

一种音频检测方法、装置、设备及存储介质

本申请要求在2018年10月10日提交中国专利局、申请号为201811178750.2的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机网络技术领域，例如一种音频检测方法、装置、设备及存储介质。

背景技术

随着计算机网络技术的快速发展，诸如个人计算机(Personal Computer, PC)、手机、平板电脑等设备越来越普及，给人们的生活、学习及工作带来了极大的便利。

作为设备的一个应用，用户可以使用设备，通过网络与其他用户进行语音沟通，如可以使用设备中所安装的带有语音聊天功能的软件，通过网络与其他用户进行语音聊天，也可以通过加入特定聊天室或聊天群参与多人的语音聊天和娱乐。用户可以通过设备将其所需要发送的语音信息发送给网络，使得其他用户可以通过网络接收到该语音信息并播放，达到语音沟通的目的。在实际聊天环境中，尤其是多人聊天时，语音信息涉及的内容范围较广，其中可能包含令人不适的语音数据，诸如刺耳、高分贝、内容不恰当等，这些语音通常是由个别用户恶意发出旨在干扰其他用户的正常使用，因而软件运营方将这类语音列为违规语音数据。

为了打击违规语音数据，保障用户的正常使用体验，避免用户流失而影响商业运营，软件运营方做了很多努力和探索，但收效有限。相关技术中，经常采用两种方案，一种是在软件上配置有举报入口，供正常用户举报违规用户，软件平台根据举报的线索作相应处理和惩罚；另一种是在平台侧部署人力，通过人工抽查或监控处理违规语音。对于具有大量活跃用户的软件平台，同一时间内经常同时并存数目极大的聊天室，各种违规语音数据很可能会大量随机出现，由此可知，上述两种方案均难以有效制止同一时间内随机出现的大量违规语音数据，也整体上很难避免违规语音影响用户正常体验的情况，并且偏向事后或者事情发生到一定程度后才介入，因此存在时间滞后，并且投入代价大。

发明内容

本申请实施例提供一种音频检测方法、系统、设备以及存储介质，通过自动检测语音违规行为，以避免传统基于人工检测违规语音数据的方法中所存在的检测时间滞后的情况。

第一方面，本申请实施例提供了一种音频检测方法，包括：获取音频文件数据；确定音频文件数据对应的属性检测数据；通过预先训练的全连接网络模型，对属性检测数据进行语音违规行为检测，生成音频文件数据对应的语音行为检测结果。

第二方面，本申请实施例还提供了一种音频检测装置，包括：音频文件数据获取模块，设置为获取音频文件数据；属性检测数据确定模块，设置为确定音频文件数据对应的属性检测数据；语音违规行为检测模块，设置为通过预先训练的全连接网络模型，对属性检测数据进行语音违规行为检测，生成音频文件数据对应的语音行为检测结果。

第三方面，本申请实施例还提供了一种设备，包括：处理器和存储器；所述存储器中存储有至少一条指令，所述指令由所述处理器执行，使得所述设备执行如第一方面所述的音频检测方法。

第四方面，本申请实施例还提供了一种计算机可读存储介质，所述存储介质中的指令由设备的处理器执行时，使得所述设备能够执行如第一方面所述的音频检测方法。

附图说明

图1是本申请一实施例中的一种音频检测方法的步骤流程示意图；

图2是本申请一实施例中的音频文件数据的检测流程示意图；

图3是本申请一实施例中的一种音频检测装置实施例的结构方框示意图；

图4是本申请一实施例中的一种设备的结构方框示意图。

具体实施方式

参照图1，示出了本申请的一实施例中的一种音频检测方法的步骤流程示意图，可以包括步骤110至步骤130。

在步骤110中，获取音频文件数据。

示例性的，本申请实施例在语音违规行为检测过程中，可以获取当前所需要检测的音频文件数据，以检测当前获取到音频文件数据是否包含语音违规行为对应的违规语音数据。其中，音频文件数据可以用于表征音频文件所包含的数据，如可以是声音文件中的语音数据。需要说明的是，声音文件可以是设备依据用户讲话声音生成的，可以用于承载表征用户所需要发送的语音数据，可以包括：语音违规行为对应的违规语音数据、符合预设语音行为规定的语音数据（即合规语音数据）等，本申请实施例对此不作限制。

在一实施例中，用户的讲话声音形成语音数据，该语音数据可以通过软件接口输入到软件中，使得软件可以将该语音数据确定为音频文件数据，即软件可以获取到音频文件数据，从而使得软件可以对该音频文件数据自动进行音频检测；或者，软件可以通过网络，将获取到的音频文件数据传输给软件平台，使得软件平台可以获取到该音频文件数据，进而可以通过软件平台对该音频文件数据进行音频检测。

在步骤 120 中，确定所述音频文件数据对应的属性检测数据。

本申请一实施例在获取到音频文件数据后，可以对该音频文件数据进行数据处理，确定音频文件数据对应的属性检测数据，以便后续可以对该属性检测数据进行语音违规行为检测。其中，属性检测数据可以包括以下至少一项：用户等级数据、分类概率数据和声纹特征数据等，本申请实施例对此不作限制。

需要说明的是，用户等级数据、分类概率数据和声纹特征数据等均可以预先获得并存储于预定位置中，以供步骤 120 执行时从预定位置中读取进行比对或确定。用户等级数据可以用于确定目标用户的用户等级，且用户等级可以依据目标用户的消费习惯、历史登录行为预先确定，如可以通过对目标用户的历史充值记录和历史登录行为数据进行归一化处理，确定出该目标用户的用户等级。其中，历史登录行为数据可以用于表征用户的历史登录行为，如作为历史登录行为数据的历史登录时长数据可以用于表征用户历史登录的时长；目标用户可以是指发送音频文件数据的用户，如可以是在语音沟通过程中，发送语音信息的用户等。目标用户的用户等级越高，可以表征该目标用户发送违规语音数据的可能性越小。

