



(12) 发明专利

(10) 授权公告号 CN 111611973 B

(45) 授权公告日 2023.05.12

(21) 申请号 202010486582.4

G06V 10/74 (2022.01)

(22) 申请日 2020.06.01

G06V 10/82 (2022.01)

(65) 同一申请的已公布的文献号

G06N 3/0464 (2023.01)

申请公布号 CN 111611973 A

H04N 17/00 (2006.01)

G06F 16/2458 (2019.01)

(43) 申请公布日 2020.09.01

审查员 罗重凡

(73) 专利权人 广州市百果园信息技术有限公司

地址 511402 广东省广州市番禺区市桥街

兴泰路274号C栋西塔5-13层

(72) 发明人 姚苏桁 战志恒 苏郑博 王怡

(74) 专利代理机构 北京三高永信知识产权代理

有限责任公司 11138

专利代理师 邢少真

(51) Int. Cl.

G06V 20/40 (2022.01)

G06V 40/16 (2022.01)

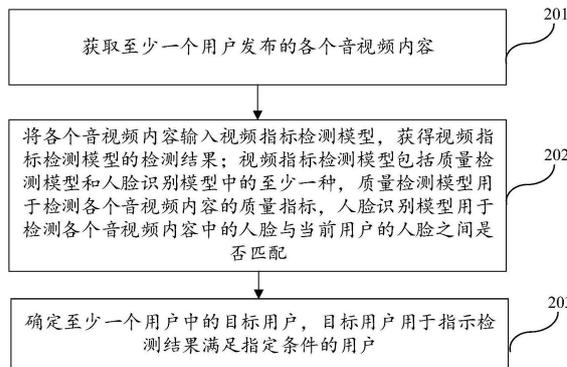
权利要求书2页 说明书12页 附图4页

(54) 发明名称

目标用户识别的方法、装置及存储介质

(57) 摘要

本公开揭示了一种目标用户识别的方法、装置及存储介质,首先获取至少一个用户发布的音视频内容,将各个音视频内容输入视频指标检测模型中,根据模型输出的检测结果确定至少一个用户是否为目标用户。通过上述方案,通过将音视频内容输入视频指标检测模型对各个用户进行筛选,可以保证在待检测的用户量级较大时可以依靠机器学习模型自动识别筛选出优质原创生产者作为目标用户,避免了依靠人工标注造成目标用户被遗漏的情况,从而提高了优质UGC生产者识别的效率和准确性。



1. 一种目标用户识别的方法,其特征在于,所述方法包括:

获取至少一个用户发布的各个音视频内容;

将所述各个音视频内容输入视频指标检测模型,获得所述视频指标检测模型的检测结果;所述视频指标检测模型包括质量检测模型和人脸识别模型中的至少一种,所述质量检测模型用于检测所述各个音视频内容的质量指标,所述人脸识别模型用于检测所述各个音视频内容中的人脸与当前用户的人脸之间是否匹配;

确定所述至少一个用户中的目标用户,所述目标用户是对应的所述检测结果满足指定条件的用户。

2. 根据权利要求1所述的方法,其特征在于,响应于所述视频指标检测模型包括所述质量检测模型,所述方法包括:

将所述各个音视频内容输入所述质量检测模型,获得所述各个音视频内容的音画质分数;

确定所述至少一个用户中,对应的音视频内容的音画质分数高于第二分数阈值的用户为所述目标用户。

3. 根据权利要求2所述的方法,其特征在于,所述将所述各个音视频内容输入所述质量检测模型,获得所述各个音视频内容的音画质分数,包括:

通过所述质量检测模型对所述各个音视频内容进行检测,得到所述各个音视频内容各自的音视频质量信息;

根据所述各个音视频内容各自的音视频质量信息获取所述各个音视频内容的质量达标率,所述质量达标率用于指示所述各个音视频内容中,对应的音视频质量信息达到质量指标的音视频内容所占的比例;

将所述各个音视频内容的质量达标率作为所述各个音视频内容的音画质分数。

4. 根据权利要求1所述的方法,其特征在于,响应于所述视频指标检测模型包括所述人脸识别模型,所述方法包括:

将所述各个音视频内容对应的图像信息,以及对应用户的人脸图像输入所述人脸识别模型,获得所述人脸识别模型输出的人脸匹配率,所述人脸匹配率用于指示所述各个音视频内容的图像信息与对应用户的人脸图像的匹配程度;

确定所述至少一个用户中,对应的人脸匹配率高于匹配率阈值的用户为所述目标用户。

5. 根据权利要求1所述的方法,其特征在于,所述获取至少一个用户发布的各个音视频内容之前,还包括:

根据初步识别标准,对各个用户的至少一种基本信息数据进行判断,获得所述各个用户各自的初步识别结果,所述初步识别结果用于指示对应的用户是否通过初步筛选;

根据所述各个用户各自的初步识别结果,从所述各个用户中筛选出通过初步筛选的所述至少一个用户。

6. 根据权利要求5所述的方法,其特征在于,所述根据初步识别标准,对各个用户的至少一种基本信息数据进行判断,获得所述各个用户各自的初步识别结果,包括:

从数据库中获取第一用户的各种类型的基本信息数据;所述第一用户是所述各个用户中的任意一个;

按照所述初步识别标准,对所述各种类型的基本信息数据分别进行指标分值转换,得到所述各种类型的基本信息数据分别对应的指标分数;

根据所述各种类型的基本信息数据分别对应的权重参数和指标分数,确定所述第一用户的综合分数;

根据所述第一用户的综合分数,获得所述第一用户的初步识别结果。

7.根据权利要求6所述的方法,其特征在于,所述根据所述第一用户的综合分数,获得所述第一用户的初步识别结果,包括:

响应于所述综合分数大于或者等于第一分数阈值,确定所述第一用户的初步识别结果为通过初步筛选;

响应于所述综合分数小于所述第一分数阈值,确定所述第一用户的初步识别结果为未通过初步筛选。

8.根据权利要求6所述的方法,其特征在于,所述基本信息数据包括消费信息数据以及用户行为信息数据中的至少一种。

9.一种目标用户识别的装置,其特征在于,所述装置包括:

内容获取模块,用于获取至少一个用户发布的各个音视频内容;

结果获取模块,用于将所述各个音视频内容输入视频指标检测模型,获得所述视频指标检测模型的检测结果;所述视频指标检测模型包括质量检测模型和人脸识别模型中的至少一种,所述质量检测模型用于检测所述各个音视频内容的质量指标,所述人脸识别模型用于检测所述各个音视频内容中的人脸与当前用户的人脸之间是否匹配;

目标用户确定模块,用于确定所述至少一个用户中的目标用户,所述目标用户是对应的所述检测结果满足指定条件的用户。

10.一种计算机设备,其特征在于,计算机设备包含处理器和存储器,所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如权利要求1至8任一所述的目标用户识别的方法。

11.一种计算机可读存储介质,其特征在于,所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现如权利要求1至8任一所述的目标用户识别的方法。

目标用户识别的方法、装置及存储介质

技术领域

[0001] 本公开涉及机器学习技术领域,特别是涉及一种目标用户识别的方法、装置及存储介质。

背景技术

[0002] 随着互联网技术的日益发展,由用户自己上传音视频内容的网站或应用也越来越多,这种由用户自己上传的音视频内容可以称为用户原创内容(User Generated Content, UGC),而上传UGC的用户可以称为UGC生产者。

