先训练的异常帐户识别模型。

[0087] 其中,样本异常帐户是指预先被确定为异常的帐户;样本异常帐户可以通过预先的人工筛选得到;样本异常帐户中的操作行为特征是指能够体现出样本异常帐户行为的特征数据,例如帐户观看视频时长、登录登出次数、帐户使用设备类型等,可以由操作行为日志记录的各个操作行为发生的次数,统计得到操作行为特征;也可以通过预先设置的埋点获取操作行为的次数、时长、频率等操作行为特征。样本异常帐户中的异常统计特征是指样本异常帐户被标记上异常标签的次数,异常标签可以是服务器根据样本异常帐户的行为所标记得到,也可以是其他帐户中记录的对待识别帐户所作出的标记。

[0088] 其中,待训练的异常帐户识别模型是指还不具有异常帐户识别能力的模型,因此需要通过样本异常帐户的特征信息对该模型进行训练,通过损失值对模型的训练参数不断调整,直到该模型对调整后的操作行为特征和异常统计特征的进行识别得到的概率值大于预设阈值,且损失值小于预设阈值时,得到可以应用的预先训练的异常帐户识别模型。

[0089] 上述实施例中,服务器通过样本异常帐户所提取出的特征数据,对模型进行训练,能够提高异常帐户识别模型的识别准确率和训练效率。

[0090] 在一示例性实施例中,如图4所示,获取操作行为特征和异常统计特征的异常程度,可以包括步骤S401-S402。

[0091] 在步骤S401中,获取目标视频数量与所有发布视频的数量之间的比值。

[0092] 在步骤S402中,获取与比值对应的权重系数,作为调整后的操作行为特征和异常统计特征的异常程度。

[0093] 权重系数a可以通过以下方式计算得到：

$$[0094] \quad a = \log \left(1 + \frac{cnt_{\text{目标视频}}}{cnt_{\text{所有发布视频}}} \right);$$

[0095] 其中,发布视频特征可以包括 $cnt_{\text{目标视频}}$ 和 $cnt_{\text{所有发布视频}}$, $cnt_{\text{目标视频}}$ 为待识别帐户中被识别为目标视频的数量, $cnt_{\text{所有发布视频}}$ 为待识别帐户中所有发布视频的数量; $\frac{cnt_{\text{目标视频}}}{cnt_{\text{所有发布视频}}}$ 为

目标视频数量与所有发布视频的数量之间的比值; $\log \left(1 + \frac{cnt_{\text{目标视频}}}{cnt_{\text{所有发布视频}}} \right)$ 为选取的标准

化处理方法所对应的函数。

[0096] 进一步地,服务器根据目标视频的具体类型,还可以分为高相似程度的视频数量,以及完全转载的视频数量,即 $cnt_{\text{目标视频}} = cnt_{\text{高相似度视频}} + cnt_{\text{转载视频}}$;高相似度视频和转载视频可以通过视频相似度算法得到,可以设置当视频相似度为100%时目标视频的具体类型为转载视频,当视频相似度大于某个阈值例如70%时目标视频的具体类型为高相似度视频。

[0097] 上述实施例,服务器能够根据待识别帐户的发布视频特征中目标视频数量与所有发布视频的数量得到比值,然后通过标准化函数对比值进行处理,得到相应的权重系数;即本公开将发布视频特征转化为权重系数的方式,对待识别帐户进行异常帐户识别,提高了异常帐户识别的准确性。

[0098] 在一示例性实施例中,步骤S203可以进一步包括:获取权重系数与操作行为特征的第一乘积,以及权重系数与异常统计特征的第二乘积;分别对第一乘积和第二乘积进行标准化处理,得到标准化处理后的第一乘积和第二乘积,标准化处理后的第一乘积作为调整后的操作行为特征,标准化处理后的第二乘积作为调整后的异常统计特征。