分类概率数据可以用于表征语音违规行为对应的分类概率，如语音违规概率、语音合规概率等。其中，语音违规概率可以是指：音频文件数据中所包含的语音数据是违规语音数据的概率；语音合规概率可以是指：音频文件数据中

所包含的语音数据是合规语音数据的概率。在一实施例中，语音违规概率的大小可以参照以下因素：刺耳、高分贝、音频内容信息等。例如，在音频文件数据含有恶意发出的刺耳、高分贝、内容不恰当等令人不适的语音数据的情况下，会增大该音频文件数据对应的语音违规概率，即音频含有用户恶意发出的刺耳、高分贝、内容不恰当等令人不适的信息，会导致该音频的违规概率增大。

声纹特征数据可以用于表征音频文件数据对应的声纹特征。该声纹特征可以是指用户的声音纹理特征。声音纹理特征可以是指：将用户的原始音频波形经过傅里叶变换及其它后处理生成的声音在时域上的频域特征。

在一实施例中，当获取到音频文件数据后，根据该音频文件数据对应的用户信息，查找或者获取该用户对应的用户等级数据、分类概率数据或者声纹特征数据等，这些数据的至少一种作为该音频文件数据的属性检测数据。因此，在本申请的一实施例中，上述步骤 120 确定的属性检测数据可以包括以下至少两项：用户等级数据、分类概率数据和声纹特征数据。其中，所述用户等级用于表征用户等级，所述分类概率数据用于表征语音违规行为对应的分类概率，声纹特征数据用于表征音频文件数据对应的声纹特征。示例性的，属性检测数据同时包含用户等级数据、分类概率数据和声纹特征数据，以便语音行为检测结果更加准确。

在步骤 130 中，通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。

本申请实施例中，在确定出音频文件数据对应的属性检测数据后，可以将该属性检测数据作为预先训练的全连接网络模型的输入，随后可将属性检测数据输入到预先训练的全连接网络模型中，以通过该全连接网络模型进行语音违规行为，生成音频文件数据对应的语音行为检测结果。在语音行为检测结果为语音违规行为检测结果的情况下，可以确定音频文件数据包含违规语音数据，即检测出语音违规行为对应的违规语音数据。在语音行为检测结果不是语音违规行为检测结果的情况下，可以确定音频文件数据不包含违规语音数据，在语音行为检测结果为语音正常行为检测结果的情况下，可以确定音频文件数据未含有语音违规行为对应的违规语音数据。

综上，本申请实施例在获取到音频文件数据后，可通过确定音频文件数据对应的属性检测数据，来进行语音违规行为检测，从而能够及时检测出语音违规行为对应的违规语音数据，确保用户的正常使用，避免了相关基于用户举报

和人工抽查导致语音违规行为检测的时间滞后的情况，保障用户的正常使用体验，满足用户需求，投入代价小。

在此基础上，当检测到用户拟发出的语音违规后，限制或暂停该用户的语音数据传输给其他用户，在一实施例中，根据该用户的音频文件数据判断出该用户违规时，给该用户的客户端发送相应信息，使该客户端暂停用户的语音聊天功能。

本申请一实施例在获取到音频文件数据后，可以对该音频文件数据进行切片处理，从而可以得到切片处理后的至少两帧的音频时域信息；随后，可以对得到的至少两帧的音频时域信息进行特征提取，得到音频特征数据，以对该音频特征数据确定出音频文件数据对应的属性检测数据。其中，音频特征数据可以用于表征音频特征，如表征音频的振幅特征、声纹特征等。

例如，在音频特征数据为振幅谱（Magnitude Spectrum, Mags）特征数据的情况下，可以依据该振幅谱特征数据生成音频文件数据对应的分类概率数据，以便后续可以对该分类概率数据进行语音违规行为检测。振幅谱（Magnitude Spectrum, Mags）特征数据可以用于表征音频的 Mags 特征。需要说明的是，申请人在实现本申请实施例过程中，经过深入分析和实验发现，Mags 特征在声音违规检测中，效果良好，因而如后文所述，本申请实施例的一些方案结合 Mags 特征作了相应扩展。

在本申请的一实施例中，上述确定所述音频文件数据对应的属性检测数据的步骤，可以包括：对所述音频文件数据进行切片处理，得到至少两帧音频时域信息；对所述至少两帧音频时域信息进行特征提取，得到振幅谱特征数据和声纹特征数据；对所述振幅谱特征数据和所述声纹特征数据进行拼接，生成特征向量数据；通过预先训练的语音分类模型，对所述特征向量数据进行语音分类处理，得到作为所述属性检测数据的分类概率数据。其中，特征提取可以包括：振幅谱特征提取、声纹特征提取等，本申请实施例对此不作限制。

例如，在语音通讯过程中，可以获取待发送的音频文件数据，并可利用预设的移动窗口对获取到的音频文件数据进行切片，得到多帧的音频时域信息。从而，可以对至少两帧的音频时域信息进行 Mags 特征提取，得到 Mags 特征数据；并且可以对至少两帧的音频时域信息进行声纹特征提取，得到所述音频文件数据对应的声纹特征数据；随后，可以将得到的 Mags 特征数据与到声纹特征数据进行拼接，生成一维的特征向量数据，并将该特征向量数据作为语音分类

