



(12) 发明专利

(10) 授权公告号 CN 119274591 B

(45) 授权公告日 2025. 05. 02

(21) 申请号 202411732447.8

G10L 15/06 (2013.01)

(22) 申请日 2024.11.29

G10L 15/07 (2013.01)

(65) 同一申请的已公布的文献号

G10L 15/16 (2006.01)

申请公布号 CN 119274591 A

G10L 15/22 (2006.01)

G10L 21/0232 (2013.01)

(43) 申请公布日 2025.01.07

(56) 对比文件

(73) 专利权人 芯知科技(江苏)有限公司

CN 113611286 A, 2021.11.05

地址 210000 江苏省南京市江宁区六朝路

CN 114299995 A, 2022.04.08

18号高铁网谷A座709室

审查员 郭忆

(72) 发明人 林钦松 韩耀华

(74) 专利代理机构 南京致飞专利代理事务所

(普通合伙) 32736

专利代理师 纪德虎

(51) Int. Cl.

G10L 25/63 (2013.01)

G10L 15/02 (2006.01)

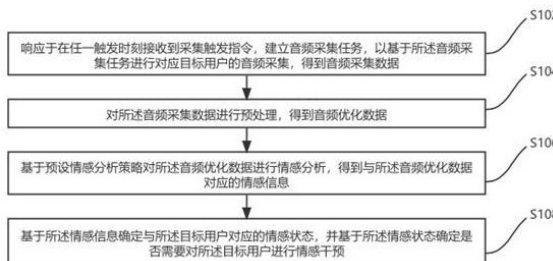
权利要求书3页 说明书22页 附图2页

(54) 发明名称

基于人工智能的语音情感分析方法及系统

(57) 摘要

本发明提供一种基于人工智能的语音情感分析方法及系统,其中,所述方法包括以下步骤:响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据;对所述音频采集数据进行预处理,得到音频优化数据;基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息;基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。本发明至少提高了客户满意度。



1. 一种基于人工智能的语音情感分析方法,其特征在于,包括以下步骤:

响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述语音采集任务进行对应目标用户的音频采集,得到音频采集数据,包括:

响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,并调取与所述音频采集任务对应的预设采集指示音频进行输出;

响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的测试音频数据;

基于所述测试音频数据对所述目标用户进行测试评分,并基于得到的音频测试分值确定与所述目标用户对应的音频属性;

基于所述测试音频数据确定所述目标用户与所述预设音频采集面之间的音频输出夹角,并基于音频输出夹角确定与所述目标用户对应的偏移评分;

基于所述测试音频数据确定与所述目标用户对应的音频输出分贝,并基于所述音频输出分贝确定与所述目标用户对应的音量评分;

对所述偏移评分以及音量评分进行融合计算,并将得到的与所述目标用户对应的音频测试分值与基准音频分值进行比较;

当所述音频测试分值小于所述基准音频分值时,将所述目标用户的输出属性确定为负向评价属性;

当所述音频测试分值大于等于所述基准音频分值时,将所述目标用户的输出属性确定为正向评价属性;

响应于所述目标用户的音频属性为正向评价属性,调取与所述音频采集任务对应的预设采集确定音频进行输出;

响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的音频采集数据;

或者,

响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出;

对所述音频采集数据进行预处理,得到音频优化数据;

基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息;

基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。

2. 根据权利要求1所述的基于人工智能的语音情感分析方法,其特征在于,

响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出,包括:

响应于所述目标用户的音频属性为负向评价属性,获取与所述目标用户对应的音频输出夹角以及音频输出分贝;

将所述音频输出夹角、音频输出分贝分别与调取的基准音频夹角、基准音频分贝进行差值计算,得到音频夹角差值以及音频分贝差值;

调取初始调整音频,其中,所述初始调整音频包括基准音频段以及穿插于所述基准音

频段中的夹角填充槽位、分贝填充槽位；

基于所述音频夹角差值、音频分贝差值分别对所述夹角填充槽位、分贝填充槽位进行填充,得到当前调整音频；

将当前调整音频作为更新后的预设采集指示音频,并将所述预设采集指示音频基于所述预设音频采集面进行对应所述基准音频分贝的输出。

3. 根据权利要求1所述的基于人工智能的语音情感分析方法,其特征在于,

对所述音频采集数据进行预处理,得到音频优化数据,包括:

基于预设频域滤波策略对所述音频采集数据进行噪声去除;

对经过噪声去除后的音频采集数据进行归一化处理;

响应于完成对所述音频采集数据的归一化处理,对完成归一化处理后的音频采集数据进行特征提取,得到与所述音频采集数据对应的音频优化数据。

4. 根据权利要求3所述的基于人工智能的语音情感分析方法,其特征在于,

基于预设频域滤波策略对所述音频采集数据进行噪声去除,包括:

基于语音活动检测对所述音频采集数据进行噪声类型的确定,得到与所述音频采集数据对应的噪声特性;

对所述音频采集数据分别进行包括数据分割、信号加窗的信号预处理,得到对应时域的音频采集数据;

对经过信号预处理的音频采集数据进行频域转换,得到对应所述频域的音频采集数据;

对对应所述频域的音频采集数据进行滤波处理,并将经过所述滤波处理后的音频采集数据进行时域转换,得到对应所述时域的音频采集数据;

对对应所述时域的音频采集数据进行性能评估,得到域所述音频采集数据对应的感知质量,并响应于所述感知质量满足预设输出条件,完成对所述音频采集数据的噪声去除。

5. 根据权利要求3所述的基于人工智能的语音情感分析方法,其特征在于,

对经过噪声去除后的音频采集数据进行归一化处理,包括:

对经过噪声去除后的音频采集数据进行分帧处理,并将经过分帧处理后的音频采集数据进行加窗处理;

对经过加窗处理后的音频采集数据进行均值、标准差的计算,并将得到的对应所述音频采集数据的均值、标准差代入公式对所述音频采集数据进行归一化处理。

6. 根据权利要求3所述的基于人工智能的语音情感分析方法,其特征在于,

响应于完成对所述音频采集数据的归一化处理,对完成归一化处理后的音频采集数据进行特征提取,得到与所述音频采集数据对应的音频优化数据,包括:

响应于完成对所述音频采集数据的归一化处理,对经过归一化处理后的音频采集数据进行分帧处理,并将经过分帧处理后的音频采集数据进行加窗处理;

对经过所述加窗处理后的音频采集数据进行快速傅立叶变换,以将对应时域的音频采集数据转换为对应频域的音频采集数据;

对对应频域的音频采集数据进行特征提取,得到对应多维特征矩阵的音频优化数据。

7. 根据权利要求1所述的基于人工智能的语音情感分析方法,其特征在于,

基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据

对应的情感信息,包括:

基于循环神经网络以及卷积神经网络构建初始情感分析模型;

获取训练数据集,其中,所述训练数据集中包括各标注了不同情感类别的语音数据;

基于所述训练数据集对所述初始情感分析模型进行训练,得到训练后的当前情感分析模型;

将所述音频优化数据输入至所述当前情感分析模型中进行情感分析,得到对应不同情感类别的各置信度;

将对应置信度最高的情感状态确定为与所述音频优化数据对应的情感信息。

8.一种基于人工智能的语音情感分析系统,其特征在于,包括:

音频采集模块,被配置为响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述语音采集任务进行对应目标用户的音频采集,得到音频采集数据,包括:

响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,并调取与所述音频采集任务对应的预设采集指示音频进行输出;

响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的测试音频数据;

基于所述测试音频数据对所述目标用户进行测试评分,并基于得到的音频测试分值确定与所述目标用户对应的音频属性;

基于所述测试音频数据确定所述目标用户与所述预设音频采集面之间的音频输出夹角,并基于音频输出夹角确定与所述目标用户对应的偏移评分;

基于所述测试音频数据确定与所述目标用户对应的音频输出分贝,并基于所述音频输出分贝确定与所述目标用户对应的音量评分;

对所述偏移评分以及音量评分进行融合计算,并将得到的与所述目标用户对应的音频测试分值与基准音频分值进行比较;

当所述音频测试分值小于所述基准音频分值时,将所述目标用户的输出属性确定为负向评价属性;

当所述音频测试分值大于等于所述基准音频分值时,将所述目标用户的输出属性确定为正向评价属性;

响应于所述目标用户的音频属性为正向评价属性,调取与所述音频采集任务对应的预设采集确定音频进行输出;

响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的音频采集数据;

或者,

响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出;