[0099] 具体地,对操作行为特征和异常统计特征进行调整可以具体为,将权重系数分别与操作行为特征以及异常统计特征相乘,得到与操作行为特征对应的第一乘积和与异常统计特征对应的第二乘积;通过预设的标准化函数分别对第一乘积以及第二乘积进行标准化处理,即得到调整后的操作行为特征和异常统计特征。预设的标准化函数可以选择min-max标准化(Min-max normalization),log函数转换,atan函数转换,z-score标准化(zero-mean normalization)等。

[0100] 上述实施例通过利用权重系数对操作行为特征和异常统计特征进行调整,巧妙地将发布视频特征分别与操作行为特征和异常统计特征进行融合,提高了异常帐户识别的准确性。

[0101] 在一示例性实施例中,步骤S204可以进一步包括:若概率大于预设概率,则确认待识别帐户为异常帐户。

[0102] 具体地,例如设置预设概率为80%,则预先训练的异常帐户识别模型输出待识别帐户属于异常帐户的概率大于等于80%时,确定待识别帐户为异常帐户;当预先训练的异常帐户识别模型输出待识别帐户属于异常帐户的概率小于80%时,确定待识别帐户为正常帐户。预设概率可以根据异常帐户识别模型的训练程度、效果进行调整,其目的是尽可能准确地识别出异常帐户。

[0103] 上述实施例,通过设置预设概率,设置了异常帐户的判断标准,预设概率可以根据模型的训练效果以及模型的实际识别效果进行调整,能够更加准确的对异常帐户作出识别。

[0104] 在一示例性实施例中,在若概率大于预设概率,则确认待识别帐户为异常帐户之后,还可以包括:标记待识别帐户为异常帐户;在预设时长范围内,拒绝响应异常帐户的视频发布请求。

[0105] 具体地,当待识别帐户被标记为异常帐户时,可以设置一定的预设时长范围限制该异常帐户的视频发布功能,具体可以通过拒绝相应异常帐户发送的视频发布请求的形式进行。例如预设时长范围设置为7天,则该帐户在被识别为异常帐户的7天内,无法发布新的视频。进一步地,还可以通过限制异常帐户的视频发布次数、发布视频长度等多种方式对异常帐户进行限制。

[0106] 上述实施例,通过拒绝相应异常帐户的视频发布请求,减少了异常帐户发布其他帐户已发布视频的数量,降低了非原创内容对相似度算法的干扰,提高了对发布非原创内容的帐户进行识别的准确率。

[0107] 在一示例性实施例中,还提供了一种异常帐户识别方法,具体包括:

[0108] 从预先被识别为是样本异常帐户中,分别获取发布视频特征、操作行为特征和异常统计特征;其中发布视频特征包括发布视频数量,转载视频的数量,以及与现有视频相似度超过一定阈值的相似视频的数量等;操作行为特征包括观看视频时长、登陆次数以及登出次数等,通过操作行为特征可以判断出帐号是否为由人工或机器批量程序化操作;异常统计特征包括被举报次数、异常关注来源等。

[0109] 首先通过转载视频的数量、与现有视频相似度超过一定阈值的视频的数量以及发布视频数量,确定出样本异常帐户所发布视频中异常视频的比例,然后利用对数函数对该比例进行标准化,得到权重系数 α 作为调整操作行为特征及异常统计特征的异常程度:

$$[0110] \quad \alpha = \log\left(1 + \frac{cnt_{\text{转载视频}} + cnt_{\text{相似视频}}}{cnt_{\text{发布视频}}}\right);$$

[0111] 其中, $cnt_{\text{转载视频}}$ 表示转载视频的数量, $cnt_{\text{相似视频}}$ 表示与现有视频相似度超过一定阈值的相似视频的数量, $cnt_{\text{发布视频}}$ 表示发布视频数量, $\log\left(1 + \frac{cnt_{\text{转载视频}} + cnt_{\text{相似视频}}}{cnt_{\text{发布视频}}}\right)$ 表示选取的

标准化处理方法所对应的函数。

[0112] 然后利用该权重系数 α 分别乘以操作行为特征以及异常统计特征,实现通过发布视频特征对操作行为特征以及异常统计特征作出调整,并将调整后的操作行为特征以及异常统计特征进行标准化处理,得到用于训练模型的特征。