[0003] 随着UGC生产者的数量不断增加,如何筛选出优质的UGC生产者是UGC发布平台的服务商急需解决的问题。在相关技术中,针对提供UGC服务的应用程序,为了识别优质的UGC生产者,该应用程序的服务器可以首先通过对内容的上传来源进行分类,然后根据内容的播放次数、内容的播放时长以及用户的点赞量等后验数据进行计算,得到候选集合,最后通过人工发掘和手动标注从候选集合中筛选出优质的UGC生产者。

[0004] 然而,相关技术中的方案识别UGC生产者的方式容易因误判内容上传方式而挖掘出大量无效账号,对人工依赖强,还容易因量级过大以致优质的UGC生产者被遗漏的情况,从而导致优质的UGC生产者识别的效率和准确性不高。

发明内容

[0005] 本公开提供一种目标用户识别的方法、装置及存储介质,可以通过多层筛选提高目标用户的识别效率和准确性,技术方案如下:

[0006] 一方面,提供了一种目标用户识别的方法,其特征在于,所述方法包括:

[0007] 获取至少一个用户发布的各个音视频内容;

[0008] 将所述各个音视频内容输入视频指标检测模型,获得所述视频指标检测模型的检测结果;所述视频指标检测模型包括质量检测模型和人脸识别模型中的至少一种,所述质量检测模型用于检测所述各个音视频内容的质量指标,所述人脸识别模型用于检测所述各个音视频内容中的人脸与所述当前用户的人脸之间是否匹配;

[0009] 确定所述至少一个用户中的目标用户,所述目标用户是对应的所述检测结果满足指定条件的用户。

[0010] 一方面,提供了一种目标用户识别的装置,其特征在于,所述装置包括:

[0011] 内容获取模块,用于获取至少一个用户发布的各个音视频内容;

[0012] 结果获取模块,用于将所述各个音视频内容输入视频指标检测模型,获得所述视频指标检测模型的检测结果;所述视频指标检测模型包括质量检测模型和人脸识别模型中的至少一种,所述质量检测模型用于检测所述各个音视频内容的质量指标,所述人脸识别模型用于检测所述各个音视频内容中的人脸与所述当前用户的人脸之间是否匹配;

[0013] 目标用户确定模块,用于确定所述至少一个用户中的目标用户,所述目标用户是对应的所述检测结果满足指定条件的用户。

[0014] 又一方面,提供了一种计算机设备,该计算机设备包含处理器和存储器,存储器中存储由至少一条指令、至少一段程序、代码集或指令集,至少一条指令、至少一段程序、代码集或指令集由处理器加载并执行以实现上述任一可选实现方式所述的目标用户识别的方法。

[0015] 另一方面,提供了一种计算机可读存储介质,该存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,至少一条指令、至少一段程序、代码集或指令集由处理器加载并执行以实现上述任一可选实现方式所述的目标用户识别的方法。

[0016] 至少一个方面涉及一种计算机程序产品,该计算机程序产品被配置为使得:在包括处理器和存储器的数据处理系统上执行时使数据处理系统执行上述方面的方法。该计算机程序产品可以被包括在有形的、非瞬时性的计算机可读介质中或在有形的、非瞬时性的计算机可读介质上提供。

[0017] 本公开的实施例提供的技术方案可以包括以下有益效果:

[0018] 本公开实施例提供的目标用户识别的方案,首先获取至少一个用户发布的音视频内容,将各个音视频内容输入视频指标检测模型中,根据模型可以输出的检测结果确定至少一个用户是否为目标用户。通过上述方案,通过将音视频内容输入视频检测模型对各个用户进行筛选,可以保证在待检测的用户量级较大时可以依靠机器学习模型自动识别筛选出优质原创生产者作为目标用户,避免了依靠人工标注造成目标用户被遗漏的情况,从而提高了优质UGC生产者识别的效率和准确性。

[0019] 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的,并不能限制本公开。

附图说明

[0020] 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,并于说明书一起用于解释本公开的原理。

[0021] 图1是根据一示例性实施例示出的一种目标用户识别系统的示意图;

[0022] 图2是根据一示例性实施例示出的一种目标用户识别的方法的流程示意图;

[0023] 图3是根据另一示例性实施例示出的一种目标用户识别的方法的流程示意图;

[0024] 图4是图3所述实施例涉及的一种目标用户识别的方法示意图;

[0025] 图5是根据一示例性实施例示出的目标用户识别装置的结构方框图;

[0026] 图6是根据一示例性实施例示出的一种计算机设备的结构示意图。

具体实施方式

[0027] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。

[0028] 图1是根据一示例性实施例示出的一种目标用户识别系统的示意图。该目标用户识别系统中包括平台服务器110、用户终端120以及数据库130。

[0029] 用户可以在用户终端120上进入平台服务器110对应的平台场景,用户可以进行在平台场景下直接或者依靠第三方平台上传用户原创内容。

[0030] 用户进入平台场景后,平台服务器110可以记录用户在平台场景下的用户数据作

为部分基本信息数据。

[0031] 其中,基本信息数据可以包括用户在场景中上传生产内容的信息数据、用户在场景中的行为信息数据以及用户的消费信息数据等等。

[0032] 平台服务器110中可以包含有存储器,该存储器可以用来存储各个用户基本信息数据。

[0033] 平台服务器110也可以将收集到的用户基本信息数据发送给数据库130,进行存储。其中,数据库130可以是用来存储该平台收集的用户基本信息数据的专用的数据库。

[0034] 用户终端120可以通过有线或者无线网络与平台服务器110进行数据传输。

[0035] 平台服务器110可以调用数据库130中的用户基本信息数据,进行处理,包括,平台服务器110中可以设置机器学习模型对数据进行挖掘。

[0036] 平台服务器110可以是一台服务器,或者可以是由若干台服务器构成的服务器集群,或者可以包含一个或多个虚拟化平台,或者也可以是一个云计算服务中心。

[0037] 图2是根据一示例性实施例示出的一种目标用户识别的方法的流程示意图。该目标用户识别的方法可以由计算机设备执行。其中,上述计算机设备可以是上述图1所示系统中的平台服务器110。如图2所示,该目标用户识别的方法可以包括以下步骤:

[0038] 在步骤201中,获取至少一个用户发布的各个音视频内容。

[0039] 在本公开实施例中,用户原创内容发布平台是向用户开放用户原创内容发布功能的网络平台,在该用户原创内容发布平台上,用户可以发布音视频内容,也可以查看其它用户发布的音视频内容。相应的,用户原创内容发布平台对应的平台服务器,可以获取至少一个用户在该平台中发布的各个音视频内容。

[0040] 其中,音视频内容为通过该平台的应用程序或者第三方应用程序进行录制并且上传的音视频文件,平台的其他用户可以通过网络下载各个音视频内容或者在线观看各个音视频内容。

[0041] 另外,在获取至少一个用户在用户原创内容UGC发布平台中发布的各个音视频内容之前,可以根据初步识别标准,对各个用户的至少一种基本信息数据进行判断,获得各个用户各自的初步识别结果,初步识别结果用于指示对应的用户是否通过初步筛选。根据各个用户各自的初步识别结果,从各个用户中筛选出通过初步筛选的至少一个用户。