数据优化模块,被配置为对所述音频采集数据进行预处理,得到音频优化数据;

情感分析模块,被配置为基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息;

情感干预模块,被配置为基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。

基于人工智能的语音情感分析方法及系统

技术领域

[0001] 本发明涉及数据处理技术,尤其涉及一种基于人工智能的语音情感分析方法及系统。

背景技术

[0002] 在现代客户服务和心理健康评估中,了解用户的情绪状态对于提供个性化服务和及时干预至关重要。目前,情感分析主要依赖于文本数据,但忽视了语音中的情感信息。随着人工智能和深度学习技术的发展,语音情感分析成为可能,但现有系统在实时性和准确性上仍存在许多不足。

发明内容

[0003] 基于上述问题,提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的一种基于人工智能的语音情感分析方法及系统。

[0004] 根据本发明的一个方面,提供一种基于人工智能的语音情感分析方法,包括以下步骤:

[0005] 响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据;

[0006] 对所述音频采集数据进行预处理,得到音频优化数据;

[0007] 基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息;

[0008] 基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。

[0009] 可选地,在根据本发明的方法中,响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据,包括:

[0010] 响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,并调取与所述音频采集任务对应的预设采集指示音频进行输出;

[0011] 响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的测试音频数据;

[0012] 基于所述测试音频数据对所述目标用户进行测试评分,并基于得到的音频测试分值确定与所述目标用户对应的音频属性;

[0013] 响应于所述目标用户的音频属性为正向评价属性,调取与所述音频采集任务对应的预设采集确定音频进行输出;

[0014] 响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的音频采集数据;

[0015] 或者,

[0016] 响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出。

[0017] 可选地,在根据本发明的方法中,基于所述测试音频数据对所述目标用户进行测试评分,并基于得到的音频测试分值确定与所述目标用户对应的音频属性,包括:

[0018] 基于所述测试音频数据确定所述目标用户与所述预设音频采集面之间的音频输出夹角,并基于音频输出夹角确定与所述目标用户对应的偏移评分;

[0019] 基于所述测试音频数据确定与所述目标用户对应的音频输出分贝,并基于所述音频输出分贝确定与所述目标用户对应的音量评分;

[0020] 对所述偏移评分以及音量评分进行融合计算,并将得到的与所述目标用户对应的音频测试分值与基准音频分值进行比较;

[0021] 当所述音频测试分值小于所述基准音频分值时,将所述目标用户的输出属性确定为负向评价属性;

[0022] 当所述音频测试分值大于等于所述基准音频分值时,将所述目标用户的输出属性确定为正向评价属性。

[0023] 可选地,在根据本发明的方法中,响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出,包括:

[0024] 响应于所述目标用户的音频属性为负向评价属性,获取与所述目标用户对应的音频输出夹角以及音频输出分贝;

[0025] 将所述音频输出夹角、音频输出分贝分别与调取的基准音频夹角、基准音频分贝进行差值计算,得到音频夹角差值以及音频分贝差值;

[0026] 调取初始调整音频,其中,所述初始调整音频包括基准音频段以及穿插于所述基准音频段中的夹角填充槽位、分贝填充槽位;

[0027] 基于所述音频夹角差值、音频分贝差值分别对所述夹角填充槽位、分贝填充槽位进行填充,得到当前调整音频;

[0028] 将当前调整音频作为更新后的预设采集指示音频,并将所述预设采集指示音频基于所述预设音频采集面进行对应所述基准音频分贝的输出。

[0029] 可选地,在根据本发明的方法中,对所述音频采集数据进行预处理,得到音频优化数据,包括:

[0030] 基于预设频域滤波策略对所述音频采集数据进行噪声去除;

[0031] 对经过噪声去除后的音频采集数据进行归一化处理;

[0032] 响应于完成对所述音频采集数据的归一化处理,对完成归一化处理后的音频采集数据进行特征提取,得到与所述音频采集数据对应的音频优化数据。

[0033] 可选地,在根据本发明的方法中,基于预设频域滤波策略对所述音频采集数据进行噪声去除,包括:

[0034] 基于语音活动检测对所述音频采集数据进行噪声类型的确定,得到与所述音频采集数据对应的噪声特性;

[0035] 对所述音频采集数据分别进行包括数据分割、信号加窗的信号预处理,得到对应时域的音频采集数据;

