



(12) 发明专利申请

(10) 申请公布号 CN 111932056 A

(43) 申请公布日 2020.11.13

(21) 申请号 202010568115.6

(22) 申请日 2020.06.19

(71) 申请人 北京文思海辉金信软件有限公司  
地址 100192 北京市海淀区西小口路66号  
东升科技园C区4号楼401室

(72) 发明人 唐文军 贾晓谦 宋子岳 王冉

(74) 专利代理机构 北京华进京联知识产权代理  
有限公司 11606

代理人 赵文静

(51) Int. Cl.

G06Q 10/06 (2012.01)

G06N 3/08 (2006.01)

G06N 3/04 (2006.01)

G10L 25/30 (2013.01)

G10L 25/63 (2013.01)

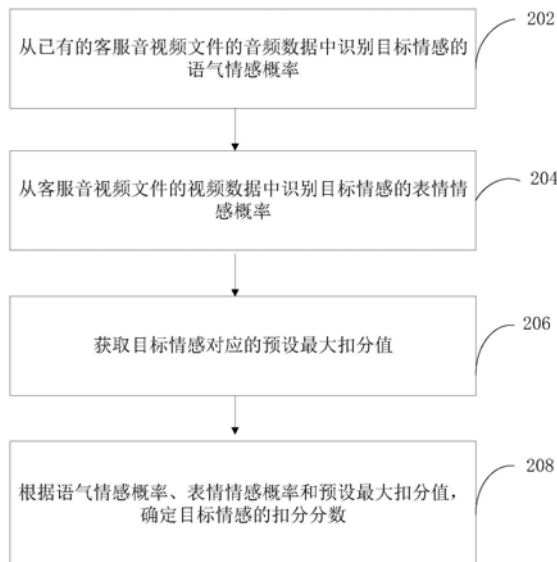
权利要求书2页 说明书10页 附图5页

(54) 发明名称

客服质量评分方法、装置、计算机设备和存储介质

(57) 摘要

本申请涉及一种客服质量评分方法、装置、计算机设备和存储介质。所述方法包括：从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率，从客服音视频文件的视频数据中识别目标情感的表情情感概率，获取目标情感对应的预设最大扣分值，根据语气情感概率、表情情感概率和预设最大扣分值，确定目标情感的扣分数。上述方法从音频和视频两个角度进行情感分析，实现音视频数据的解耦和情感分析判断的领域专业化，提高情感判断的准确度，并且，能够有效检测说话者的说话态度及情绪状态，丰富评分维度，提高评分准确度。



1. 一种客服质量评分方法,其特征在于,所述方法包括:  
从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;  
从所述客服音视频文件的视频数据中识别所述目标情感的表情情感概率;  
获取所述目标情感对应的预设最大扣分值;  
根据所述语气情感概率、所述表情情感概率和所述预设最大扣分值,确定所述目标情感的扣分分数。
2. 根据权利要求1所述的方法,其特征在于,所述从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率,包括:  
将所述音频数据进行切片处理,得到音频切片,所述音频切片的长度与预训练语气情感识别模型的样本数据一致,以使不同长度的音频数据能够适用于所述预训练语气情感识别模型;  
根据所述预训练语气情感识别模型计算所述音频切片对应的语气情感概率。
3. 根据权利要求2所述的方法,其特征在于,所述根据所述预训练语气情感识别模型计算所述音频切片对应的语气情感概率,包括:  
获取所述音频切片的音频特征向量,将所述音频特征向量输入至已训练的基于LSTM神经网络的语气情感分析模型进行语气情感分析,得到所述音频切片对应的语气情感概率。
4. 根据权利要求1所述的方法,其特征在于,所述从所述客服音视频文件的视频数据中识别所述目标情感的表情情感概率,包括:  
通过人脸识别模块识别所述视频数据的每一个帧图像中是否包含人脸图像;  
若是,则根据预训练表情情感识别模型计算所述帧图像对应的所述表情情感概率。
5. 根据权利要求4所述的方法,其特征在于,根据预训练表情情感识别模型计算所述帧图像对应的所述表情情感概率,包括:  
将所述帧图像输入至基于Xception神经网络的表情情感识别模型进行表情情感分析,得到所述帧图像对应的表情情感概率。
6. 根据权利要求1至4中任一项所述的方法,其特征在于,根据所述语气情感概率、所述表情情感概率和所述最大扣分值,确定所述目标情感的扣分分数,包括:  
计算所述音频数据中预设时域范围内的音频切片对应的语气情感概率均值;  
计算所述视频数据中所述预设时域范围内的包含人脸图像的全部帧图像对应的表情情感概率均值;  
将所述语气情感概率均值与所述表情情感概率均值的加和或乘积与所述最大扣分值相乘,得到所述目标情感的扣分分数。
7. 根据权利要求6所述的方法,其特征在于,所述方法还包括:  
获取用户选择的评分策略;  
根据所述用户选择的评分策略,确定所述预设时间范围,其中,若所述评分策略为全局评分策略,则所述预设时间范围为整个音视频文件对应的时间范围,若所述评分策略为局部评分策略,则所述预设时间范围为用户选择的一个时间切片对应的时间范围。
8. 一种客服质量评分装置,其特征在于,所述装置包括:  
语气情感概率识别模块,用于从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;

表情情感概率识别模块,用于从所述客服音视频文件的视频数据中识别所述目标情感的表情情感概率;

扣分值获取模块,用于获取所述目标情感对应的预设最大扣分值;

质量评分模块,用于根据所述语气情感概率、所述表情情感概率和所述预设最大扣分值,确定所述目标情感的扣分分数。

9.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述的方法的步骤。

10.一种计算机可读存储介质,其上存储有计算机程序,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至7中任一项所述的方法的步骤。

## 客服质量评分方法、装置、计算机设备和存储介质

### 技术领域

[0001] 本申请涉及情感分析技术领域,特别是涉及一种客服质量评分方法、装置、计算机设备和存储介质。

### 背景技术

[0002] 随着人工智能技术的发展,出现了情感分析(或情感识别)技术。现如今,情感分析技术已在多个领域得到广泛应用。基于情感分析,进而实现产品的准确的用户定位和营销,也受到大家的青睐。例如在商品零售领域,用户的评价对于零售商和生产商都是非常重要的反馈信息,通过对海量用户的评价进行情感分析,可以量化用户对产品及其竞品的褒贬程度,从而了解用户对于产品的诉求以及自己产品与竞品的对比优劣,进而进行产品的情感包装,以实现产品各项经营目标。