[0042] 在一种可能的实现方式中,可以从数据库中获取第一用户的各种类型的基本信息数据,然后按照初步识别标准,对各种类型的基本信息数据分别进行指标分值转换,得到各种类型的基本信息数据分别对应的指标分数,接着根据各种类型的基本信息数据分别对应的权重参数和指标分数,确定第一用户的综合分数,最后根据第一用户的综合分数,获得第一用户的初步识别结果。

[0043] 其中,第一用户可以是各个用户中的任意一个。

[0044] 其中,当综合分数大于或者等于第一分数阈值时,可以确定第一用户的初步识别结果为通过初步筛选;当综合分数小于第一分数阈值时,可以确定第一用户的初步识别结果为未通过初步筛选。

[0045] 在步骤202中,将各个音视频内容输入视频指标检测模型,获得视频指标检测模型的检测结果;视频指标检测模型包括质量检测模型和人脸识别模型中的至少一种,质量检测模型用于检测各个音视频内容的质量指标,人脸识别模型用于检测各个音视频内容中的

人脸与当前用户的人脸之间是否匹配。

[0046] 其中,质量检测模型可以是用来检测音频质量以及画面质量的一种机器学习模型;质量指标可以是用来衡量音频质量以及画面质量的指标数值。

[0047] 在本公开实施例中,为了筛选出各个用户中的优质原创生产者作为目标用户,便于平台对目标用户进行重点关注和培养,可以通过对各个用户原创内容的各个方面的质量进行检测,获得检测结果。

[0048] 其中,对用户原创内容的检测可以分为音视频的质量检测以及音视频的原创性检测。

[0049] 其中,对音视频的原创性检测可以通过判断当前用户的音视频内容中出现的人像是否与当前用户的人像相对应,若音视频内容中的人像与当前用户的人像一致性较高,则当前用户为原创内容用户的几率较高。

[0050] 在一种可能的实现方式中,视频指标检测模型可以用于检测音视频音质和画质以及音视频中的人像与当前用户的人像的一致性。视频指标检测模型可以分为两个模型分别进行两个方面的检测。

[0051] 其中,质量检测模型可以通过输入各个音视频文件,输出各个音视频文件的音频质量与画面质量情况;人脸识别模型可以通过输入各个音视频文件以及当前用户的人脸信息,输出各个音视频内容中的人脸与当前用户的人脸之间的一致程度。

[0052] 在一种可能的实现方式中,响应于视频指标检测模型包括该质量检测模型,可以将各个音视频内容输入质量检测模型,获得该各个音视频内容的音画质分数。

[0053] 其中,平台服务器可以通过质量检测模型对各个音视频内容进行检测,得到各个音视频内容各自的音视频质量信息,然后根据各个音视频内容各自的音视频质量信息获取各个音视频内容的质量达标率,平台服务器可以将各个音视频内容的质量达标率作为各个音视频内容的音画质分数。

[0054] 其中,质量达标率可以用于指示各个音视频内容中,对应的音视频质量信息达到质量指标的音视频内容所占的比例。

[0055] 在另一种可能的实现方式中,响应于视频指标检测模型包括人脸识别模型,平台服务器可以将各个音视频内容对应的图像信息,以及对应用户的人脸图像输入人脸识别模型,获得人脸识别模型输出的人脸匹配率。

[0056] 其中,人脸匹配率可以用于指示各个音视频内容中,针对于任一用户对应的图像信息与人脸图像匹配的音视频内容所占的比例。

[0057] 其中,人脸匹配率也可以用于指示至少一个用户对应的各个音视频内容的图像信息与对应用户的人脸图像的匹配程度。

[0058] 在步骤203中,确定至少一个用户中的目标用户,目标用户用于指示检测结果满足指定条件的用户。

[0059] 在本公开实施例中,通过上述步骤获取的检测结果,包括音画质检测结果以及人脸一致性的检测结果,可以将检测结果与指定条件进行比较,确定至少一个用户的检测结果满足指定条件,从而确定至少一个用户为目标用户。

[0060] 在一种可能的实现方式中,确定至少一个用户中,对应的音视频内容的音画质分数高于第二分数阈值的用户为目标用户。

[0061] 其中,第二分数阈值可以用于指示确定为目标用户的音视频内容的质量达标率指标。

[0062] 在另一种可能的实现方式中,确定至少一个用户中,对应的人脸匹配率高于匹配率阈值的用户为目标用户。

[0063] 综上所述,本公开实施例提供的目标用户识别的方案,首先获取至少一个用户发布的音视频内容,将各个音视频内容输入视频指标检测模型中,根据模型输出的检测结果确定至少一个用户是否为目标用户。通过上述方案,通过将音视频内容输入视频检测模型对各个用户进行筛选,可以保证在待检测的用户量级较大时可以依靠机器学习模型自动识别筛选出优质原创生产者作为目标用户,避免了依靠人工标注造成目标用户被遗漏的情况,从而提高了优质UGC生产者识别的效率和准确性。

[0064] 图3是根据另一示例性实施例示出的一种目标用户识别的方法的流程示意图。该目标用户识别的方法可以由计算机设备执行。其中,上述计算机设备可以是上述图1所示系统中的平台服务器110。如图3所示,该目标用户识别的方法可以包括以下步骤:

[0065] 在步骤301中,从数据库中获取第一用户的各种类型的基本信息数据。

[0066] 在本公开实施例中,从数据库中分别获取各个用户中的任意一个对应的基本信息数据。

[0067] 在一种可能的实现方式中,用户对应的基本信息数据可以为用户对应的用户行为信息数据以及用户对应的账号消费信息数据中的至少一种。

[0068] 其中,用户行为信息数据可以用来指示用户发布音视频内容(比如UGC)时对应的用户操作数据以及时间数据等。

[0069] 在一种可能的实现方式中,用户行为信息数据包括用户对应的生产内容的活跃度、用户上传生产内容的生产来源、用户对应的用户账号的粉丝数量以及粉丝质量的粉丝情况中的至少一种。

[0070] 比如,当用户行为信息数据是用户对应的生产内容的活跃度时,生产内容可以用来指示用户在内容发布平台发布的音视频内容。例如,用户对应的生产内容的活跃度表示为在指定一段时间内用户上传生产内容到平台服务器的次数或者上传生产内容的数量。

[0071] 当用户行为信息数据是用户上传生产内容的生产来源时,用户上传生产内容的生产来源表示为用户通过何种渠道生产的内容以及通过直接或者间接的方式上传生产内容的生产来源信息数据。

[0072] 当用户行为信息数据是用户对应的用户账号的粉丝数量以及粉丝质量的粉丝情况时,用户对应的用户账号的粉丝数量以及粉丝质量的粉丝情况表示为用户对应的用户帐号中当前实际拥有的粉丝的数量以及各个粉丝的活跃程度。

[0073] 其中,粉丝质量可以通过检测关注该用户的全部用户打开用户的生产内容次数与观看总时长或者是平均时长进行判断;比如,当关注用户的全部用户观看用户的生产内容的平均时长超过了指定阈值,则可以判断用户的粉丝质量达到了平均标准。