[0036] 对经过信号预处理的音频采集数据进行频域转换,得到对应所述频域的音频采集数据;

[0037] 对对应所述频域的音频采集数据进行滤波处理,并将经过所述滤波处理后的音频采集数据进行时域转换,得到对应所述时域的音频采集数据;

[0038] 对对应所述时域的音频采集数据进行性能评估,得到域所述音频采集数据对应的感知质量,并响应于所述感知质量满足预设输出条件,完成对所述音频采集数据的噪声去除。

[0039] 可选地,在根据本发明的方法中,对经过噪声去除后的音频采集数据进行归一化处理,包括:

[0040] 对经过噪声去除后的音频采集数据进行分帧处理,并将经过分帧处理后的音频采集数据进行加窗处理;

[0041] 对经过加窗处理后的音频采集数据进行均值、标准差的计算,并将得到的对应所述音频采集数据的均值、标准差代入以下公式对所述音频采集数据进行归一化处理:

[0042] $[X_{\text{norm}}] = \frac{X - \mu}{\sigma};$

[0043] 其中,(X) 为经过加窗处理的音频采集数据,(μ) 为均值,(σ) 为标准差,(X_{norm}) 为经过归一化处理后的音频采集数据。

[0044] 可选地,在根据本发明的方法中,响应于完成对所述音频采集数据的归一化处理,对完成归一化处理后的音频采集数据进行特征提取,得到与所述音频采集数据对应的音频优化数据,包括:

[0045] 响应于完成对所述音频采集数据的归一化处理,对经过归一化处理后的音频采集数据进行分帧处理,并将经过分帧处理后的音频采集数据进行加窗处理;

[0046] 对经过所述加窗处理后的音频采集数据进行快速傅立叶变换,以将对应时域的音频采集数据转换为对应频域的音频采集数据;

[0047] 对对应频域的音频采集数据进行特征提取,得到对应多维特征矩阵的音频优化数据。

[0048] 可选地,在根据本发明的方法中,基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息,包括:

[0049] 基于循环神经网络以及卷积神经网络构建初始情感分析模型;

[0050] 获取训练数据集,其中,所述训练数据集中包括各标注了不同情感类别的语音数据;

[0051] 基于所述训练数据集对所述初始情感分析模型进行训练,得到训练后的当前情感分析模型;

[0052] 将所述音频优化数据输入至所述当前情感分析模型中进行情感分析,得到对应不同情感类别的各置信度;

[0053] 将对应置信度最高的情感状态确定为与所述音频优化数据对应的情感信息。

[0054] 根据本发明的又一个方面,提供一种基于人工智能的语音情感分析系统,包括:

[0055] 音频采集模块,被配置为响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据;

[0056] 数据优化模块,被配置为对所述音频采集数据进行预处理,得到音频优化数据;

[0057] 情感分析模块,被配置为基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息;

[0058] 情感干预模块,被配置为基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。

[0059] 根据本发明的方案,本发明可以通过获取目标用户的音频采集数据,并进一步对音频采集数据进行后续的预处理,以根据得到的音频优化数据来进行对目标用户的情感分析,从而对目标用户的情感状态进行确定,同时,根据获取的情感状态还能够确定是否需要进行相应的情感干预,以帮助目标用户能够处于正面情绪,提高对于目标用户的心理纠正效果;另外,对于客户服务方面来说,本发明能够实时监控客户情绪,优化客服沟通策略,提升客户满意度;对于心理健康方面来说,本发明能够监控和分析用户的情绪变化,辅助心理健康诊断和干预;对于教育领域来说,本发明能够通过分析学生的情感状态,优化教学方法,提高教学效果;对于智能家居来说,本发明能够通过理解和响应用户的情感,提高人机交互的自然性和亲和力;而对于市场调研来说,本发明则可以通过分析消费者的情感反应,改进产品和服务。

附图说明

[0060] 图1示出了根据本发明一个实施例的基于人工智能的语音情感分析方法的流程图;

[0061] 图2示出了本实施例的应用场景图;

[0062] 图3示出了根据本发明另一个实施例的基于人工智能的语音情感分析系统的结构框图。

具体实施方式

[0063] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0064] 在现代客户服务和心理健康评估中,了解用户的情绪状态对于提供个性化服务和及时干预至关重要。目前,情感分析主要依赖于文本数据,但忽视了语音中的情感信息。随着人工智能和深度学习技术的发展,语音情感分析成为可能,但现有系统在实时性和准确性上仍存在许多不足。