模型的输入，通过该语音分类模型进行语音分类处理，得到所述音频文件数据对应的分类概率数据。在一实施例中，语音分类模型可以根据不同的音频输入提取音频的特征，即根据输入的特征向量数据进行特征提取，得到输入特征；随后可根据输入特征的分布，分配该音频输入一个概率值，即为输入的特征向量数据分配一个分类概率数据，并输出，以作为语音违规行为检测的属性检测数据。在输入特征与预设的违规样本的特征相似的情况下，语音分类模型会分配给输入的特征向量数据一个高违规概率，例如输出 90% 作为特征向量数据的语音违规概率。在输入特征与预设的正常样本的特征相似的情况下，语音分类模型会分配给输入的特征向量数据一个低违规概率，例如输出 1% 作为特征向量数据的语音违规概率，即特征向量数据的语音合规概率为 99%。

在一实施例中，对所述至少两帧音频时域信息进行特征提取，得到振幅谱特征数据，可以包括：对所述至少两帧音频时域信息进行频域变换，得到音频频域信息；基于所述音频频域信息进行振幅谱特征提取，得到振幅谱特征数据。其中，频域变换可以包括傅里叶变换，如快速傅氏变换（Fast Fourier Transformation, FFT）等，本申请实施例对此不作限制。

作为本申请的一个示例，在利用预设的移动窗口将音频文件数据切成至少两个小段后，可以将每一小段称为一帧音频时域信息；随后，可以对得到的每帧音频时域信息进行傅里叶变换，得到每帧音频时域信息对应的音频频域信息，可以基于该音频频域信息进行振幅谱特征提取，得到所述音频文件数据对应的振幅谱特征数据，如对音频频域信息的 Mags 取均值和方差，然后将取到的均值和方差作为振幅谱特征数据，以基于振幅谱特征数据生成特征向量数据，从而可以对该特征向量数据进行语音分类处理，得到所述音频文件数据对应的分类概率数据。其中，相邻的两帧音频时域信息可以有交叠部分，即帧与帧之间可以有交叠，如在一帧音频时域信息的帧长为 25 毫秒（ms），帧移为 10 毫秒时，两帧之间可以有 15 毫秒的交叠。需要说明的是，帧长和帧移可以依据精确度要求进行设置，本示例对此不作限制。

当然，也可以将得到的音频特征数据作为属性检测数据，以对该音频特征数据进行语音违规行为检测，如在音频特征数据为声纹特征数据的情况下，可以将该声纹特征数据作为属性检测数据，以对该音频特征数据进行语音违规行为检测，本申请实施例对此不作限制。

在本申请的一实施例中，上述确定所述音频文件数据对应的属性检测数据

的步骤，可以包括：对所述音频文件数据进行切片处理，得到至少两帧音频时域信息；对所述至少两帧音频时域信息进行频域变换，得到音频频域信息；对所述音频频域信息进行均值处理，得到第一定长数据；基于所述第一定长数据，通过预先训练的神经网络模型进行声纹特征提取，得到作为所述属性检测数据的声纹特征数据。例如，在获取到待检测的音频文件数据后，可以利用移动窗口对该音频文件数据进行切片处理，得到至少两帧音频时域信息，随后对得到的至少两帧音频时域信息进行 FFT 变换，得到音频频域信息，对该音频频域信息进行均值处理，如对该音频频域信息取均值，得到定长的数据，并可将得到的定长的数据确定为第一定长数据，以将该第一定长数据输入到预先训练的神经网络模型中进行声纹特征提取，得到声纹特征数据，然后可将该声纹特征数据作为音频文件数据对应的属性检测数据，输入到全连接网络模型中，以对声纹特征数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。其中，待检测的音频文件数据可以包括待传输的音频文件数据、待播放的音频文件数据等；待传输的音频文件数据可以用于表征语音通讯过程中待发送的音频文件数据；待播放的音频文件数据可以用于表征待播放语音的文件数据。

本申请一实施例可以基于用户消费习惯对用户进行分级，以基于分级后得到的用户级别进行语音违规行为检测，以预测出语音违规行为。在一实施例中，本申请实施例提供的音频检测方法还可以包括：获取目标用户的历史行为数据；根据所述历史行为数据得到作为所述属性检测数据的用户等级数据。其中，所述历史行为数据包括以下至少一项：历史登录数据、用户消费行为数据、违规历史数据、用户消费行为数据和充值历史数据等。

示例性的，在需要确定某一个用户的用户等级时，可以将该用户确定为目标用户，并可以根据该目标用户的用户标识，如用户账号、用户名等，从数据库中获取该目标用户的历史行为数据，以根据获取到的历史行为数据确定出该目标用户的用户等级，即确定出目标用户的用户等级数据。随后可以将用户等级数据存储到的数据库中，以在音频检测时从该数据库中获取到该用户等级数据作为属性检测数据。

当然，本申请实施例也可在音频检测时，针对获取到的音频文件数据，获取该音频文件数据对应目标用户的历史行为数据，以根据获取到的历史行为数据实时确定出作为属性检测数据的用户等级数据，从而可以采用实时确定出的

用户等级数据进行语音违规行为检测,提高语音违规行为检测的准确性。例如,上述确定所述音频文件数据对应的属性检测数据的步骤,包括:针对所述音频文件数据,获取目标用户的历史行为数据;对所述历史行为数据进行归一化处理,得到作为所述属性检测数据的用户等级数据。