[0074] 比如,用户对应的账号消费信息数据也可以是用户对应账户的消费信息数据,用户对应账户的消费信息数据可以包括用户对应的用户账号历史消费信息数据以及平台会员等级信息数据等。

[0075] 在一种可能的实现方式中,数据库可以是在平台服务器中用来收集用户对应的基

本信息数据的数据库,或者是在平台服务器之外的其他存储设备中用来进行基本信息数据存储的数据库。

[0076] 在一种可能的实现方式中,平台服务器可以根据用户对应的用户账户检索对应该用户的数据库中的各个基本信息数据,或者,平台服务器也可以根据基本信息数据的种类,分类对各个用户在数据库中进行检索,获得各个种类的基本信息数据对应的各个用户的数据。

[0077] 在步骤302中,按照初步识别标准,对各种类型的基本信息数据分别进行指标分值转换,得到各种类型的基本信息数据分别对应的指标分数。

[0078] 在本公开实施例中,根据获取到的各个用户的基本信息数据按照基本信息数据的种类获取对应种类的指标分值。通过指标分值对各个用户的各类基本信息数据进行比较判断,得到各个用户的各个基本信息数据对应的判断结果。

[0079] 其中,指标分值可以用来指示基本信息数据达到通过初步筛选的指标对应的分数。

[0080] 在一种可能的实现方式中,各个种类的基本信息数据的指标分值可以根据实时从数据库中获取的数据,进行计算得到。

[0081] 比如,从数据库中获取到若干个用户的生产活跃度信息数据,计算在指定时限内,比如30天内,若干个用户上传生产内容的次数的平均值,当计算得到若干个用户在三十天内生产内容数量的平均值为30条时,可以确定生产活跃度信息数据对应的指标分值可以为30,通过将各个用户的生产活跃度信息数据与指标分值进行比较,可以得到若干个用户的各个基本信息数据对应的判断结果,即在三十天内生产内容数量超过30条的用户判断结果为达标,在三十天内生产内容数量不超过30条的用户判断结果为不达标。

[0082] 在一种可能的实现方式中,将各用户对应的实际的基本信息数据与计算得到的指标分值进行比较,得到与指标分值相差的数据值或者超过指标分值的数据值,根据数据值得到各用户对应的该基本信息数据的单种基本信息数据的分数。

[0083] 比如,对于用户的生产活跃度信息数据,根据用户发布平台的实际属性,可以指定用户的生产活跃度信息数据的指标分数为三十天内生产内容的数量为三十条对应该种基本信息数据的分数为60分,可以指定用户三十天内距离三十条生产内容的数据量相差1条,则分数减少3分,生产内容的数据量超过1条,则分数增加3分,根据指定规则得到用户的生产活跃度信息数据的分数。

[0084] 在步骤303中,根据各种类型的基本信息数据分别对应的权重参数和指标分数,确定第一用户的综合分数。

[0085] 在本公开实施例中,根据各个用户中的任意一个对应的各个种类的基本信息数据的指标分值,结合各个种类基本信息数据对应的权重参数,确定各个用户中的任意一个的综合分数。

[0086] 其中,权重参数可以根据该用户所处的UGC发布平台的属性进行预先设置,可以对各个种类的基本信息数据设置不同的权重参数。

[0087] 其中,UGC发布平台的属性可以用来指示该UGC发布平台对各种基本信息数据的分析侧重情况。

[0088] 在一种可能的实现方式中,根据产品平台对各种基本信息数据对应的影响UGC发

布平台的属性的大小设置对应的权重参数。

[0089] 比如,在一种UGC发布平台中,平台需要获取对平台提供的价值与热度较高的用户,则可以适当的提高消费情况信息数据与粉丝情况信息数据所占的权重参数,使在计算的过程中,对上述的两种基本信息数据的依赖程度更高,更加准确的得到符合标准的用户。

[0090] 在一种可能的实现方式中,通过预先设定的各种基本信息数据对应的权重,对各个种类的基本信息数据对应的分数进行加权计算,得到各个用户对应的综合分数。

[0091] 比如,当一种平台的平台服务器预先设置需要进行计算的基本信息数据为两种,分别是消费情况信息数据与粉丝情况信息数据。当按照初步识别标准,对两种类型的基本信息数据分别进行指标分值转换,可以得到当前用户的消费情况信息数据对应的分数为50,粉丝情况信息数据对应的分数为60,该平台中消费情况信息数据与粉丝情况信息数据分别设置的权重参数是0.3和0.7,则可以计算得到当前用户对应的综合分数为 $50*0.3+60*0.7$ 。

[0092] 在步骤304中,根据第一用户的综合分数,获得第一用户的初步识别结果。

[0093] 在本公开实施例中,平台服务器可以根据各个用户中的任意一个的综合分数,确定各个用户对应的初步识别结果。

[0094] 在一种可能的实现方式中,响应于综合分数大于或者等于第一分数阈值,确定第一用户的初步识别结果为通过初步筛选,响应于综合分数小于该第一分数阈值,确定第一用户的初步识别结果为未通过初步筛选。

[0095] 其中,通过初步筛选的用户的数据信息可以继续获取处理,并且进行下一步的筛选,未通过初步筛选的用户的数据信息可以直接进行舍弃,即不进行后续处理。

[0096] 在步骤305中,根据各个用户各自的初步识别结果,从各个用户中筛选出通过初步筛选的至少一个用户。

[0097] 在本公开实施例中,获取通过初步筛选的用户发布在平台服务器中的各个音视频内容,进行接下来的音视频内容的检测,对各个用户进行进一步筛选。

[0098] 其中,初始目标用户用于指示初步识别结果指示通过初步筛选的用户。

[0099] 在步骤306中,获取至少一个用户发布的各个音视频内容,响应于视频指标检测模型包括质量检测模型,将各个音视频内容输入质量检测模型,获得各个音视频内容的音画质分数。

[0100] 在本公开实施例中,将通过初步筛选的至少一个用户对应的音视频内容输入到质量检测模型中,得到各个音视频内容对应的音画质分数。

[0101] 其中,音视频内容包括音频内容以及图像内容中的至少一种。

[0102] 在一种可能的实现方式中,获取通过初步筛选的各个用户的全部上传到平台服务器中的音视频内容,将其输入到质量检测模型中,进行音频质量和画面质量的检测。

[0103] 其中,质量检测模型可以是一种图像分割模型,是一种视觉领域的深度学习模型,比如,可以是应用卷积神经网络(Convolutional Neural Networks,CNN)的机器学习模型。

[0104] 比如,当获取到通过初步筛选的用户的生产内容为短视频时,可以将用户对应的各段短视频分别输入到,计算机设备中建立完成的质量检测模型中,通过质量检测模型可以输出得到各个用户的各段短视频对应的表示音质情况的分数以及表示画面图像情况的分数作为音画质分数。

[0105] 在一种可能的实现方式中,通过将经过初步筛选的用户的音视频内容输入到质量检测模型中,得到各个音视频内容对应的检测分数作为音画质分数。

[0106] 在一种可能的实现方式中,经过初步筛选的用户对应的音视频内容可以是用户对上传的各个音视频内容或者在指定期限内上传的各个音视频内容。

[0107] 其中,当音视频内容为用户上传的各个音视频内容或者在指定期限内上传的各个音视频内容时,可以分别进行如下两种步骤。

[0108] 1) 当初步筛选后的用户对应的音视频内容用来指示该用户上传的各个音视频内容时,通过质量检测模型对初步筛选后的用户上传的各个音视频内容进行检测,输出总体达标率作为音画质分数。