[0113] 使用机器学习中的传统二分类模型xgboost结合标准化后的特征进行训练,当该模型满足一定训练条件后,例如识别率达到预设阈值时,即判断该模型训练完成,得到可以对异常帐号进行识别的异常帐号识别模型。

[0114] 上述实施例主要通过从用户使用帐号的维度,即通过操作行为特征和异常统计特征对异常帐号进行识别;同时结合帐号自身的发布视频特征得到权重系数 α ,对操作行为特征和异常统计特征进行调整,提高了异常帐户识别模型的准确率以及召回率,提高了异常帐户识别模型的鲁棒性。综合考虑了发布视频特征、操作行为特征和异常统计特征,使得对

异常帐户进行识别的准确性更高;弥补了传统技术避免了仅通过帐户所发布视频进行识别而导致的对异常帐户识别的准确性不高的问题,从而提高了异常帐户识别的准确性。

[0115] 应该理解的是,虽然图2-4的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,图2-4中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0116] 图5是根据一示例性实施例示出的一种异常帐户识别装置的框图。参照图5,该装置包括信息获取单元501,系数获取单元502,特征调整单元503和结果确定单元504。

[0117] 信息获取单元501,被配置为执行获取待识别帐户的特征信息;特征信息包括发布视频特征、操作行为特征和异常统计特征;发布视频特征用于表征待识别帐户的所有发布视频中,与其他帐户的发布视频相似的目标视频数量;异常统计特征用于表征待识别帐户被标记上异常标签的次数。

[0118] 系数获取单元502,被配置为执行获取操作行为特征和异常统计特征的异常程度;异常程度根据所有发布视频的数量和目标视频数量确定,用于表征待识别帐户的视频转载程度。

[0119] 特征调整单元503,被配置为执行根据异常程度,调整操作行为特征和异常统计特征。

[0120] 结果确定单元504,被配置为执行根据调整后的操作行为特征和异常统计特征,确定待识别帐户属于异常帐户的概率,根据概率确定对待识别帐户的识别结果。

[0121] 在一示例性实施例中,结果确定单元504,进一步被配置为执行拼接调整后的操作行为特征和异常统计特征,得到待识别帐户的目标特征;将目标特征输入预先训练的异常帐户识别模型,异常帐户识别模型对输入的目标特征进行分类处理,输出包括类别标签和类别标签对应的预测概率;其中,类别标签中异常类别标签对应的预测概率为待识别帐户属于异常帐户的概率。

[0122] 在一示例性实施例中,异常帐户识别装置中还包括模型训练单元,模型训练单元被配置为执行获取样本异常帐户的调整后的操作行为特征和异常统计特征;将样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型,得到样本异常帐户属于异常帐户的预测概率;根据样本异常帐户属于异常帐户的预测概率和实际概率之间的差值,得到损失值;在损失值大于等于预设阈值时调整待训练的异常帐户识别模型的模型参数,得到调整后的异常帐户识别模型,并重复执行将样本异常帐户的调整后的操作行为特征和异常统计特征输入待训练的异常帐户识别模型、根据样本异常帐户属于异常帐户的预测概率和实际概率之间的差值得到损失值以及在损失值大于等于预设阈值时调整待训练的异常帐户识别模型的模型参数的过程,直至损失值小于预设阈值,则将训练后的异常帐户识别模型,作为预先训练的异常帐户识别模型。

[0123] 在一示例性实施例中,系数获取单元502,进一步被配置为执行获取目标视频数量与所有发布视频的数量之间的比值;获取与比值对应的权重系数,作为调整后的操作行为特征和异常统计特征的异常程度。

[0124] 在一示例性实施例中,特征调整单元503,进一步被配置为执行获取权重系数与操作行为特征的第一乘积,以及权重系数与异常统计特征的第二乘积;分别对第一乘积和第二乘积进行标准化处理,得到标准化处理后的第一乘积和第二乘积,标准化处理后的第一乘积作为调整后的操作行为特征,标准化处理后的第二乘积作为调整后的异常统计特征。