作为本申请的一个示例,在获取到音频文件数据后,可以将发送该音频文件数据的用户确定为目标用户,随后可针对该音频文件数据,依据目标用户的用户标识,获取该目标用户的历史行为数据,如用户消费行为数据,历史登录数据、违规历史数据和充值历史数据等其中至少一项数据。其中,用户消费行为数据可以用于确定目标用户的消费行为习惯信息;违规历史数据可以用于确定目标用户的语音违规历史信息,如确定目标用户是否有违规历史,或者确定目标用户的违规历史次数等;充值历史数据可以用于确定目标用户的充值历史信息,如目标用户的充值次数,历史充值金额等;用户历史登录数据可以用于确定目标用户的历史登录行为,包括:登录次数、登录时长、登录地址等。登录次数可以用于表征目标用户的登录数量;登录时长可以用于表征目标用户的历史登录时长,如可以包括目标用户每一次登录对应的登录时长;登录地址可以用于确定用户每一次登录的地址,如可以是目标用户登录时所使用的设备的互联网协议(Internet Protocol, IP)地址、媒体访问控制(Medium Access Control, MAC)地址等,本示例对此不作限制。

在获取到目标用户的历史行为数据后,可对获取到的历史行为数据进行归一化处理,如对目标用户的目标用户的登陆数量、登陆时长、是否有违规历史、充值历史等信息进行数值化和归一化,从而可以基于归一化处理结果确定出目标用户的用户等级数据,随后可将该用户等级数据作为属性检测数据,输入到预先训练的全连接网络模型中进行语音违规行为检测,生成语音行为检测结果。

本申请一实施例中,通过预先训练的全连接网络模型,对所述属性检测数据进行语音违规行为检测,生成所述音频文件数据对应的语音行为检测结果,包括:将所述属性检测数据输入到所述全连接网络模型进行检测;将所述全连接网络模型输出的检测结果作为所述语音行为检测结果。

作为本申请的一实施例,在获取到音频文件数据后,如图2所示,可以基于诸如卷积神经网络(Convolutional Neural Network, CNN)进行声音纹理提取,得到该音频文件数据对应的声纹特征数据;并可以基于Mags特征进行声音分类,即采用音频文件数据对应的Mags特征数据生成特征向量数,并将该特征向量数

输入到语音分类模型进行语音分类处理，得到所述音频文件数据对应的分类概率数据；以及可以针对当前获取到的音频文件数据，基于用户消费习惯对应的消费行为习惯信息进行用户等级预测，确定出用户等级数据。随后，可以基于用户等级数据、分类概率数据以及声纹特征数据进行融合检测，即将用户等级数据、分类概率数据以及声纹特征数据这三种属性检测数据输入到预先训练好的全连接网络模型，以通过全连接网络模型融合用户等级数据、分类概率数据以及声纹特征数据这三种属性检测数据，来进行语音违规行为检测，得到全连接网络模型输出的检测结果，然后将全连接网络模型输出的检测结果作为语音行为检测结果，从而可以基于该语音行为检测结果确定出音频文件数据是否包含违规语音数据，实现了语音违规行为的预测，避免了相关技术中语音违规行为的检测时间滞后的情况。

本申请一实施例提供的音频检测方法还可以包括：在所述语音为检测结果为语音违规行为检测结果的情况下，确定所述音频文件数据包含违规语音数据；禁止传输或播放所述违规语音数据。

示例性的，在语音通讯过程中，可以获取目标用户当前所需要发送的音频文件数据进行语音违规行为检测，以判断该目标用户发送的音频文件数据是否包含有语音违规行为对应的违规语音数据。语音行为检测结果分为语音违规行为检测结果和语音正常行为检测结果，在全连接网络模型输出的语音行为检测结果为语音正常行为检测结果的情况下，可以确定当前获取到的音频文件数据不包含违规语音数据，随后可以基于语音正常行为检测结果发送该音频文件数据，使得与该目标用户进行语音通讯的其他用户可以接收到该音频文件数据并播放，达到语音通讯的目的；在全连接网络模型输出的语音行为检测结果为语音违规行为检测结果的情况下，可以确定当前获取到的音频文件数据包含违规语音数据，随后可以基于语音违规行为检测结果，禁止该音频文件数据的发送，如驳回该音频文件数据对应的发送请求，以禁止传输该音频文件数据中所包含的违规语音数据，从而可以避免违规语音数据所带来的负面影响，确保用户的正常使用。

当然，本申请实施例中获取到的音频文件数据还可以是其他音频文件数据，如可以是待播放的音频文件数据等。在检测到待播放的音频文件数据对应的语音行为检测结果为语音违规行为检测结果的情况下，可以基于语音违规行为检测结果，禁止该音频文件数据的播放，示例性的，在软件在检测到待播放的音

音频文件数据对应的语音行为检测结果为语音违规行为检测结果的情况下，可以丢弃或忽略该音频文件数据，即不对该音频文件数据进行播放，以禁止播放该音频文件数据中所包含的违规语音数据；在检测到待播放的音频文件数据对应的语音行为检测结果为语音正常行为检测结果后，可以基于语音正常行为检测结果对音频文件数据进行播放等。

此外，本申请实施例在确定出音频文件数据包含违规语音数据后，可以屏蔽该违规语音数据对应用户的语音输入。在一实施例中，用户可以在使用软件过程中，可以通过该软件的语音输入接口进行语音输入，使得软件 and 该软件对应的软件平台中至少之一可以获取到其所输入的语音数据，从而可以基于获取到的语音数据形成音频文件数据，随后可基于该音频文件数据进行音频检测，以确定语音数据是否是违规语音数据。在软件或者软件平台检测到音频文件数据对应的语音行为检测结果为语音违规行为检测结果的情况下，可以确定音频文件数据包含违规语音数据，即确定该用户输入的语音数据为违规语音数据，然后可以针对该用户关闭软件的语音输入接口，使得该用户不可以通过该软件的语音输入接口进行语音输入，以屏蔽该用户的语音输入。当然，软件和软件平台中至少之一也可以采用其他方式来屏蔽用户的语音输入，如可以通过关闭软件的语音输入功能来实现语音输入的屏蔽等，本申请实施例对此不作限制。