[0003] 目前,现有的用户情感分析技术,大多是从文本中分析出人们对于实体及其属性所表达的情感倾向以及观点,主要是对文本建立分类模型或通过极性词匹配度计算情感得分,从而判断说话者的态度是积极还是消极。

[0004] 然而,现有的对文本的情感建模,分析维度比较单一,对于客服质量评价这类有专业细分的领域,分析的业务话术中包含大量的中性词及专业词汇,使得情感判断的准确率较低,且判断结果无法反映用户的真实情绪态度,使得评分准确度较低。

### 发明内容

[0005] 基于此,有必要针对上述技术问题,提供一种评分准确度高的客服质量评分方法、装置、计算机设备和存储介质。

[0006] 一种客服质量评分方法,所述方法包括:

[0007] 从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;

[0008] 从客服音视频文件的视频数据中识别目标情感的表情情感概率;

[0009] 获取目标情感对应的预设最大扣分值;

[0010] 根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分分数。

[0011] 在其中一个实施例中,从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率,包括:

[0012] 将音频数据进行切片处理,得到音频切片,音频切片的长度与预训练语气情感识别模型的样本数据一致,以使不同长度的音频数据能够适用于预训练语气情感识别模型;

[0013] 根据预训练语气情感识别模型计算音频切片对应的语气情感概率。

[0014] 在其中一个实施例中,根据预训练语气情感识别模型计算音频切片对应的语气情感概率,包括:

[0015] 获取音频切片的音频特征向量,将音频特征向量输入至已训练的基于LSTM神经网络的语气情感分析模型进行语气情感分析,得到音频切片对应的语气情感概率。

[0016] 在其中一个实施例中,从客服音视频文件的视频数据中识别目标情感的表情情感

概率,包括:

[0017] 通过人脸识别模块识别视频数据的每一个帧图像中是否包含人脸图像;

[0018] 若是,则根据预训练表情情感识别模型计算帧图像对应的表情情感概率。

[0019] 在其中一个实施例中,根据预训练表情情感识别模型计算帧图像对应的表情情感概率,包括:

[0020] 将帧图像输入至基于Xception神经网络的表情情感识别模型进行表情情感分析,得到帧图像对应的表情情感概率。

[0021] 在其中一个实施例中,根据语气情感概率、表情情感概率和最大扣分值,确定目标情感的扣分分数,包括:

[0022] 计算音频数据中预设时域范围内的音频切片对应的语气情感概率均值;

[0023] 计算视频数据中预设时域范围内的包含人脸图像的全部帧图像对应的表情情感概率均值;

[0024] 将语气情感概率均值与表情情感概率均值的加和或乘积与最大扣分值相乘,得到目标情感的扣分分数。

[0025] 在其中一个实施例中,方法还包括:

[0026] 获取用户选择的评分策略;

[0027] 根据用户选择的评分策略,确定预设时间范围,其中,若评分策略为全局评分策略,则预设时间范围为整个音视频文件对应的时间范围,若评分策略为局部评分策略,则预设时间范围为用户选择的一个时间切片对应的时间范围。

[0028] 一种服务质量评分装置,所述装置包括:

[0029] 语气情感概率识别模块,用于从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;

[0030] 表情情感概率识别模块,用于从客服音视频文件的视频数据中识别目标情感的表情情感概率;

[0031] 扣分值获取模块,用于获取目标情感对应的预设最大扣分值;

[0032] 质量评分模块,用于根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分分数。

[0033] 一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现以下步骤:

[0034] 从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;

[0035] 从客服音视频文件的视频数据中识别目标情感的表情情感概率;

[0036] 获取目标情感对应的预设最大扣分值;

[0037] 根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分分数。

[0038] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现以下步骤:

[0039] 从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;

[0040] 从客服音视频文件的视频数据中识别目标情感的表情情感概率;

[0041] 获取目标情感对应的预设最大扣分值;

[0042] 根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分分数。

[0043] 上述客服质量评分方法、装置、计算机设备和存储介质,基于客服音视频文件,对其中的音频数据和视频数据进行处理,并从音频和视频两个角度独立进行情感分析,得到目标情感的语气情感概率以及表情情感概率,再获取目标情感对应的最大扣分值,结合语气情感概率以及表情情感概率,得到相应的质量评分,实现音视频数据的解耦和情感分析判断的领域专业化,提高情感判断的准确度,并且,能够有效检测说话者的说话态度及情绪状态,丰富评分维度,提高评分准确度。

#### 附图说明

- [0044] 图1为一个实施例中客服质量评分方法的应用环境图;
- [0045] 图2为一个实施例中客服质量评分方法的流程示意图;
- [0046] 图3为一个实施例中客服质量评分方法的详细流程示意图;
- [0047] 图4为另一个实施例中客服质量评分方法的流程示意图;
- [0048] 图5为另一个实施例中计算客服质量评分步骤的流程示意图;
- [0049] 图6为一个实施例中客服质量评分装置的结构框图;
- [0050] 图7为另一个实施例中客服质量评分装置的结构框图;
- [0051] 图8为一个实施例中计算机设备的内部结构图。

#### 具体实施方式

[0052] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0053] 本申请提供的音视频数据的客服质量评分方法,可以应用于如图1所示的应用环境中。其中,终端102通过网络与服务器104进行通信。用户通过终端102上传需要进行客服质量评分的客服音视频文件(即客服话术音视频数据)至服务器104。然后,于终端102进行操作,通过终端102发送客服质量评分请求至服务器104,服务器104响应该请求,从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率,从客服音视频文件的视频数据中识别目标情感的表情情感概率,获取目标情感对应的预设最大扣分值,根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分分数。其中,终端102可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0054] 在一个实施例中,如图2所示,提供了一种客服质量评分方法,以该方法应用于图1中的服务器为例进行说明,包括以下步骤:

[0055] 步骤202,从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率。

[0056] 客服音视频文件可以包含客服人员说话场景的音频数据和视频数据。具体实施时,以对客服人员进行话术质量评分为例,音视频数据可以是一段包含有客服人员进行话术培训的数据。在实际应用中,当用户于终端102上传完待评分的客服音视频数据后,点击“客服质量评分”按钮,终端102生成客服质量评分请求,并将请求发送至服务器104。响应该客服质量评分请求,获取该音视频文件,识别音视频文件中音频数据的目标情感(如生气、紧张或难过等)的语气情感概率。具体的,识别音频数据中的目标情感的语气概率,可以是

根据预训练的语气情感识别模型对音频数据进行包括生气、紧张、难过、反感、害怕、惊讶等不符合专业性要求的音频情绪识别,得到语气情感概率。该语气情感识别模型基于历史音频数据训练得到。

[0057] 如图3所示,在其中一个实施例中,步骤202包括:

[0058] 步骤222,将音频数据进行切片处理,得到音频切片,根据预训练语气情感识别模型计算音频切片对应的语气情感概率。

[0059] 音频切片是指将音频数据进行切割得到的一段时长的音频流。本实施例中,音频切片的长度与预训练语气情感识别模型的样本数据一致,以使不同长度的音频数据能够适用于预训练语气情感识别模型。具体实施时,可以是先提取音视频文件中的音频数据和视频数据,再根据预训练语气情感识别模型计算音频切片对应的语气情感概率。其中,音视频数据中的音频数据和视频数据可以借助媒体数据处理工具提取。例如,可以借助FFmpeg(结合Python语句)提取音视频源数据中的音频数据和视频数据。

[0060] 如图4所示,在其中一个实施例中,步骤222中的根据预训练语气情感识别模型计算音频切片对应的语气情感概率,包括:获取音频切片的音频特征向量,将音频特征向量输入至已训练的基于LSTM神经网络的语气情感分析模型进行语气情感分析,得到音频切片对应的语气情感概率。

[0061] 本实施例中,音频特征向量即为基于梅尔频率倒谱系数(MFCC)构建的MFCC特征向量。具体实施时,音频切片可以是时长与音频切片样本数据一致的标准数据。对目标音频切片数据进行语音信号处理,得到MFCC的过程可以是:先对音频切片数据进行预加重处理,使用高通滤波器提升高频信息,输出经过分帧处理,引入一定的重叠率保证音频信号的帧内平稳性,再使用汉明窗进行加窗处理,使帧与帧之间变得平滑。通过快速傅里叶变换将短时域信号转换到频域,再使用梅尔刻度滤波器过滤,梅尔刻度的滤波器组在低频部分的分辨率高,高频部分的分辨率低,而人耳对于低频部分更加敏感,因此可以使每个频率段得到的数值最大限度还原人类的听觉特性。再进行对数能量计算,加入每一帧的音高特征,最后进行离散余弦变换,将能量集中到变换结果的低频部分,实现音频信号的压缩,得到最终的梅尔频率倒谱系数,即MFCC,再基于MFCC数据构建MFCC特征向量。

[0062] 其中,语气情感分析模型基于历史音频数据(包括音频切片数据及其对应的MFCC特征向量)训练得到。具体实施时,语气情感分析模型可以是基于LSTM(Long Short-Term Memory,长短期记忆)神经网络的语气情感分析模型,该模型的训练过程可以是将已标注的音频样本处理成时长、采样率、声道一致的标准数据,通过高通滤波器进行预加重处理,经过音频分帧、加窗、频域转换、梅尔刻度滤波器过滤、对数运算、DCT变换,最终得到每个音频样本的MFCC系数,进而构建MFCC特征向量,将MFCC特征向量作为该语气情感分析模型的输入特征,进行音频情绪识别。音频情绪识别采用长短期记忆单元,将上步音频特征按照时序输入LSTM单元,经由dropout和全连接层,进行梯度下降过程,直至收敛,得到语气情感分析模型。语气情感分析过程可以是将得到的目标音频切片数据的MFCC特征向量按照时序先后循环输入到LSTM单元,发掘音频数据的前后依赖关系,通过Dropout层防止过拟合,输出依次经过三个全连接网络,顺序通过relu、tanh激活,通过softmax层得到各类情感的音频情感类别的概率分布。再基于概率分布,得到如生气、紧张、难过等情感的语气情感概率。

[0063] 步骤204,从客服音视频文件的视频数据中识别目标情感的表情情感概率。

[0064] 具体实施时,可以是基于历史视频数据(包括人脸图像的视频帧图像)训练得到的表情情感识别模型识别目标情感(如生气、紧张或难过)的表情情感概率。

[0065] 在其中一个实施例中,步骤204包括:步骤224,通过人脸识别模块识别视频数据的每一个帧图像中是否包含人脸图像,若是,则根据预训练表情情感识别模型计算帧图像对应的表情情感概率。

[0066] 帧图像,即为对视频数据进行分帧处理,得到的帧图像,将所有视频帧图像连续播放能够得到分帧处理前的视频数据。在实际应用中,可以是先对视频数据进行分帧处理,得到连续的帧图像,再通过人脸识别模型(人脸识别模型)对每一帧图像进行人脸识别检测,再将包含人脸图像的帧图像输入预训练表情情感识别模型,针对各类目标情感,对输入的包含人脸图像的帧图像进行包括生气、紧张、难过等表情的表情识别,计算对应的表情情感概率。具体实施时,表情情感识别模型可以包括人脸识别模型以及基于Xception神经网络的图片情感分析模型。

[0067] 本实施例中,对人脸图像进行人脸识别包括:使用哈尔特征模板提取图像的边缘特征、线性特征、中心特征、对角特征,特征值为白色矩形像素和减去黑色矩形像素和,提取出人脸的灰度变化特征;具体计算过程采用积分图方法进行,将特征像素区域的累加和计算转化为特征模板顶点像素和的加减运算,提升运算效率。输出的人脸特征值通过多个AdaBoost强分类器,只有在通过每个强分类器都被认为是正脸的情况下才会判断为有人脸,否则判断为无人脸。