[0125] 在一示例性实施例中,结果确定单元504,进一步被配置为执行若概率大于预设概率,则确认待识别帐户为异常帐户。

[0126] 在一示例性实施例中,结果确定单元504,还被配置为执行标记待识别帐户为异常帐户;在预设时长范围内,拒绝响应异常帐户的视频发布请求。

[0127] 关于上述实施例中的装置,其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0128] 图6是根据一示例性实施例示出的一种用于执行上述的异常帐户识别方法的设备600的框图。例如,设备600可以为服务器。参照图6,设备600包括处理组件620,其进一步包括一个或多个处理器,以及由存储器622所代表的存储器资源,用于存储可由处理组件620的执行的指令,例如应用程序。存储器622中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外,处理组件620被配置为执行指令,以执行上述异常帐户识别方法。

[0129] 设备600还可以包括一个电源组件624被配置为执行设备600的电源管理,一个有线或无线网络接口626被配置为将设备600连接到网络,和一个输入输出(I/O)接口628。设备600可以操作基于存储在存储器622的操作系统,例如Windows Server™,Mac OS X™,Unix™,Linux™,FreeBSD™或类似。

[0130] 在示例性实施例中,还提供了一种包括指令的存储介质,例如包括指令的存储器622,上述指令可由设备600的处理器执行以完成上述方法。存储介质可以是非临时性计算机可读存储介质,例如,该非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0131] 在示例性实施例中,还提供了一种包括指令的计算机可读存储介质,例如包括指令的存储器504,上述指令可由设备500的处理器520执行以完成上述方法。例如,计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0132] 在示例性实施例中,还提供一种计算机程序产品,该计算机程序产品包括计算机程序,该计算机程序存储在计算机可读存储介质中,设备的至少一个处理器从该计算机可读存储介质读取并执行该计算机程序,使得设备执行本公开的任一项实施例中的异常帐户识别方法。

[0133] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其它实施方案。本公开旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和精神由下面的权利要求指出。

[0134] 应当理解的是,本公开并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本公开的范围仅由所附的权利要求来限制。