本申请一实施例在确定出音频文件数据对应的属性检测数据后，可以将该属性检测数据存储到一个训练集中，作为待训练属性检测数据，以便在训练全连接网络模型时可以从训练集中获取到该待训练属性检测数据进行训练。示例性的，本申请实施例提供的音频检测方法还可以包括：获取待训练属性检测数据；对所述待训练属性检测数据进行训练，得到全连接网络模型。其中，所述待训练属性检测数据包括从训练集获取到的各种属性检测数据，如用户等级数据、分类概率数据和声纹特征数据等。

例如，在确定出音频文件数据对应的用户等级数据、分类概率数据和声纹特征数据后，可以将用户等级数据、分类概率数据以及声纹特征数据作为全连接网络模型的训练数据，即将用户等级数据、分类概率数据以及声纹特征数据作为待训练属性检测数据，然后可按照预设的全连接网络结构，采用分类概率数据、用户等级数据以及声纹特征数据进行模型训练，得到全连接网络模型。该全连接网络模型可以用于对输入的属性检测数据进行语音违规行为检测，输出语音行为检测结果。该语音行为检测结果可以用于判断是否存在语音违规行

为，以确定音频文件数据是否包含语音违规行为对应的违规语音数据。

当然，本申请实施例也可以将音频文件数据作为训练数据，以基于该音频文件数据进行训练，得到相应的网络模型。该网络模型可以用于确定音频文件数据对应的属性检测数据，包括：神经网络模型和语音分类模型等，本申请实施例对此不作限制。其中，神经网络模型可以用于确定音频文件数据对应的声纹特征数据；语音分类模型可以用于确定音频文件数据对应的分类概率数据。

在本申请的一实施例中，上述音频检测方法还可以包括：从预设的训练集中，获取待训练音频文件数据；采用预设的移动窗口，对所述待训练音频文件数据进行切片，得到帧时域信息；对所述帧时域信息进行频域变换，得到帧频域信息；对所述帧频域信息进行均值处理，得到第二定长数据；基于所述第二定长数据和所述音频文件数据对应的标签数据，按照预设的神经网络算法进行训练，得到所述神经网络模型。其中，频域变换可以包括傅里叶变换、快速傅里叶变换等。

本申请一实施例可以预先将需要进行训练的音频文件数据存储到训练集中，并可将训练集中存储的音频文件数据作为待训练音频文件数据。在模型训练过程中，可以从该训练集中获取待训练音频文件数据，然后可采用预设的移动窗口对该待训练音频文件数据进行切片，得到至少两帧的时域信息，即得到帧时域信息，随后可对帧时域信息进行频域变换，如对至少两帧的时域信息进行FFT变换，得到帧频域信息；可以对该帧频域信息进行均值处理，如对帧频域信息取均值，得到定长的数据，并可以将该数据作为第二定长数据。

此外，可以为待训练音频数据设置对应的标签数据，从而采用该标签数据和第二定长数据，按照预设的神经网络算法进行训练，如按照预设的CNN算法进行网络训练，直到网络收敛。在网络收敛的情况下，可以基于训练得到的网络参数构建对应的神经网络模型，以便后续可以通过该神经网络模型进行声纹特征提取。其中，神经网络模型可以包括：网络参数和至少两个网络层，如卷积层、全连接层等，本申请实施例对此不作限制。

例如，在神经网络模型的训练过程中，可以将得到的第二定长数据和音频文件数据对应的标签数据输入到CNN模型中，以训练该CNN模型的网络参数，直到网络收敛。其中，标签数据可以用于标记待训练音频数据是否包含语音违规行为对应的违规语音数据。

本申请一实施例在训练过程中，可以基于提取到的声纹特征数据进行训练，

以训练出语音分类模型。示例性的，上述音频检测方法还可以包括：采用预设的移动窗口，对获取到的待训练音频文件数据进行切片，得到帧时域信息；对所述帧时域信息进行特征提取，得到振幅谱特征训练数据和声纹特征训练数据，其中，所述特征提取包括：振幅谱特征提取和声纹特征提取；对所述振幅谱特征训练数据进行均值处理，得到第三定长数据；对所述振幅谱特征训练数据和声纹特征训练数据进行拼接，生成特征向量训练数据；对第三定长数据和所述特征向量训练数据进行训练，得到所述语音分类模型。

本申请一实施例在获取到待训练音频文件数据后，可以预设的移动窗口对该待训练音频文件数据进行切片，得到帧时域信息，然后可对该帧时域信息进行振幅谱特征提取和声纹特征提取，得到振幅谱特征训练数据和声纹特征训练数据。例如，可以通过对帧时域信息进行 FFT 变换，得到帧频域信息；然后，可对该帧频域信息进行振幅谱特征，得到待训练音频文件数据对应的振幅谱特征训练数据，并可基于该帧频域信息进行声纹特征提取，得到待训练音频文件数据对应的声纹特征训练数据。