[0109] 其中,总体达标率可以用来指示检测为达标的音视频内容在全部音视频内容中的占比情况。

[0110] 2) 当初步筛选后的用户对应的音视频内容用来指示在指定期限内上传的音视频内容时,通过质量检测模型对初步筛选后的用户在指定期限内上传的该音视频内容进行检测,输出近期达标率作为音画质分数。

[0111] 其中,近期达标率可以用来指示检测为达标的音视频内容在指定期限内上传的音视频内容中的占比情况。

[0112] 比如,将需要检测的用户对应的短视频输入到质量检测模型中进行检测,短视频可以分为声音部分和图像画面部分,质量检测模型可以分别对声音的质量和图像画面质量进行检测,根据声音的频率以及画面的分辨率帧数等作为参数进行检测,可以获得短视频是否在音画质方面达标,当对用户的全部检测内容资源进行检测之后,可以得到通过检测达标的短视频的数量,将通过检测达标的短视频的数量与用户进行检测的短视频数量作比值,得到达标短视频的占比情况,平台服务器根据占比情况可以获得通过进一步筛选得到的用户。

[0113] 在一种可能的实现方式中,将各个用户的音画质分数进行排序得到其中的平均值,将音画质分数高于平均值的对应的部分用户,确定为通过音画质检测筛选的用户。

[0114] 在另一种可能的实现方式中,根据预先设置的音画质分数的预设指标对各个用户的音画质分数进行比较判断,将音画质分数高于预设的音画质分数指标的用户,确定为通过音画质检测筛选的用户。

[0115] 在步骤307中,响应于视频指标检测模型包括人脸识别模型,将各个音视频内容对应的图像信息,以及对应用户的人脸图像输入人脸识别模型,获得人脸识别模型输出的人脸匹配率。

[0116] 在本公开实施例中,将通过初次筛选达标的至少一个用户对应的音视频内容与各个用户对应的实际人脸图像信息输入到人脸识别模型中。

[0117] 在一种可能的实现方式中,各个用户对应的实际人脸图像信息可以从平台服务器的存储器或者数据库中调取获得。音视频内容可以包括用户的生产内容中的图像内容。

[0118] 其中,人脸识别模型可以是一种图像分割模型,是一种视觉领域的深度学习模型,比如,可以是应用卷积神经网络(Convolutional Neural Networks, CNN)的机器学习模型。

[0119] 在一种可能的实现方式中,通过人脸识别模型获得各个用户的对应的各段音视频内容对应的人脸匹配率以及各个用户对应的整体音视频内容对应的人脸匹配率作为识别

分数。

[0120] 其中,识别分数可以包括总体一致率或者近期一致率。

[0121] 在一种可能的实现方式中,总体一致率可以通过将用户各个音视频内容中的图像部分输入到人脸识别模型中,得到人脸识别通过的音视频内容的数量,将人脸识别通过的音视频内容的数量与整体音视频内容的数量作比值,得到总体一致率作为识别分数。

[0122] 在另一种可能的实现方式中,近期一致率可以通过将指定期限内的用户对应的音视频内容中的图像部分输入到人脸识别模型中,得到人脸识别通过的音视频内容的数量,将人脸识别通过的音视频内容的数量与检测的音视频内容数量作比值,得到近期一致率作为识别分数。

[0123] 在步骤308中,确定至少一个用户中的目标用户。

[0124] 在本公开实施例中,通过质量检测模型和人脸识别模型中的至少一个,对经过初次筛选通过后的用户对应的音视频内容进行检测,根据检测结果,得到满足指定条件的目标用户。

[0125] 其中,目标用户可以用于指示检测结果满足指定条件的用户。

[0126] 在一种可能的实现方式中,根据预先设置或者计算确定的分数指标,确定通过模型检测的用户作为目标用户。

[0127] 比如,通过将各个用户的识别分数或者音画质分数进行排序,选择符合筛选属性训练得到的识别分数,或者取平均值,或者取指定阈值作为分数指标,获取识别分数,或者音画质分数高于分数指标对应的用户作为通过模型筛选的目标用户。

[0128] 在一种可能的实现方式中,通过初次筛选的用户可以通过质量检测模型和人脸识别模型中的至少一种进行接下来的筛选,确定目标用户。

[0129] 其中,用户通过初次筛选之后,可以进行质量检测模型和人脸识别模型两个机器学习模型的筛选。

[0130] 在一种可能的实现方式中,先将经过初次筛选的用户通过音画质模型进行二次筛选,再通过人脸识别模型进行三次筛选,确定目标用户。

[0131] 比如,图4是本公开实施例涉及的一种目标用户识别的方法示意图。如图4所示,在实际场景下,为了对所有的用户原创内容进行自动筛选,可以从数据指标、音画质检测以及人脸图像识别三个维度进行数据挖掘。该目标用户识别的方法包括的步骤如下:

[0132] 在步骤401中,启动识别目标用户的任务,并且通过查询数据库获取识别目标用户所需的各个用户的各种类型的基本信息数据。

[0133] 在步骤402中,获得包括消费信息数据、粉丝情况、生产内容的活跃度以及生产来源在内的多种各个用户的基本信息数据,根据初步识别标准判断各个用户的基本信息数据是否满足初步识别标准。

[0134] 经过第一维度,可以按照各个用户的基本信息数据是否满足初步识别标准,将各个用户进行初步筛选,得到符合初步识别标准的用户以及不符合初步识别标准的用户。

[0135] 在步骤403中,将符合初步识别标准的用户的生产内容输入到用于画面以及音频质量检测的机器模型中,通过质量检测模型对用户的生产内容的音画质情况进行输出,发掘音画质情况达到指标的用户进行接下来的筛选。

[0136] 通过第二维度,质量检测模型对通过初步筛选的用户进行二次筛选,得到音画质

情况达到指标的用户进行接下来的筛选。

[0137] 在步骤404中,将经过音画质检测达到指标的用户的生产内容与用户的用户信息中的实际人像图输入到人脸识别模型中,确定生产内容与用户人脸图像的一致性程度,将一致性程度达到指标的用户筛选出。

[0138] 通过第三维度,人脸识别模型对生产者进行三次筛选,得到一致性程度达到指标的用户,输出最终检测结果,完成目标用户的数据挖掘。

[0139] 在另一种可能的实现方式中,计算机设备可以先将通过初次筛选的用户对应的音视频内容通过人脸识别模型进行二次筛选,再通过质量检测模型进行三次筛选,确定目标用户。

[0140] 比如,将经过初次筛选通过的用户对应的音视频内容与用户信息中的实际人像图像输入人脸识别机器模型中,确定音视频内容与用户人脸图像的一致性程度,将一致性程度达到标准的用户筛选出。通过第二维度,人脸识别模型对用户进行第二次筛选,得到一致性程度达到指标的用户,然后将经过第二次筛选的用户的音视频内容输入到质量检测模型中,对用户的音视频内容的音画质情况进行输出,发掘音画质情况达到指标的用户为目标用户。通过第三维度,质量检测模型对用户进行第三次筛选,得到音画质情况达到标准的用户,完成目标用户的数据挖掘。