[0068] 具体的,人脸识别模型的训练过程可以是采用Haar分类器,将已标注的人脸图片样本进行Haar-like特征检测,使用积分图对特征值求算进行加速。对特征值进行筛选,训练针对优选特征值的AdaBoost弱分类器,将所有弱分类器进行组合构成强分类器。训练多个强分类器进行级联,得到最终的人脸识别模型。对于任意图片,级联分类器中任意一个强分类器判断其不具备人脸特征,则检测为非人脸图像。基于Xception神经网络的图片情感分析模型的训练过程可以是采用Xception网络结构,将已标注的表情图片样本输入Xception网络中,采用adam步长优化策略,计算交叉熵损失,直至收敛,得到图片情感分析模型。

[0069] 如图4所示,在其中一个实施例中,步骤224中根据预训练表情情感识别模型计算帧图像对应的表情情感概率,包括:

[0070] 将帧图像输入至基于Xception神经网络的表情情感识别模型进行表情情感分析,得到帧图像对应的表情情感概率。

[0071] 具体实施时,采用基于Xception神经网络的图片情感分析模型对人脸图像进行表情识别包括:将输入的人脸图像的单通道像素点经过两层3x3卷积核提取8个通道的特征图,每层的具体操作过程为:图像各像素点和8个3x3卷积核进行卷积运算,输出8个通道的特征图,分别进行归一化,再通过Relu函数激活,得到8个通道输出,重复进行两次该操作,输出进入Xception残差结构单元。单元内依次进入两个3x3深度可分离卷积单元,一个最大池化层,对特征图的空间相关性和通道相关性进行分离计算并降维,输出与外围经过1x1卷积核提取的通道相关性原始特征进行加和,深度可分离卷积单元操作具体为:将8通道特征图通过16个1x1卷积核进行卷积运算,输出的16个通道特征图分别与3x3的卷积核进行卷积,将得到的16个输出拼接得到16个通道特征图。深度可分离卷积操作结果经过归一化并



通过Relu函数激活,输入第二个深度可分离卷积单元,经过再次归一化及3x3最大池化操作,得到结果与1x1卷积核卷积后的原始特征图进行加和,输出重复上述步骤,经过两个Xception残差结构单元,卷积核数以倍数增长。输出64通道特征图,再经过64个卷积核的深度可分离卷积操作,归一化并由Relu激活,重复三次此操作。输出经过Xception残差网络单元,经过128个卷积核输出128通道的特征图,再次通过128核深度可分离卷积、归一化、Relu激活,输入6通道深度可分离卷积,输出经过归一化、Relu激活、平均池化,最终输入softmax层得到情感分类概率分布。网络采取多个小卷积核替换大卷积核策略,大量使用深度可分离卷积单元,将空间相关性和通道相关性分开处理,减少参数量,提高运行效率,残差单元可以有效抑制梯度衰减,提升模型表现。该网络主体结构基于Xception,对卷积核数按照实验效果进行了调整,提升运算性能和识别效果。

[0072] 步骤206,获取目标情感对应的预设最大扣分值。

[0073] 具体实施时,针对各类目标情感,系统预设有相应的预设最大扣分值。目标情感以生气为例,系统预设的生气情绪的最大扣分阈值为 $SCORE_{total}$ ,当得到生气情绪对应的语气情感概率和表情情感概率后,便相应获取生气情绪对应的预设最大扣分值。

[0074] 步骤208,根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分数。

[0075] 当针对各类目标情感,得到相应的表情情感概率和语气情感概率后,可以是结合系统预设好的各类情绪的扣分值,进行加权得到相应的扣分数,进而得到总体客服质量评分。在另一个实施例中,基于各类语气情感概率值以及各类表情情感概率值,计算客服质量评分之后,还包括:根据客服质量评分,构建用户情感画像。具体的,可基于客服质量评分,得到说话者是积极态度还是消极态度,进而构建用户情感画像。以客服评价为例,可根据客服质量评分,对客服的服务质量进行评价,得到其情感画像。

[0076] 上述客服质量评分方法中,基于客服音视频文件,对其中的音频数据和视频数据进行处理,并从音频和视频两个角度独立进行情感分析,得到目标情感的语气情感概率以及表情情感概率,再获取目标情感对应的最大扣分值,结合语气情感概率以及表情情感概率,得到相应的质量评分,实现音视频数据的解耦和情感分析判断的领域专业化,提高情感判断的准确度,并且,能够有效检测说话者的说话态度及情绪状态,丰富评分维度,提高评分准确度。

[0077] 在其中一个实施例中,方法还包括:获取用户选择的评分策略,根据用户选择的评分策略,确定预设时间范围,提取预设时间范围的音频切片和帧图像。其中,若评分策略为全局评分策略,则预设时间范围为整个音视频文件对应的时间范围,若评分策略为局部评分策略,则预设时间范围为用户选择的一个时间切片对应的时间范围。

[0078] 本实施例中,提供两种评分策略,包括局部评分策略(亦可称局部客服质量评分策略)和全局评分策略(亦可称全局客服质量评分策略)。局部客服质量评分策略即为适合用户有实时评分要求或在线评分要求采取的评分策略,具体是指针对一段音视频数据,陆续提取部分音视频数据,如5秒或10秒内音视频数据,实时对提取出的部分音视频数据进行情感分析,得到局部客服质量评分,并实时显示客服质量评分,例如可每10秒显示一次。全局评分策略即为适合用户对评分实时性要求不高,而注重总体评分时选取的评分策略。其具体为在音视频录制完成之后,对整体的音视频数据进行情感分析,得到累计客服质量评分,

对说话者建立总体的客服质量评分及用户情感画像。具体实施时,可根据客服质量评分策略,对音频数据和视频数据进行的预处理可包含全局处理和/或局部处理。在实际应用中,当用户于终端102上传完待评分的话术培训音视频数据后,选取相应的客服质量评分策略即局部客服质量评分策略和/或全局客服质量评分策略,接收用户客服质量评分策略选取操作,进而根据用户选取的客服质量评分策略,提取预设时间范围的音频切片数据和视频帧图像。