[0065] 为解决上述现有技术中存在的问题,发明人提出本发明的方案。本发明的一个实施例提供了一种基于人工智能的语音情感分析方法,该方法可以在计算设备中执行,其中,计算设备可以被理解为具有数据处理功能的终端,例如手机或者电脑。

[0066] 如图1所示,本实施例提出的基于人工智能的语音情感分析方法始于步骤S102,在步骤S102中,包括以下内容:

[0067] 响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据。

[0068] 例如,在本实施例中,对于任一目标用户进行情感分析主要是依据目标用户的语音来进行的,因此,在需要进行相应情感分析时,需要对目标用户的语音进行相应采集,而为了保证语音采集的准确性在进行采集前,目标用户可以发送相应的采集触发指令来建立相应的音频采集任务,从而使得能够进一步的基于该音频采集任务来进行音频采集,从而得到对应的音频采集数据。

[0069] 可以理解的是,在本实施例中,本实施例的语音情感分析方法可被烧结于相应的处理芯片中,并集成在对应的装置单元内,装置单元可具体为便携式音响,从而便于使用及收纳;上述的发送相应的采集触发指令可以具体为语音指令或者按键指令,当采集触发指令为语音指令时,可预先设置响应于的语音唤醒词,从而便于对处理芯片进行唤醒并进行相应的语音采集;而采集触发指令为按键指令时,则可以通过设置相应的物理按键来对处理芯片进行唤醒并进行相应的语音采集。

[0070] 例如,图2示出了本实施例的应用场景图,基于图2内容可以看出用户输出的音频被装置单元进行接收。

[0071] 进一步,在本实施例中,上述的“响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据”,还可以包括以下步骤:

[0072] 响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,并调取与所述音频采集任务对应的预设采集指示音频进行输出;

[0073] 响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的测试音频数据;

[0074] 基于所述测试音频数据对所述目标用户进行测试评分,并基于得到的音频测试分值确定与所述目标用户对应的音频属性;

[0075] 响应于所述目标用户的音频属性为正向评价属性,调取与所述音频采集任务对应的预设采集确定音频进行输出;

[0076] 响应于完成对所述预设采集指示音频的输出,触发预设音频采集面进行对应目标用户的音频采集,得到与所述目标用户对应的音频采集数据;

[0077] 或者,

[0078] 响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出。

[0079] 例如,在本实施例中,在进行相应情感分析时,处理芯片会持续处于工作状态,而当在任一触发时刻接收到对应的采集触发指令后,处理芯片会相应的建立音频采集任务,这时,也即表明开始准备进行相应的情感分析工作;在完成对于音频采集任务的建立后,则可以基于音频采集任务来调取对应的预设采集指示音频进行输出,在这里,预设采集指示音频可以被理解为用于指示目标用户发出相应声音以确定当前环境或者用户当前位置是否适合进行相应的情感评价,因此预设采集指示音频可以包括对应“请进行测试讲话”的音频内容;

[0080] 在完成对预设采集指示音频的输出后,对应的预设音频采集面即可开始进行音频采集,从而得到与目标用户对应的测试音频数据,由上述内容可知,处理芯片可以被集成在对应的装置单元内,因此,预设音频采集面可以被理解为对应装置单元的采集面;

[0081] 在得到与目标用户对应的测试音频数据后,则可以基于该内容进行相应的测试评分,在这里,测试音频数据可以被理解为目标用户随意讲话而形成的语音内容,其主要是用于测试获取的音频数据是否清楚或者准确,可以说明的是,由于声音的传播是具有方向性的,因此,当目标用户距离预设音频采集面过远,或者过于偏离时,都会影响相应的音频采集效果;

[0082] 在完成对应的测试评分后,则可以根据得到的音频测试分值来确定与目标用户对应的音频属性,其中,音频属性可以包括正向评价属性以及负向评价属性,正向评价属性可以被理解为清楚或者准确,而负向评价属性则可以被理解为不清楚或不准确的音频数据;