随后，对振幅谱特征训练数据和声纹特征训练数据进行拼接，形成特征向量训练数据，并可将该特征向量训练数据作为语音分类模型的训练数据，以采用特征向量训练数据进行语音分类模型训练。例如，在声纹特征训练数据是 1 维向量 (1, 1024)，Mags 特征训练数据是 1 维向量 (1, 512) 时，可以通过将这两个向量拼接在一起，组成 1 维特征向量 (1, 1536)，并可以将该 1 维特征向量 (1, 1536) 作为特征向量训练数据，输入到预设的全连接网络进行训练，从而可以训练出一个 2 层全连接网络模型，进而可以将训练出的这个 2 层全连接网络模型作为语音分类模型，以便后续可以采用该语音分类模型进行语音分类处理。

综上，本申请实施例可以通过神经网络模型进行声纹特征提取，得到音频文件数据对应的声纹特征数据，并可通过语音分类模型进行语音分类处理，得到音频文件数据对应的分类概率数据，通过对目标用户的历史行为数据进行归一化处理，确定用户等级数据，从而可以依据声纹特征数据、分类概率数据以及用户等级数据进行语音违规行为检测，即融合音频文件数据对应的多种属性检测数据进行语音违规行为检测，能够有效避免相关技术基于人工检测语音违规行为中所存在的时间滞后、代价大等情况，减少语音违规行为检测的投入代价，提高语音违规行为检测的准确度。

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请实施例并不受所描述的动作顺序的限制，因为依据本申请实施例，某些步骤可以采用其他顺序或者同时进行。

参照图 3，示出了本申请一实施例中的一种音频检测装置实施例的结构方框示意图，该音频检测装置包括音频文件数据获取模块 310、属性检测数据确定模块 320 以及语音违规行为检测模块 330。

音频文件数据获取模块 310，设置为获取音频文件数据。

属性检测数据确定模块 320，设置为确定所述音频文件数据对应的属性检测数据。

语音违规行为检测模块 330，设置为通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。

在本申请的一实施例中，属性检测数据确定模块 320 可以包括切片处理子模块、特征提取子模块、数据拼接子模块以及分类处理子模块。

切片处理子模块，设置为对所述音频文件数据进行切片处理，得到至少两帧音频时域信息。

特征提取子模块，设置为对所述至少两帧音频时域信息进行特征提取，得到所述音频文件数据对应的振幅谱特征数据和声纹特征数据。

数据拼接子模块，设置为对所述振幅谱特征数据和所述声纹特征数据进行拼接，生成特征向量数据。

分类处理子模块，设置为通过预先训练的语音分类模型，对所述特征向量数据进行语音分类处理，得到作为所述属性检测数据的分类概率数据。

在一实施例中，所述特征提取子模块包括频域变换单元以及振幅谱特征提取单元。

频域变换单元，设置为对所述至少两帧音频时域信息进行频域变换，得到音频频域信息。

振幅谱特征提取单元，设置为基于所述音频频域信息进行振幅谱特征提取，得到所述音频文件数据对应的振幅谱特征数据。

在一实施例中，分类处理子模块包括均值处理单元以及分类处理单元。

均值处理单元，设置为对所述振幅谱特征数据进行均值处理，得到定长数

据。

分类处理单元, 设置为基于所述定长数据, 通过预先训练的神经网络模型, 对所述特征向量数据进行语音分类处理, 得到所述音频文件数据对应的分类概率数据。

在本申请的一实施例中, 属性检测数据确定模块 320 包括切片处理子模块、频域变换子模块、频域均值处理子模块以及声纹特征提取子模块。

切片处理子模块, 设置为对所述音频文件数据进行切片处理, 得到至少两帧音频时域信息。

频域变换子模块, 设置为对所述至少两帧音频时域信息进行频域变换, 得到音频频域信息。

频域均值处理子模块, 设置为对所述音频频域信息进行均值处理, 得到第一定长数据。

声纹特征提取子模块, 设置为基于所述第一定长数据, 通过预先训练的神经网络模型进行声纹特征提取, 得到所述音频文件数据对应作为所述属性检测数据的声纹特征数据。

在本申请的一实施例中, 音频检测装置还可以包括历史行为数据获取模块以及户等级数据确定模块。

历史行为数据获取模块, 设置为获取目标用户的历史行为数据。

用户等级数据确定模块, 设置为根据所述历史行为数据得到作为所述属性检测数据的用户等级数据。其中, 所述历史行为数据包括以下至少一项: 历史登录数据、用户消费行为数据、违规历史数据和充值历史数据。

本申请一实施例中, 属性检测数据确定模块 320 可以包括行为数据获取子模块以及归一化处理子模块。

行为数据获取子模块, 设置为针对所述音频文件数据, 获取目标用户的历史行为数据。

归一化处理子模块, 设置为对所述历史行为数据进行归一化处理, 确定所述目标用户的用户等级数据。

本申请实施例中, 上述历史登录数据包括: 登录次数、登录时长、登录地址等等。示例性的, 上述属性检测数据可以包括以下至少两项: 用户等级数据、分类概率数据和声纹特征数据, 所述用户等级数据用于表征用户等级, 所述分类概率数据用于表征语音违规行为对应的分类概率, 声纹特征数据用于表征音