[0079] 具体实施时,当用户选取的评分策略为局部客服质量评分策略时,在保证评分的同步性,需要将音频数据和视频数据时间片进行同步调整,然后基于同步调整后的音频数据和视频数据,提取出预设时域范围(如10秒)的局部音频切片数据,再对应选取预设时域范围对应的即局部音频切片数据对应的视频帧图像。在另一个实施例中,还可以是提取当前时刻对应的预设时域范围的局部音视频切片数据。如提取当前时间内10秒的局部音视频切片数据。当用户选取的客服质量评分策略为全局客服质量评分策略时,则将整段特定长时的音频数据和视频数据作为整体,分别对音频数据进行切片处理,对视频数据进行分帧处理,得到多段音频切片数据和多张视频帧图像,如得到M段音频切片数据,每个音频切片时长为S以及总帧数为W的视频帧图像。针对局部和全局两种评分策略,对音视频数据进行相应处理,能够有针对性的得到两种评分策略下的各类情感概率值。本实施例中,通过提供多种话术评分策略,能够提高用户选择度和客户满足度。

[0080] 如图5所示,在其中一个实施例中,根据语气情感概率、表情情感概率和最大扣分值,确定目标情感的扣分分数,包括:

[0081] 步骤228,计算音频数据中预设时域范围内的音频切片对应的语气情感概率均值;

[0082] 步骤248,计算视频数据中预设时域范围内的包含人脸图像的全部帧图像对应的表情情感概率均值;

[0083] 步骤268,将语气情感概率均值与表情情感概率均值的加和或乘积与最大扣分值相乘,得到目标情感的扣分分数。

[0084] 具体实施时,可根据不同的评分策略对客服质量进行评分。当客服质量评分策略为局部客服质量评分时,即针对局部的音频切片S1,切片S1内的生气情绪的概率为 $P_A$ ,切片S1内,视频总帧数为W,每一帧图像对应生气情绪的概率为 $P_{V1}, P_{V2}, \dots, P_{Vs}$ ,然后,针对每一类情感类别(如生气、紧张),统计每一帧视频帧图像对应的各类表情情感概率总值,如生气

概率总值则为 $\sum_{i=1}^{i=s} P_{Vi}$ ,再基于总值计算W帧图像的生气概率均值。基于上述数据,同时可设置

两种扣分模式,第一种为宽松的情感计分方式,局部音频切片S1对应的生气情绪的扣分

$$\text{score}_1 = P_A * \frac{\sum_{i=1}^{i=s} P_{Vi}}{W} * \text{SCORE}_{\text{total}},$$
 只要语气情感分析模型判定的概率值或者表情情感识别模

型判定的每一帧图像情感概率值中任一个为0,则此切片S1内的扣分概率为0,即此切片不扣分。另一种为严格的情感计分方式,局部音频切片S1对应的生气情绪的扣分

$$\text{score}_1 = \frac{P_A + \sum_{i=1}^{i=s} P_{Vi}}{2} * \text{SCORE}_{\text{total}},$$
 即只有在语气情感分析模型判定的概率值和表情情感识别

模型判定的每一帧图像情感概率值均为0时,则此切片S1内的扣分概率为0。同理,按照上述方式可计算其他情绪(难过、紧张及害怕等)对应的情感扣分,然后统计各类情感的扣分,得到局部情感扣分,进而得到局部服务质量评分。局部服务质量评分与局部音频切片数据对应的各类语气情感概率值以及各类表情情感概率总值成正相关。可以理解的是,计算局部服务质量评分的方式包括但不限于上述方式,在其他实施例中,根据各类语气情感概率值和各类表情情感概率值,还可以有其他的评分方式。

[0085] 具体实施时,服务质量评分策略以全局服务质量评分为例,情感类别以生气为例,系统预设的生气情绪的扣分阈值(最大扣分)为 $SCORE_{total}$ ,即音频切片总个数为M,每个切片时长为S,语气情感分析模型判定得到的M个音频切片对应生气情绪的概率分别为 $P_{A1}$ ,

$P_{A2}, \dots, P_{AM}$ ,统计音频切片对应的各类语气情感概率总值,如生气概率总值为 $\sum_{j=1}^{j=M} P_{Aj}$ ,进而

得到生气情绪的音频概率均值。视频帧图像的总帧数为W,表情情感识别模型判定得到的每一帧图像对应生气情绪的概率为 $P_{V1}, P_{V2}, \dots, P_{Vs}$ ,然后,针对每一类情感类别(如生气、紧张)

张),统计视频帧图像对应的各类表情情感概率总值,如生气概率总值则为 $\sum_{i=1}^{i=s} P_{Vi}$ ,进而得到

生气情绪的视频概率均值。基于上述数据,同样,可设置两种扣分模式,第一种为宽松的情

感计分方式,总体音视频数据对应的生气情绪的总扣分 
$$score = \frac{\sum_{j=1}^{j=M} P_{Aj}}{M} * \frac{\sum_{i=1}^{i=s} P_{Vi}}{W} * SCORE_{total},$$

只要语气情感分析模型判定的每一个切片的情感概率均值或者表情情感识别模型判定的每一帧图像的情感概率均值中任一个为0,则总扣分为0,即不扣分。另一种为严格的情感计

分方式,总体音视频数据的总扣分 
$$score = \frac{\sum_{j=1}^{j=M} P_{Aj}}{M} + \frac{\sum_{i=1}^{i=s} P_{Vi}}{W} * SCORE_{total},$$
 即只有在语气情感分

析模型判定的每一个切片的情感概率均值和表情情感识别模型判定的每一帧图像的情感概率均值均为0时,则总扣分为0。同理,按照上述方式可计算其他情绪(难过、紧张及害怕等)对应的情感扣分,然后统计各类情感的扣分,得到全局服务质量评分,再基于全局服务质量评分,构建用户情感画像。全局服务质量评分与各类语气情感概率总值和各类表情情感概率总值呈正相关。可以理解的是,计算全局服务质量评分的方式包括但不限于上述方式,在其他实施例中,根据各类语气情感概率值和各类表情情感概率值,得到相应的情感扣分也还可以有其他方式。

[0086] 通过两种模式扣分策略的引入,可以将针对音视频数据的情感分析模型的判断结果转化为服务质量的量化评分。同步时序切片加权评分策略(局部服务质量评分策略)适用于实时、准实时、在线的服务评价需求,更多的考量音视频的实时相关性,其严格的评分模式可以实现音视频协同的情绪判断,增强结果的准确性、可信度,是细粒度的计分方式;全局时序分布加权评分策略(全局服务质量评分策略)适用于对实时性要求不高、注重总体的服务评价需求,更多的考量音视频的全局情绪分布,是粗粒度的计分方式;通过调节各情绪