[0141] 另外,也可以将通过初次筛选的用户同时进行质量检测模型以及人脸识别模型的筛选,通过比较确定同时通过质量检测模型以及人脸识别模型的用户,确定同时通过质量检测模型以及人脸识别模型的用户为目标用户。

[0142] 比如,将经过初次筛选通过的用户对应的音视频内容同时输入到质量检测模型以及人脸识别模型两个模型中,质量检测模型以及人脸识别模型分别对经过初次筛选通过的用户对应的音视频内容进行识别检测,可以分别得到经过质量检测模型达标的用户以及经过人脸识别模型达标的用户,比较分别获得的两种用户,确定其共同达标的用户作为目标用户。

[0143] 综上所述,本公开实施例提供的目标用户识别的方案,首先获取至少一个用户发布的音视频内容,将各个音视频内容输入视频指标检测模型中,根据模型输出的检测结果确定至少一个用户是否为目标用户。通过上述方案,通过将音视频内容输入视频检测模型对各个用户进行筛选,可以保证在待检测的用户量级较大时可以依靠机器学习模型自动识别筛选出优质原创生产者作为目标用户,避免了依靠人工标注造成目标用户被遗漏的情况,从而提高了优质UGC生产者识别的效率和准确性。

[0144] 图5是根据一示例性实施例示出的目标用户识别装置的结构方框图。该目标用户识别装置可以通过硬件或者软硬结合的方式实现为服务器中的全部或者部分,以执行图2或图3对应实施例所示的方法的全部或部分步骤。该目标用户识别装置可以包括:

[0145] 内容获取模块510,用于获取至少一个用户发布的各个音视频内容;

[0146] 结果获取模块520,用于将所述各个音视频内容输入视频指标检测模型,获得所述视频指标检测模型的检测结果;所述视频指标检测模型包括质量检测模型和人脸识别模型中的至少一种,所述质量检测模型用于检测所述各个音视频内容的质量指标,所述人脸识别模型用于检测所述各个音视频内容中的人脸与所述当前用户的人脸之间是否匹配;

[0147] 目标用户确定模块530,用于确定所述至少一个用户中的目标用户,所述目标用户

是对应的所述检测结果满足指定条件的用户。

[0148] 在一种可能的实现方式中,响应于所述视频指标检测模型包括所述质量检测模型,所述装置包括:

[0149] 分数获取模块,用于将所述各个音视频内容输入所述质量检测模型,获得所述各个音视频内容的音画质分数;

[0150] 第一目标确定模块,用于确定所述至少一个用户中,对应的音视频内容的音画质分数高于第二分数阈值的用户为所述目标用户。

[0151] 在一种可能的实现方式中,所述分数获取模块,包括:

[0152] 质量信息获取子模块,用于通过所述质量检测模型对所述各个音视频内容进行检测,得到所述各个音视频内容各自的音视频质量信息;

[0153] 达标率获取子模块,用于根据所述各个音视频内容各自的音视频质量信息获取所述各个音视频内容的质量达标率,所述质量达标率用于指示所述各个音视频内容中,对应的音视频质量信息达到质量指标的音视频内容所占的比例;

[0154] 分数确定子模块,用于将所述各个音视频内容的质量达标率作为所述各个音视频内容的音画质分数。

[0155] 在一种可能的实现方式中,响应于所述视频指标检测模型包括所述人脸识别模型,所述装置包括:

[0156] 匹配率获取模块,用于将所述各个音视频内容对应的图像信息,以及对应用户的人脸图像输入所述人脸识别模型,获得所述人脸识别模型输出的人脸匹配率,所述人脸匹配率用于指示所述各个音视频内容的图像信息与对应用户的人脸图像的匹配程度;

[0157] 第二目标确定模块,用于确定所述至少一个用户中,对应的人脸匹配率高于匹配率阈值的用户为所述目标用户。

[0158] 在一种可能的实现方式中,所述装置还包括:

[0159] 初步结果获取模块,用于根据初步识别标准,对各个用户的至少一种基本信息数据进行判断,获得所述各个用户各自的初步识别结果,所述初步识别结果用于指示对应的用户是否通过初步筛选;

[0160] 内容获取子模块,用于根据所述各个用户各自的初步识别结果,从所述各个用户中筛选出通过初步筛选的所述至少一个用户。

[0161] 在一种可能的实现方式中,所述初步结果获取模块,包括:

[0162] 数据获取子模块,用于从数据库中获取第一用户的各种类型的基本信息数据;所述第一用户是所述各个用户中的任意一个;

[0163] 指标分数获取子模块,用于按照所述初步识别标准,对所述各种类型的基本信息数据分别进行指标分值转换,得到所述各种类型的基本信息数据分别对应的指标分数;

[0164] 综合分数获取子模块,用于根据所述各种类型的基本信息数据分别对应的权重参数和指标分数,确定所述第一用户的综合分数;

[0165] 初步结果获取子模块,用于根据所述第一用户的综合分数,获得所述第一用户的初步识别结果。

[0166] 在一种可能的实现方式中,所述初步结果获取子模块,包括:

[0167] 第一结果获取单元,用于响应于所述综合分数大于或者等于第一分数阈值,确定

所述第一用户的初步识别结果为通过初步筛选；

[0168] 第二结果获取单元,用于响应于所述综合分数小于所述第一分数阈值,确定所述第一用户的初步识别结果为未通过初步筛选。

[0169] 在一种可能的实现方式中,所述基本信息数据包括消费信息数据以及用户行为信息数据中的至少一种。

[0170] 综上所述,本公开实施例提供的目标用户识别的方案,首先获取至少一个用户发布的音视频内容,将各个音视频内容输入视频指标检测模型中,根据模型输出的检测结果确定至少一个用户是否为目标用户。通过上述方案,通过将音视频内容输入视频检测模型对各个用户进行筛选,可以保证在待检测的用户量级较大时可以依靠机器学习模型自动识别筛选出优质原创生产者作为目标用户,避免了依靠人工标注造成目标用户被遗漏的情况,从而提高了优质UGC生产者识别的效率和准确性。

[0171] 图6是根据一示例性实施例示出的一种计算机设备的结构示意图。所述计算机设备600包括中央处理单元(Central Processing Unit,CPU)601、包括随机存取存储器(Random Access Memory, RAM)602和只读存储器(Read-Only Memory, ROM)603的系统存储器604,以及连接系统存储器604和中央处理单元601的系统总线605。所述计算机设备600还包括帮助计算机设备内的各个器件之间传输信息的基本输入/输出系统(Input/Output, I/O系统)606,和用于存储操作系统608、应用程序609和其他程序模块610的大容量存储设备607。

[0172] 存储器还包括一个或者一个以上的程序,一个或者一个以上程序存储于存储器中,中央处理单元601通过执行该一个或一个以上程序来实现图2或图3所示的方法的全部或者部分步骤。

[0173] 本公开实施例还提供了一种计算机设备存储介质,用于储存为上述测试装置所用的计算机设备软件指令,其包含用于执行上述目标用户识别的方法所设计的程序。

[0174] 本公开实施例还提供了一种计算机程序产品,该计算机程序产品存储有至少一条指令,所述至少一条指令由处理器加载并执行以实现上述图2或图3对应实施例所示的方法的全部或者部分步骤。