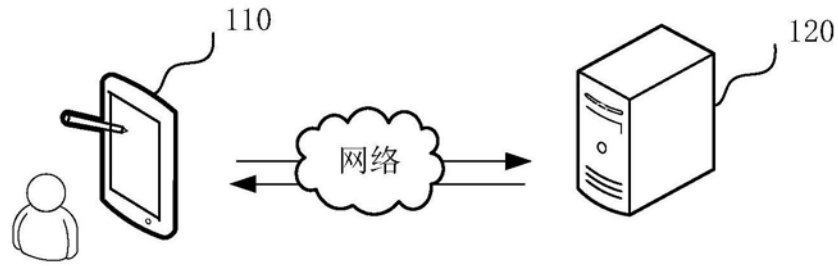


图1

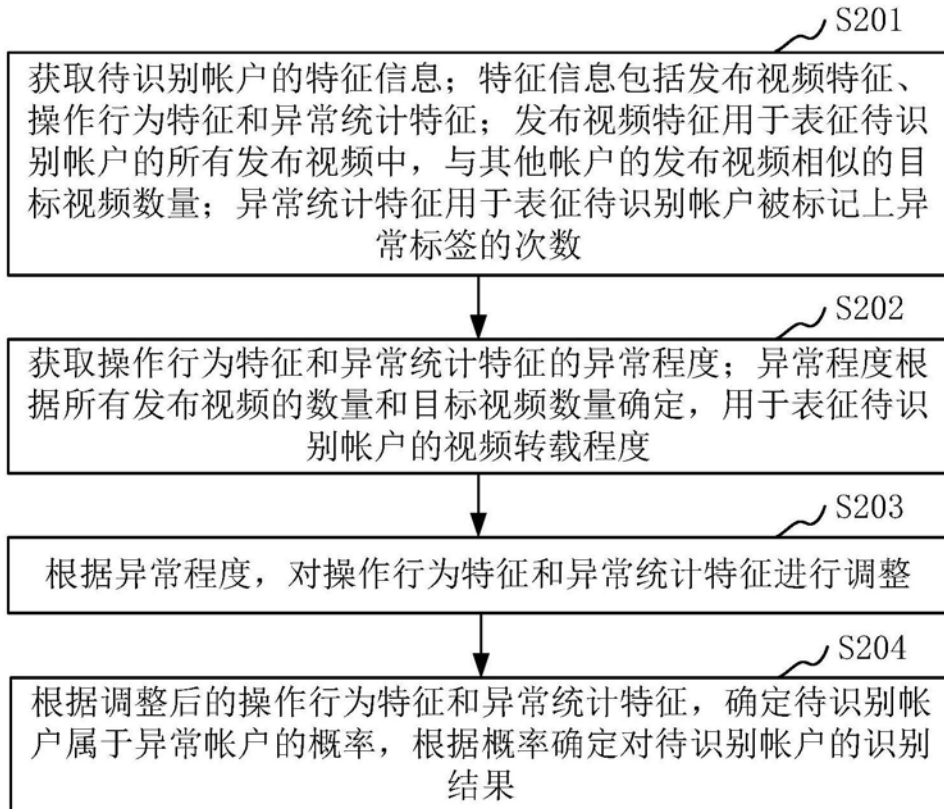


图2

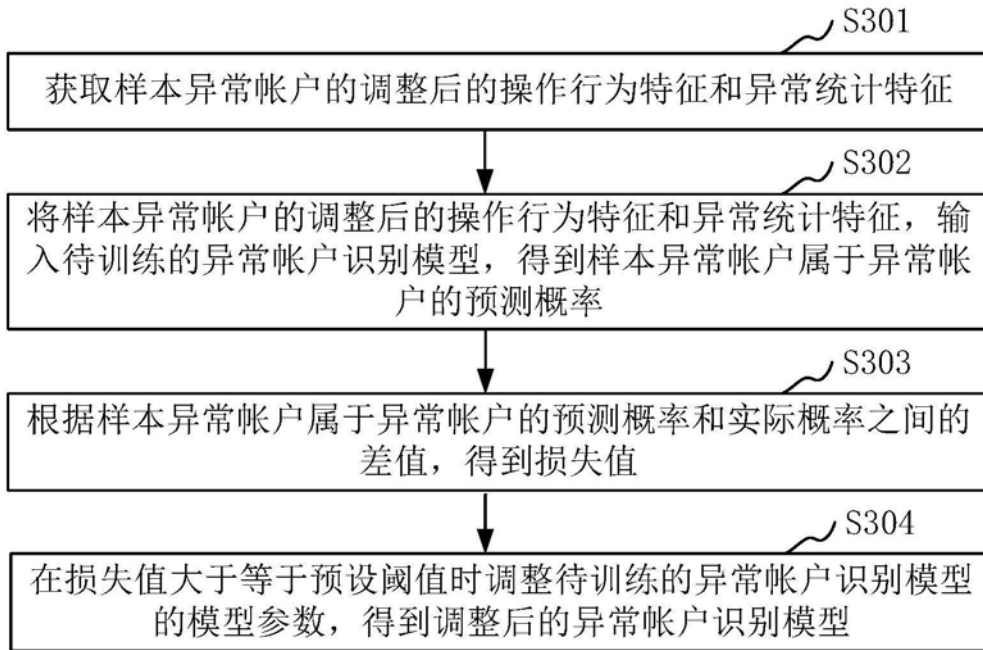


图3