的最大扣分数,可以动态调整客服质量评分对于评分系统的影响程度,增强系统的灵活性。

[0087] 应该理解的是,虽然图2-5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2-5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。

[0088] 在其中一个实施例中,如图6所示,提供了一种客服质量评分装置,包括:语气情感概率识别模块510、表情情感概率识别模块520、扣分值获取模块530和质量评分模块540,其中:

[0089] 语气情感概率识别模块510,用于从已有的客服音视频文件的音频数据中识别目标情感的语气情感概率;

[0090] 表情情感概率识别模块520,用于从客服音视频文件的视频数据中识别目标情感的表情情感概率;

[0091] 扣分值获取模块530,用于获取目标情感对应的预设最大扣分值;

[0092] 质量评分模块540,用于根据语气情感概率、表情情感概率和预设最大扣分值,确定目标情感的扣分数。

[0093] 在其中一个实施例中,语气情感概率识别模块510还用于将音频数据进行切片处理,得到音频切片,根据预训练语气情感识别模型计算音频切片对应的语气情感概率,音频切片的长度与预训练语气情感识别模型的样本数据一致,以使不同长度的音频数据能够适用于预训练语气情感识别模型。

[0094] 在其中一个实施例中,语气情感概率识别模块510还用于获取音频切片的音频特征向量,将音频特征向量输入至已训练的基于LSTM神经网络的语气情感分析模型进行语气情感分析,得到音频切片对应的语气情感概率。

[0095] 在其中一个实施例中,表情情感概率识别模块520还用于通过人脸识别模块识别视频数据的每一个帧图像中是否包含人脸图像,若是,则根据预训练表情情感识别模型计算帧图像对应的表情情感概率。

[0096] 在其中一个实施例中,表情情感概率识别模块520还用于将帧图像输入至基于Xception神经网络的表情情感识别模型进行表情情感分析,得到帧图像对应的表情情感概率。

[0097] 在其中一个实施例中,质量评分模块540还用于计算音频数据中预设时域范围内的音频切片对应的语气情感概率均值,计算视频数据中预设时域范围内的包含人脸图像的全部帧图像对应的表情情感概率均值,将语气情感概率均值与表情情感概率均值的加和或乘积与最大扣分值相乘,得到目标情感的扣分数。

[0098] 如图7所示,在其中一个实施例中,装置还包括评分策略处理模块550,用于获取用户选择的评分策略,根据用户选择的评分策略,确定预设时间范围,其中,若评分策略为全局评分策略,则预设时间范围为整个音视频文件对应的时间范围,若评分策略为局部评分策略,则预设时间范围为用户选择的一个时间切片对应的时间范围。

[0099] 关于客服质量评分装置的具体限定可以参见上文中对于客服质量评分方法的限

定,在此不再赘述。上述客服质量评分装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0100] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图8所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储音视频数据、各类情感扣分阈值等。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种客服质量评分方法。

[0101] 本领域技术人员可以理解,图8中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0102] 在其中一个实施例中,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述客服质量评分方法中的步骤。

[0103] 在其中一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现上述客服质量评分方法中的步骤。

[0104] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。

[0105] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0106] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