[0175] 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。

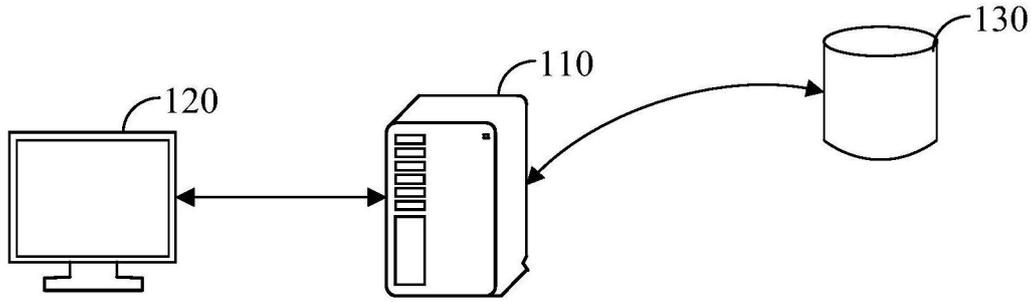


图1

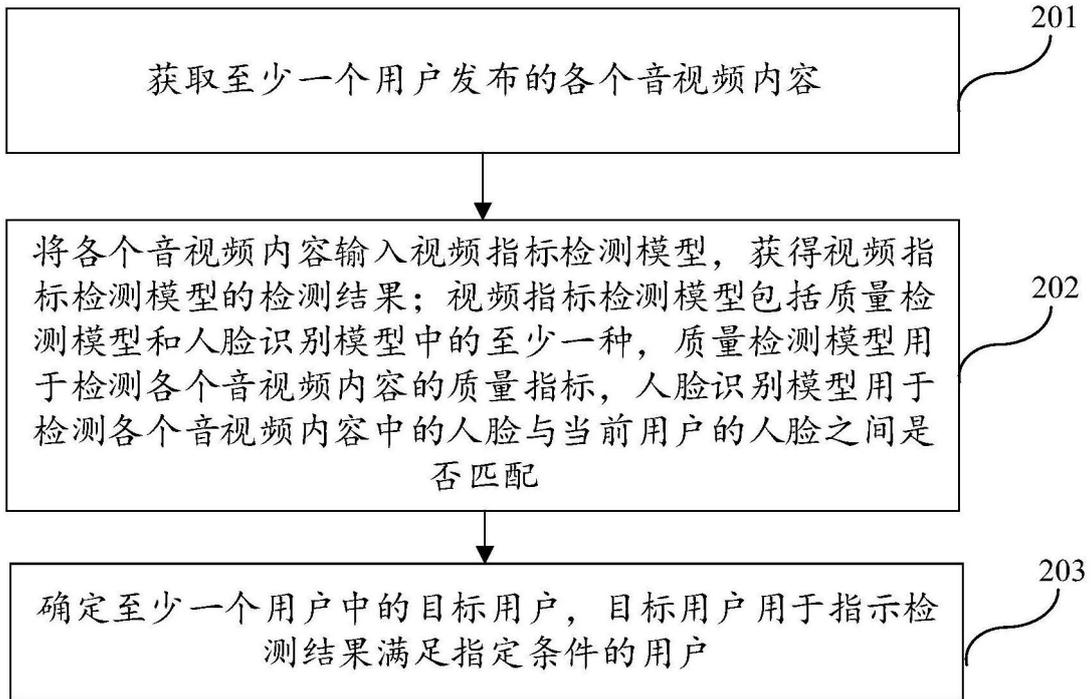


图2