频文件数据对应的声纹特征。

在本申请的一实施例中，语音违规行为检测模块 330 包括输入子模块以及输出子模块。

输入子模块，设置为将所述属性检测数据输入到所述全连接网络模型进行检测。

输出子模块，设置为将所述全连接网络模型输出的检测结果作为所述语音行为检测结果。

在一实施例中，该音频检测装置还可以包括违规语音数据确定模块、禁止传输模块、静止播放模块以及语音输入屏蔽模块。

违规语音数据确定模块，设置为在所述语音为检测结果为语音违规行为检测结果的情况下，确定所述音频文件数据包含违规语音数据。

禁止传输模块，设置为禁止传输所述违规语音数据。

静止播放模块，设置为禁止播放违规语音数据。

语音输入屏蔽模块，设置为屏蔽所述违规语音数据对应用户的语音输入。

在以实施例中，该音频检测装置还可以包括训练数据获取模块、切片模块、频域变换模块、均值处理模块以及神经网络训练模块。

训练数据获取模块，设置为从预设的训练集中，获取待训练音频文件数据。

切片模块，设置为采用预设的移动窗口，对所述待训练音频文件数据进行切片，得到帧时域信息。

频域变换模块，设置为对所述帧时域信息进行频域变换，得到帧频域信息。

均值处理模块，设置为对所述帧频域信息进行均值处理，得到第二定长数据。

神经网络训练模块，设置为基于所述第二定长数据和所述音频文件数据对应的标签数据，按照预设的神经网络算法进行训练，得到所述神经网络模型。

在本申请的一实施例中，音频检测装置还可以包括切片模块、特征提取模块、均值处理模块、训练数据拼接模块以及语音分类模型训练模块。

切片模块，设置为采用预设的移动窗口，对获取到的待训练音频文件数据进行切片，得到帧时域信息。

特征提取模块，设置为对所述帧时域信息进行特征提取，得到振幅谱特征训练数据和声纹特征训练数据，其中，所述特征提取包括：振幅谱特征提取和声纹特征提取。

均值处理模块，设置为对所述振幅谱特征训练数据进行均值处理，得到第三定长数据。

训练数据拼接模块，设置为对所述振幅谱特征训练数据和声纹特征训练数据进行拼接，生成特征向量训练数据。

语音分类模型训练模块，设置为对第三定长数据和所述特征向量训练数据进行训练，得到所述语音分类模型。

在一实施例中，该音频检测装置还可以包括：全连接网络模型训练模块。该全连接网络模型训练模块，设置为获取待训练属性检测数据；对所述待训练属性检测数据进行训练，得到全连接网络模型。其中，所述属性检测数据包括以下至少一项：用户等级数据、分类概率数据和声纹特征数据等，本申请实施例对此不作限制。

需要说明的是，上述提供的音频检测装置可执行本申请任意实施例所提供的音频检测方法。

在一实施例中，上述音频检测装置可以集成在设备中。该设备可以是至少两个物理实体构成，也可以是一个物理实体构成，如设备可以是PC、电脑、手机、平板设备、个人数字助理、服务器、消息收发设备、游戏控制台等。

本申请实施例还提供一种设备，包括：处理器和存储器。存储器中存储有至少一条指令，且指令由所述处理器执行，使得所述设备执行如上述方法实施例中所述的音频检测方法。

参照图4，示出了本申请一实施例中的一种设备的结构方框示意图。如图4所示，该设备包括：处理器40、存储器41、具有触摸功能的显示屏42、输入装置43、输出装置44以及通信装置45。该设备中处理器40的数量可以是至少一个，图4中以一个处理器40为例。该设备中存储器41的数量可以是至少一个，图4中以一个存储器41为例。该设备的处理器40、存储器41、显示屏42、输入装置43、输出装置44以及通信装置45可以通过总线或者其他方式连接，图4中以通过总线连接为例。

存储器41作为一种计算机可读存储介质，可设置为存储软件程序、计算机可执行程序以及模块，如本申请任意实施例所述的音频检测方法对应的程序指令/模块（例如，音频检测装置中的音频文件数据获取模块310、属性检测数据确定模块320以及语音违规行为检测模块330等）。存储器41可主要包括存储程序区和存储数据区，其中，存储程序区可存储操作装置、至少一个功能所需

的应用程序；存储数据区可存储根据设备的使用所创建的数据等。此外，存储器 41 可以包括高速随机存取存储器，还可以包括非易失性存储器，例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实例中，存储器 41 可包括相对于处理器 40 远程设置的存储器，这些远程存储器可以通过网络连接至设备。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

显示屏 42 为具有触摸功能的显示屏 42，其可以是电容屏、电磁屏或者红外屏。一般而言，显示屏 42 设置为根据处理器 40 的指示显示数据，还设置为接收作用于显示屏 42 的触摸操作，并将相应的信号发送至处理器 40 或其他装置。在一实施例中，在显示屏 42 为红外屏的情况下，其还包括红外触摸框，该红外触摸框设置在显示屏 42 的四周，该红外触摸框还设置为接收红外信号，并将该红外信号发送至处理器 40 或者其他设备。

通信装置 45，设置为与其他设备建立通信连接，其可以是有线通信装置和无线通信装置中至少一种。

输入装置 43 设置为接收输入的数字或者字符信息，以及产生与设备的用户设置以及功能控制有关的键信号输入，还设置为获取图像的摄像头以及获取音频数据的拾音设备。输出装置 44 可以包括扬声器等音频设备。需要说明的是，输入装置 43 和输出装置 44 的组成可以根据实际情况设定。

处理器 40 通过运行存储在存储器 41 中的软件程序、指令以及模块，从而执行设备的各种功能应用以及数据处理，即实现上述的音频检测方法。

在一实施例中，处理器 40 执行存储器 41 中存储的至少一个程序时，实现如下操作：获取音频文件数据；确定所述音频文件数据对应的属性检测数据；通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。

本申请实施例还提供一种计算机可读存储介质，所述存储介质中的指令由设备的处理器执行时，使得设备能够执行如上述方法实施例所述的音频检测方法。示例性的，该音频检测方法包括：获取音频文件数据；确定所述音频文件数据对应的属性检测数据；通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。