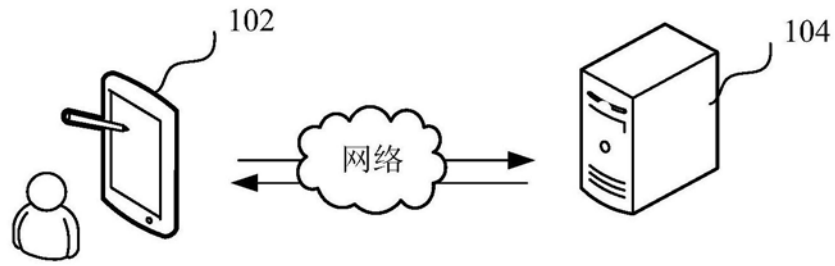


图1

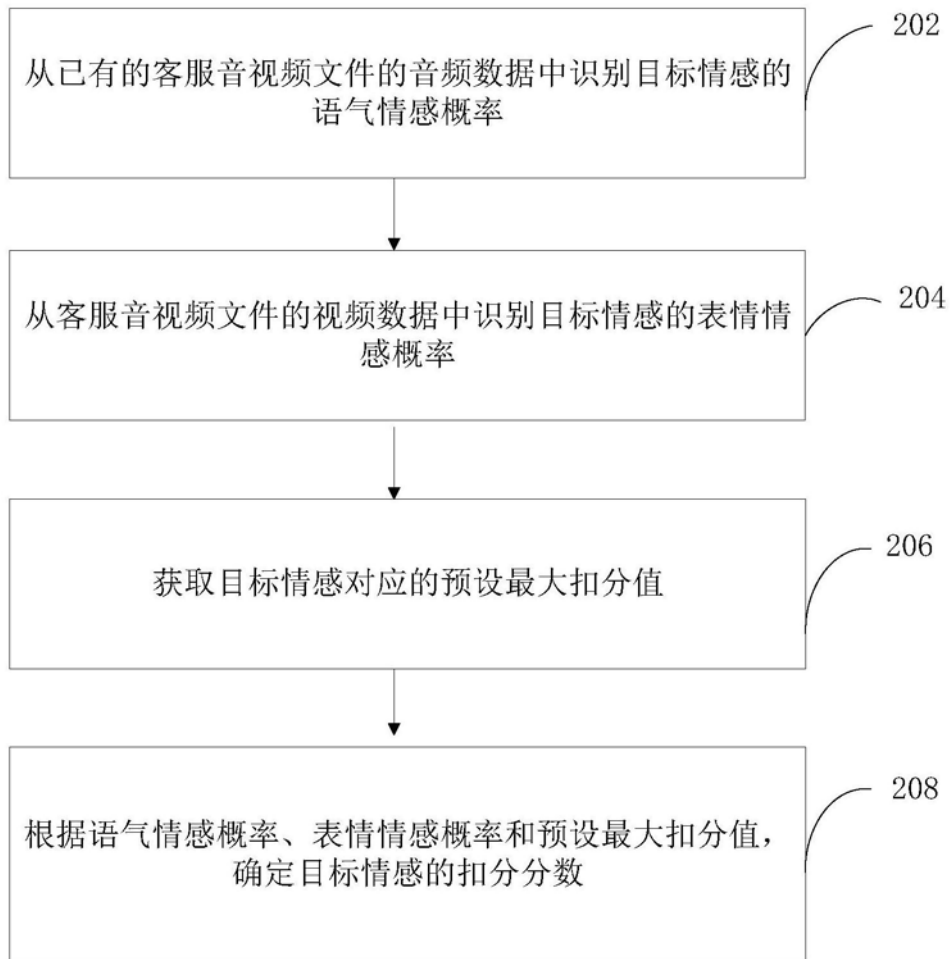


图2

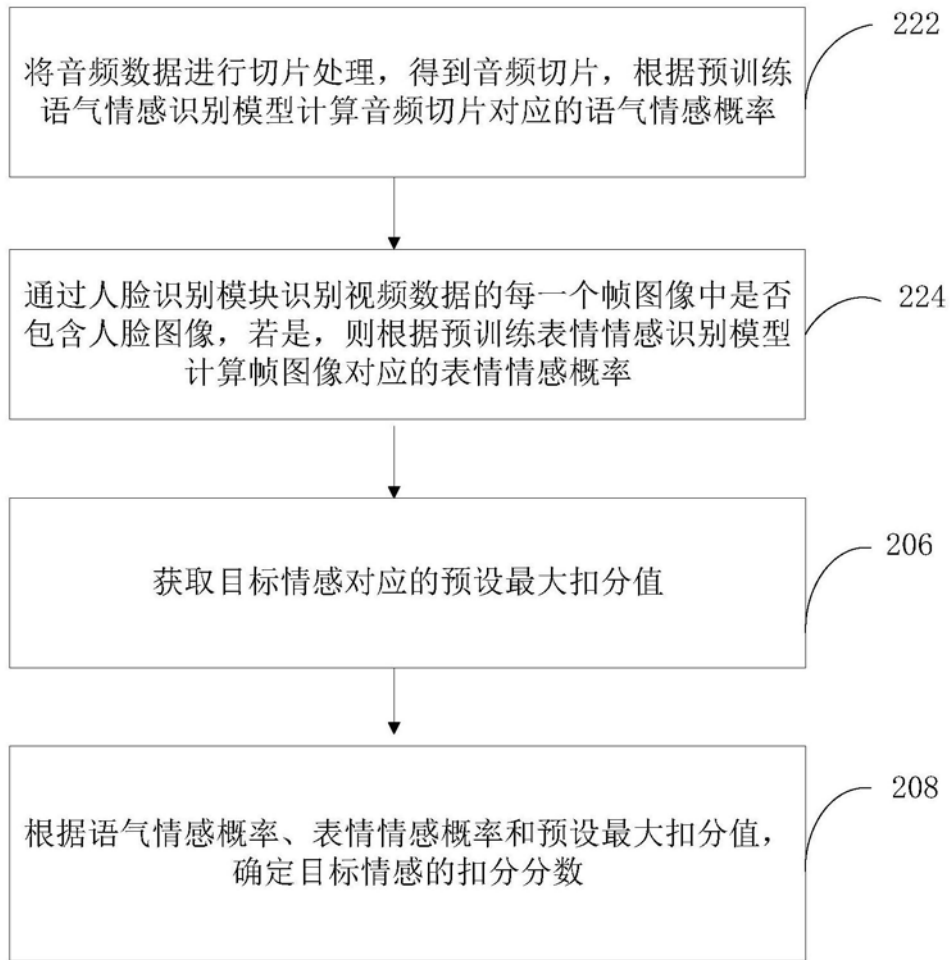


图3

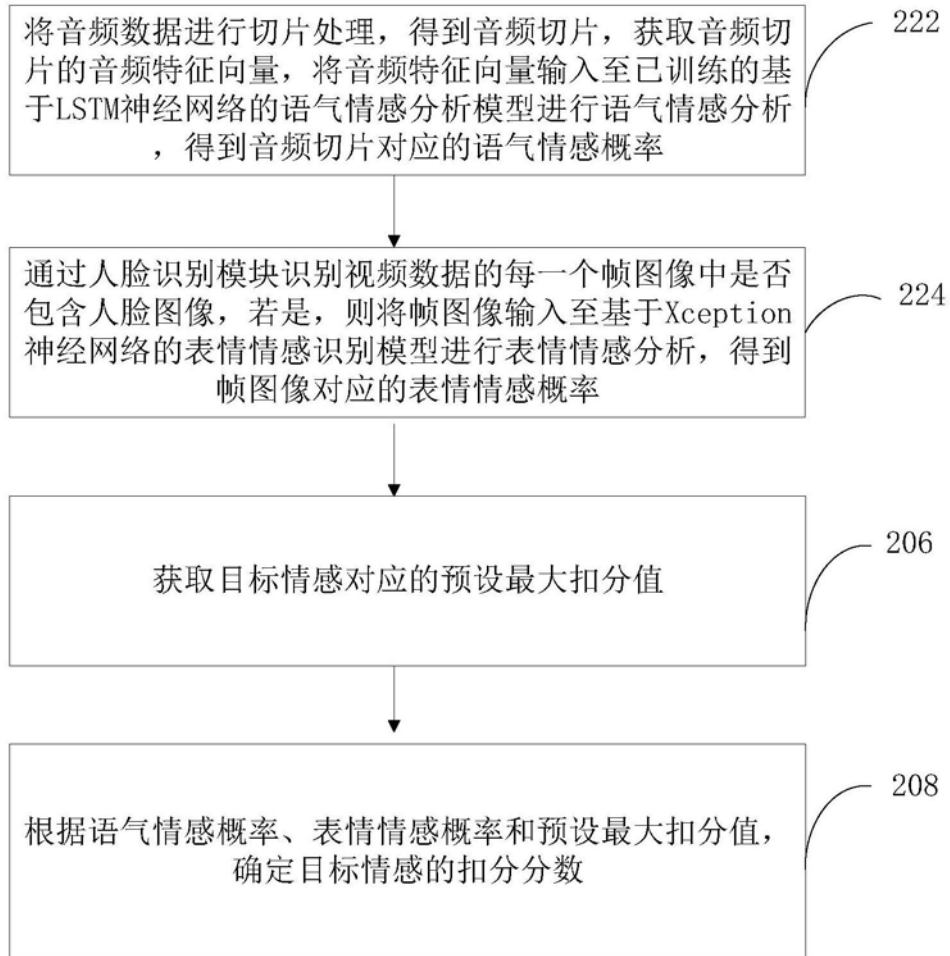


图4