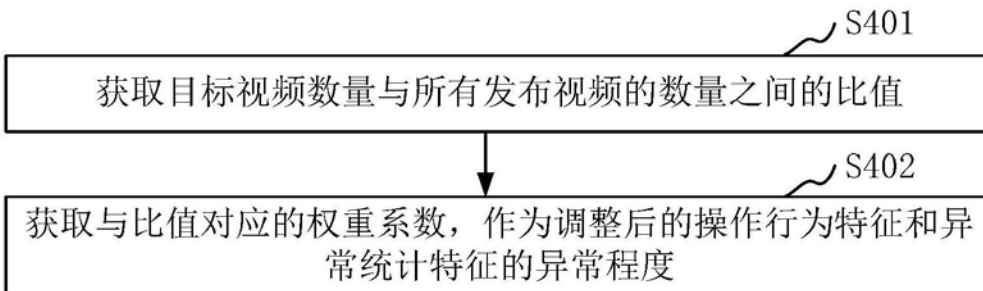


图4

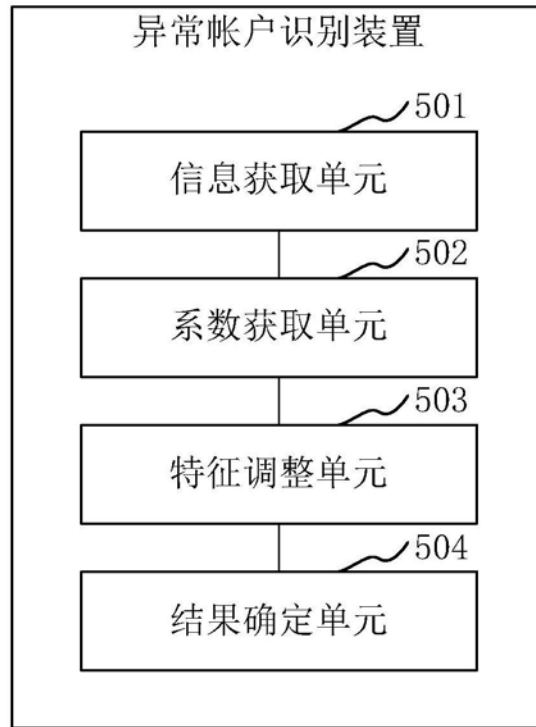


图5

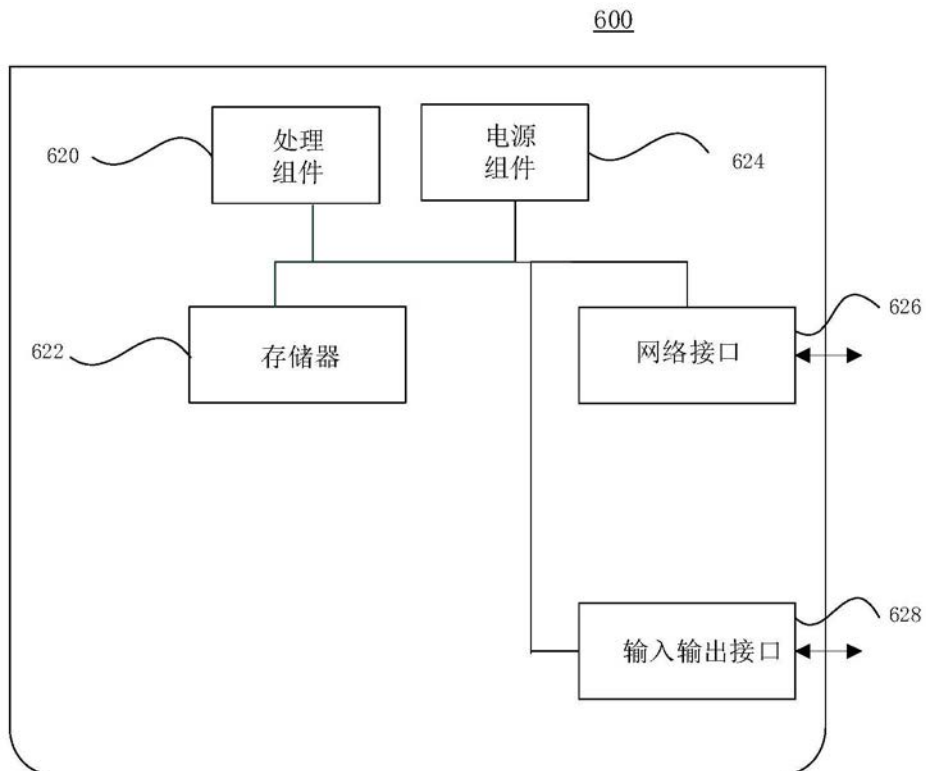


图6