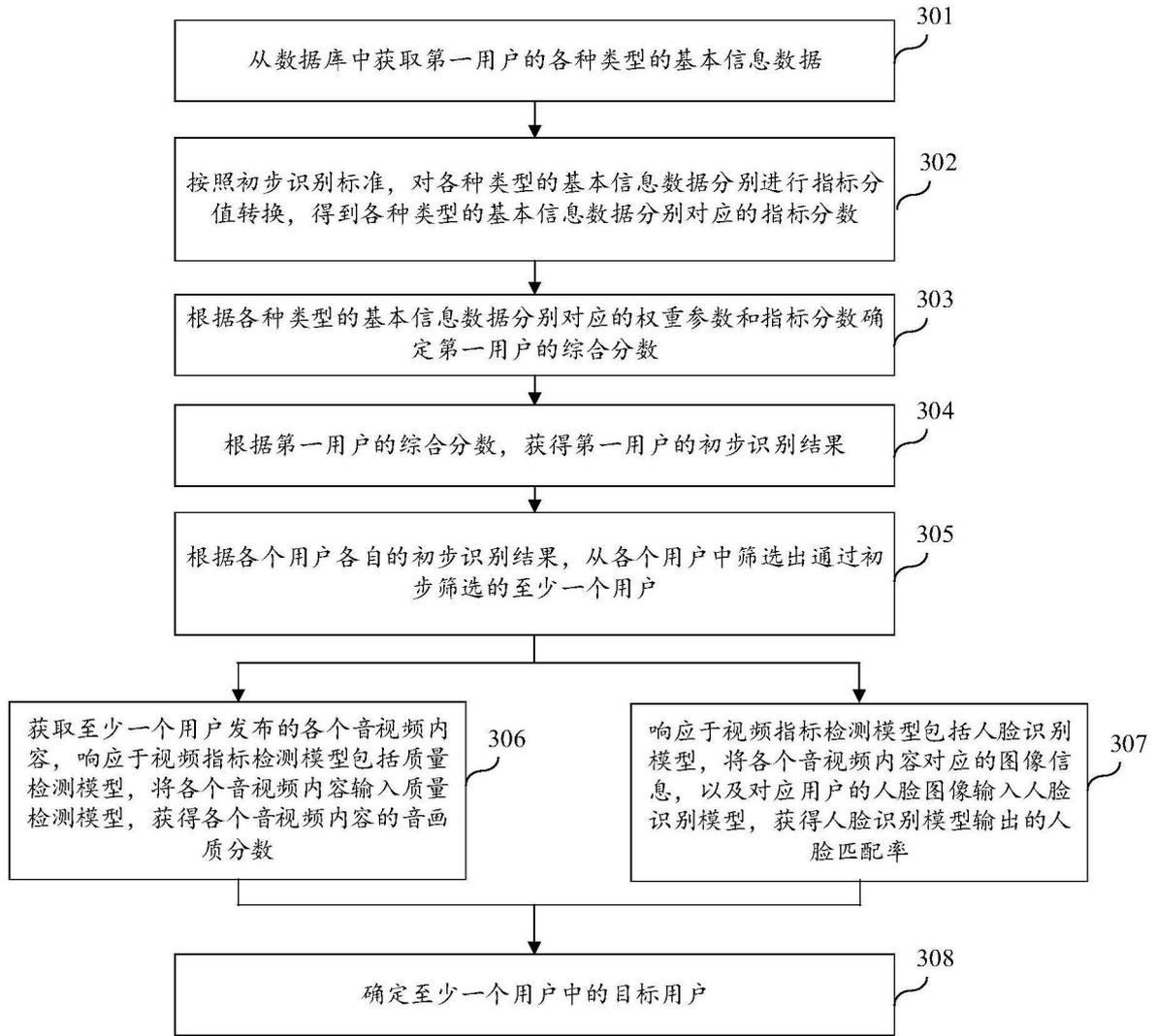


图3

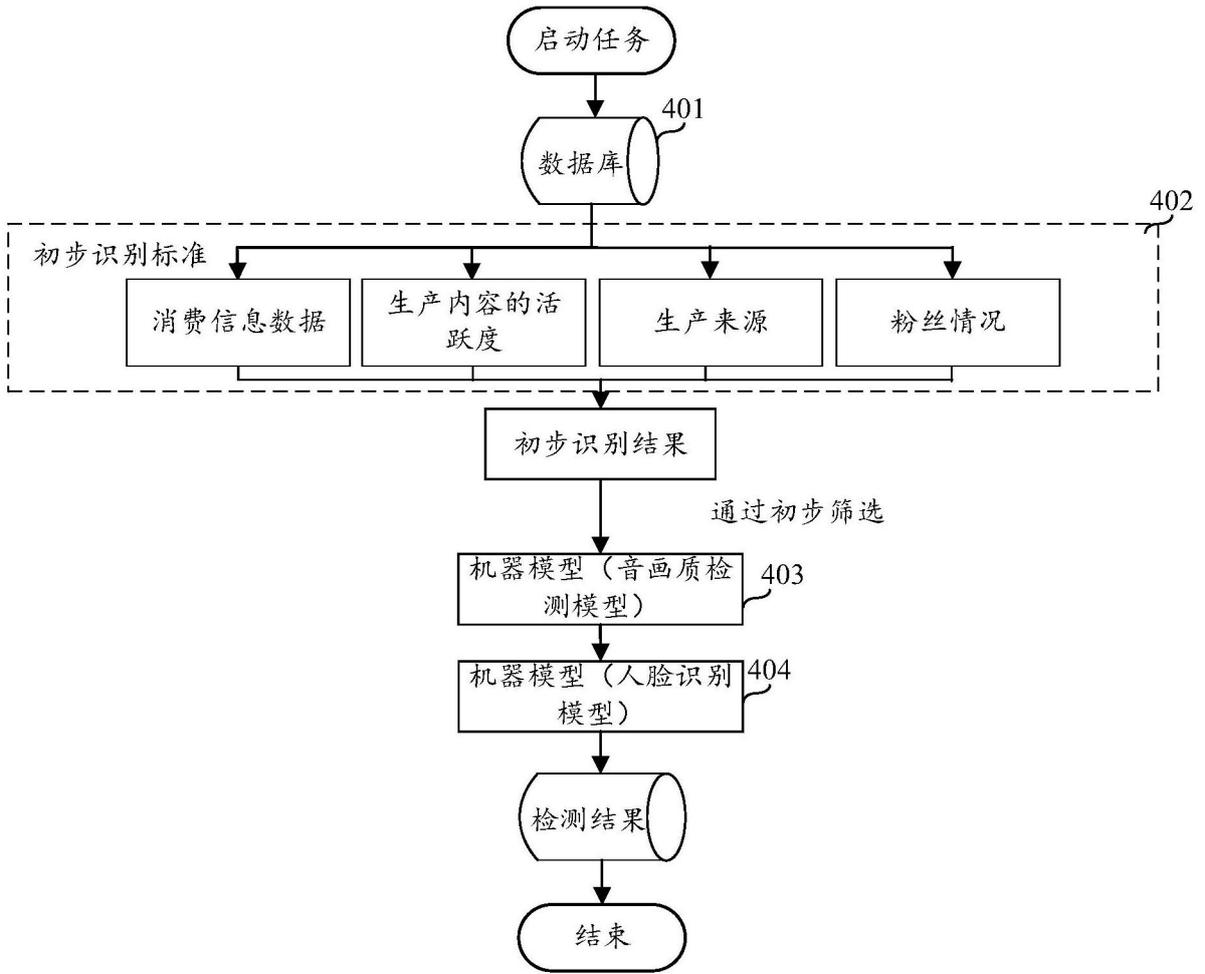


图4

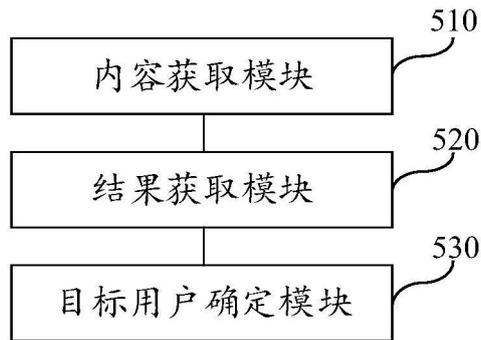


图5

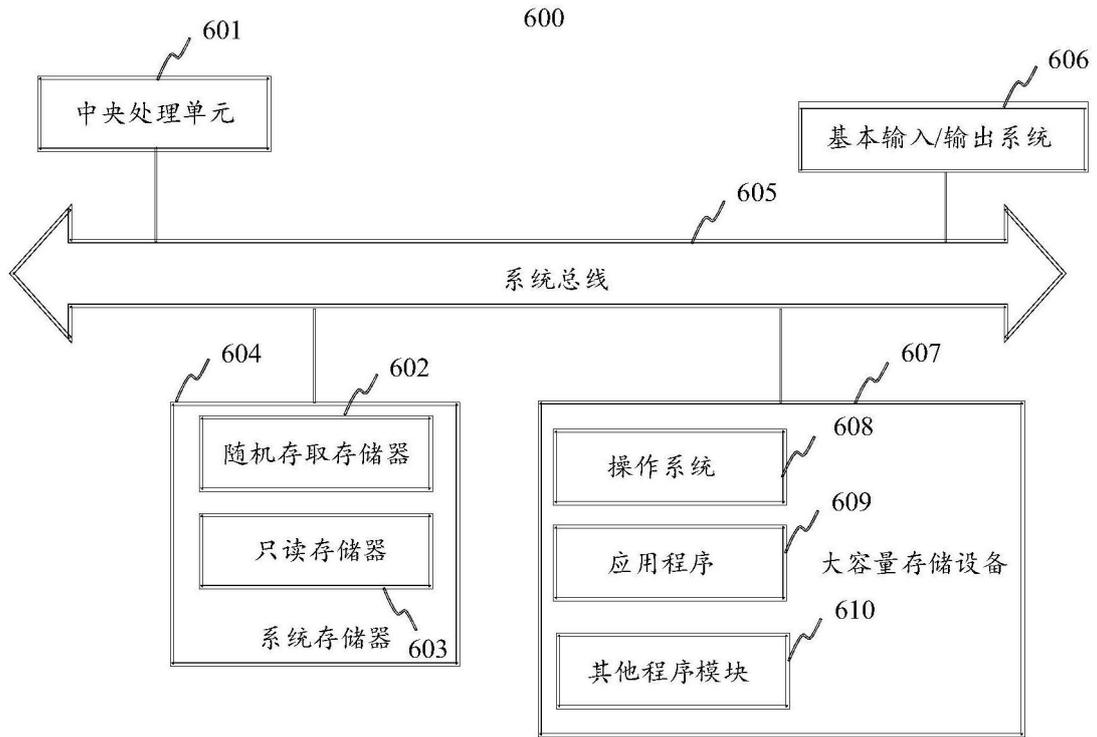


图6