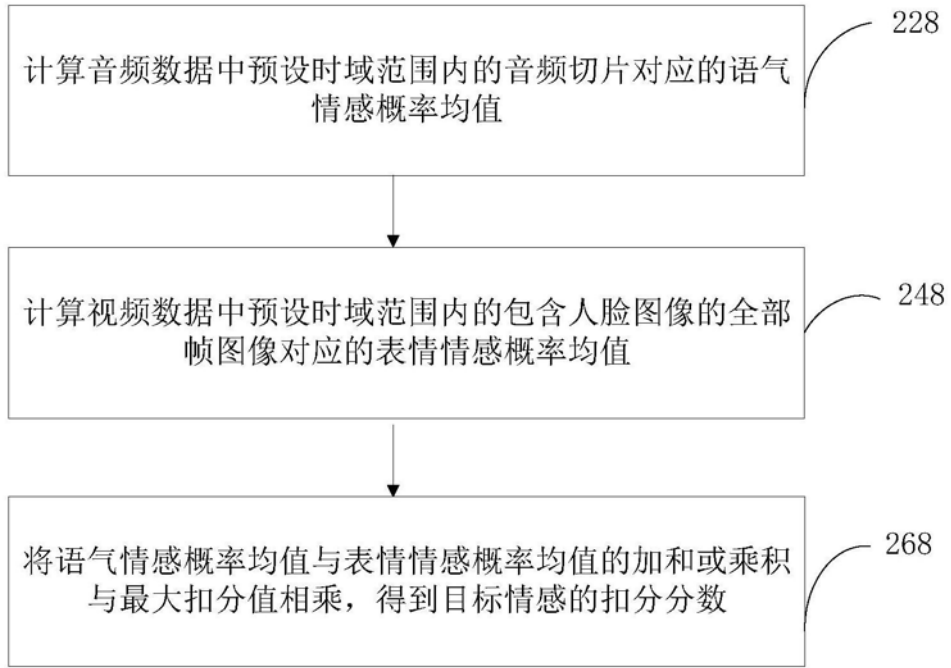


图5

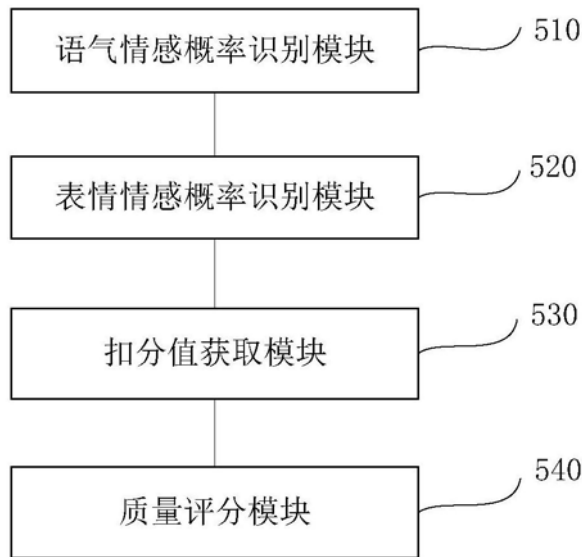


图6

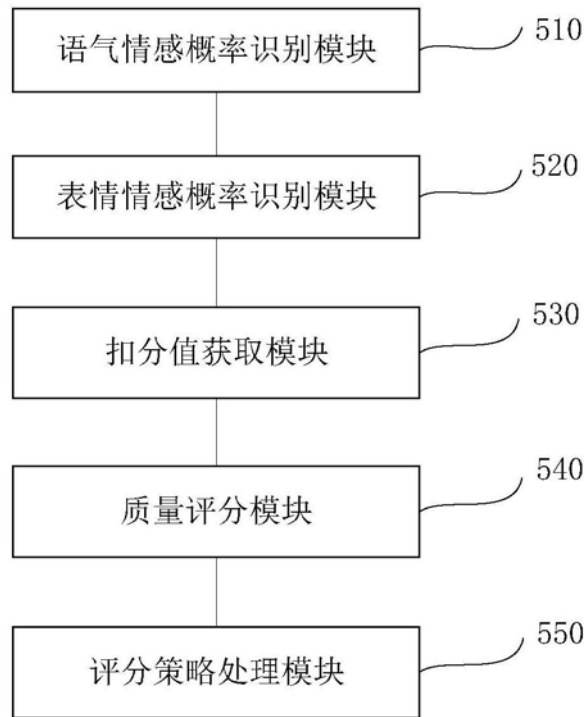


图7

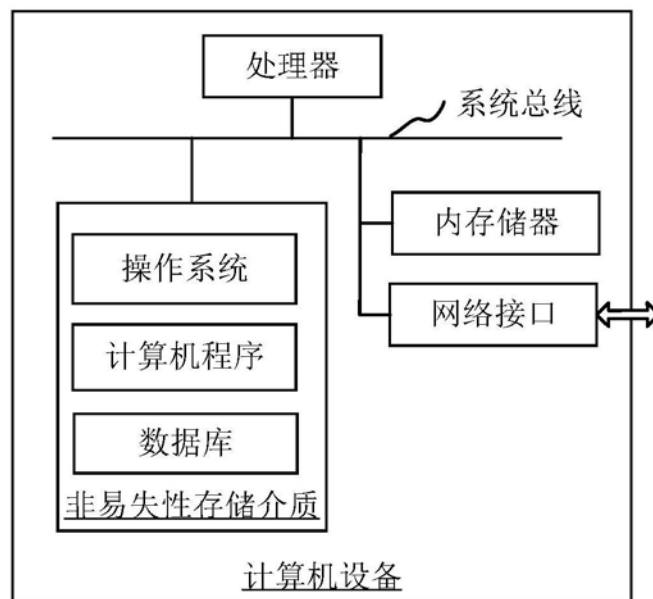


图8