需要说明的是，对于装置、设备、存储介质实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即

可。

通过以上关于实施方式的描述，所属领域的技术人员可以清楚地了解到，本申请可借助软件及必需的通用硬件来实现，当然也可以通过硬件实现。基于这样的理解，本申请的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品可以存储在计算机可读存储介质中，如计算机的软盘、只读存储器（Read-Only Memory, ROM）、随机存取存储器（Random Access Memory, RAM）、闪存（FLASH）、硬盘或光盘等，包括多个指令用以使得一台计算机设备（可以是机器人，个人计算机，服务器，或者网络设备等等）执行本申请任意实施例所述的音频检测方法。

值得注意的是，上述音频检测装置中，所包括的每个单元和模块只是按照功能逻辑进行划分的，但并不局限于上述的划分，只要能够实现相应的功能即可；另外，每个功能单元的具体名称也只是为了便于相互区分，并不用于限制本申请的保护范围。

应当理解，本申请的每个部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行装置执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列（Programmable Gate Array, PGA），现场可编程门阵列（Field Programmable Gate Array, FPGA）等。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“一实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的至少一个实施例或示例中以合适的方式结合。

权利要求书

1、一种音频检测方法，包括：

获取音频文件数据；

确定所述音频文件数据对应的属性检测数据；

通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。

2、根据权利要求1所述的方法，其中所述确定所述音频文件数据对应的属性检测数据，包括：

对所述音频文件数据进行切片处理，得到至少两帧音频时域信息；

对所述至少两帧音频时域信息进行特征提取，得到振幅谱特征数据和声纹特征数据；

对所述振幅谱特征数据和所述声纹特征数据进行拼接，生成特征向量数据；

通过预先训练的语音分类模型，对所述特征向量数据进行语音分类处理，得到作为所述属性检测数据的分类概率数据。

3、根据权利要求2所述的方法，其中所述对所述至少两帧音频时域信息进行特征提取，得到振幅谱特征数据，包括：

对所述至少两帧音频时域信息进行频域变换，得到音频频域信息；

基于所述音频频域信息进行振幅谱特征提取，得到所述振幅谱特征数据。

4、根据权利要求1所述的方法，其中所述确定所述音频文件数据对应的属性检测数据，包括：

对所述音频文件数据进行切片处理，得到至少两帧音频时域信息；

对所述至少两帧音频时域信息进行频域变换，得到音频频域信息；

对所述音频频域信息进行均值处理，得到第一定长数据；

基于所述第一定长数据，通过预先训练的神经网络模型进行声纹特征提取，得到作为所述属性检测数据的声纹特征数据。

5、根据权利要求4所述的方法，还包括：

从预设的训练集中，获取待训练音频文件数据；

采用预设的移动窗口，对所述待训练音频文件数据进行切片，得到帧时域信息；

对所述帧时域信息进行频域变换，得到帧频域信息；

对所述帧频域信息进行均值处理，得到第二定长数据；

基于所述第二定长数据和所述音频文件数据对应的标签数据，按照预设的

神经网络算法进行训练，得到所述神经网络模型。

6、根据权利要求1所述的方法，还包括：

获取目标用户的历史行为数据，其中，所述历史行为数据包括以下至少一项：历史登录数据、用户消费行为数据、违规历史数据，和充值历史数据；

根据所述历史行为数据得到作为所述属性检测数据的用户等级数据。

7、根据权利要求1所述的方法，其中，所述属性检测数据包括以下至少两项：用户等级数据、分类概率数据，和声纹特征数据，所述用户等级数据用于表征用户等级，所述分类概率数据用于表征语音违规行为对应的分类概率，所述声纹特征数据用于表征音频文件数据对应的声纹特征。

8、根据权利要求1至7任一项所述的方法，其中，所述通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果，包括：

将所述属性检测数据输入到所述全连接网络模型进行检测；

将所述全连接网络模型输出的检测结果作为所述语音行为检测结果。

9、根据权利要求8所述的方法，还包括：

在所述语音行为检测结果为语音违规行为检测结果的情况下，确定所述音频文件数据包含违规语音数据；

禁止传输或播放所述违规语音数据；或者，屏蔽所述违规语音数据对应用户的语音输入。

10、根据权利要求2或3所述的方法，还包括：

采用预设的移动窗口，对获取到的待训练音频文件数据进行切片，得到帧时域信息；

对所述帧时域信息进行特征提取，得到振幅谱特征训练数据和声纹特征训练数据，其中，所述特征提取包括：振幅谱特征提取和声纹特征提取；

对所述振幅谱特征训练数据进行均值处理，得到第三定长数据；

对所述振幅谱特征训练数据和声纹特征训练数据进行拼接，生成特征向量训练数据；

对所述第三定长数据和所述特征向量训练数据进行训练，得到所述语音分类模型。

11、根据权利要求1至7任一项所述的方法，还包括：

获取待训练属性检测数据；

对所述待训练属性检测数据进行训练，得到全连接网络模型。

12、一种音频检测装置，包括：

音频文件数据获取模块，设置为获取音频文件数据；

属性检测数据确定模块，设置为确定所述音频文件数据对应的属性检测数据；

语音违规行为检测模块，设置为通过预先训练的全连接网络模型，对所述属性检测数据进行语音违规行为检测，生成所述音频文件数据对应的语音行为检测结果。

13、一种设备，包括：处理器和存储器；

所述存储器中存储有至少一条指令，所述指令由所述处理器执行，使得所述设备执行如权利要求 1 至 11 任一项所述的音频检测方法。

14、一种计算机可读存储介质，所述存储介质中的指令由设备的处理器执行时，使得所述设备能够执行如权利要求 1 至 11 任一项所述的音频检测方法。