[0083] 当对应的音频属性为正向评价属性时,则表明目标用户输出的测试音频数据已经达标,可以开始正式的音频采集,这时,则可以触发预设音频采集面来进行对应目标用户的音频采集,得到用于进行情感评价的音频采集数据;

[0084] 而当对应的音频属性为负向评价属性时。则表明目标用户输出的测试音频数据的对应音频质量没有达标,这时,因可以基于该测试音频数据来创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出,帮助目标用户进行再次的音频测试。

[0085] 更进一步的,在本实施例中,上述的“基于所述测试音频数据对所述目标用户进行测试评分,并基于得到的音频测试分值确定与所述目标用户对应的音频属性”,还可以包括以下步骤:

[0086] 基于所述测试音频数据确定所述目标用户与所述预设音频采集面之间的音频输出夹角,并基于音频输出夹角确定与所述目标用户对应的偏移评分;

[0087] 基于所述测试音频数据确定与所述目标用户对应的音频输出分贝,并基于所述音频输出分贝确定与所述目标用户对应的音量评分;

[0088] 对所述偏移评分以及音量评分进行融合计算,并将得到的与所述目标用户对应的音频测试分值与基准音频分值进行比较;

[0089] 当所述音频测试分值小于所述基准音频分值时,将所述目标用户的输出属性确定为负向评价属性;

[0090] 当所述音频测试分值大于等于所述基准音频分值时,将所述目标用户的输出属性确定为正向评价属性。

[0091] 例如,在本实施例中,可以说明的是,音频质量的影响条件包括音频距离以及音频偏移,其中,音频距离会影响对应的音频分贝,而音频偏移则会影响音频的立体感,基于此,则在目标用户的音频属性为负向评价属性时,可以获取与目标用户对应的音频输出夹角以及音频输出分贝,在这里,对应的音频输出夹角以及音频输出分贝都可以基于前述获取的测试音频数据进行获取;

[0092] 在得到对应的音频输出夹角以及音频输出分贝后,则可以基于音频输出夹角以及音频输出分贝来分别进行偏移、音量的评分,从而得到对应的偏移评分以及音量评分,并进一步的进行融合计算,以此得到与目标用户对应的音量评分,在这里,对于偏移评分以及音量评分过程可基于随着偏移量变大、随着音量变大、而导致对应的评分变大的规律来进行评分;

[0093] 完成对于音频测试分值的获取后,通过调取预先存储的基准音频分值,并将音频测试分值与基准音频分值进行分值比较,从而来确定与目标用户对应的输出属性。

[0094] 更进一步的,在本实施例中,上述的“响应于所述目标用户的音频属性为负向评价属性,基于所述测试音频数据创建当前调整音频,以作为更新后的预设采集指示音频再次进行输出”,还可以包括以下步骤:

[0095] 响应于所述目标用户的音频属性为负向评价属性,获取与所述目标用户对应的音频输出夹角以及音频输出分贝;

[0096] 将所述音频输出夹角、音频输出分贝分别与调取的基准音频夹角、基准音频分贝进行差值计算,得到音频夹角差值以及音频分贝差值;

[0097] 调取初始调整音频,其中,所述初始调整音频包括基准音频段以及穿插于所述基准音频段中的夹角填充槽位、分贝填充槽位;

[0098] 基于所述音频夹角差值、音频分贝差值分别对所述夹角填充槽位、分贝填充槽位进行填充,得到当前调整音频;

[0099] 将当前调整音频作为更新后的预设采集指示音频,并将所述预设采集指示音频基于所述预设音频采集面进行对应所述基准音频分贝的输出。

[0100] 例如,在本实施例中,当目标用户对应的输出属性为负向评价属性时,则表明目标用户输出的测试音频数据是不达标的,这时,则需要对该目标用户进行相应的指示,以确保后续的情感分析具有较高的精准性,而对应的指示方式来如下所述:

[0101] 首先,可以基于前述得到的音频输出夹角、音频输出分贝来分别与调取的基准音频夹角、基准音频分贝进行差值,从而得到对应的音频夹角差值以及音频分贝差值,需要说明的是,在正常情况下,当目标用户正对于预设音频采集面进行音频输出时,所获取的音频数据的立体感是最优的,因此,可以基于此来设置相应的基准音频夹角;同理,对应的基准音频分贝可设置为正常水平下的音频数值即可,例如为50分贝;

[0102] 接着,在得到对应的音频夹角差值以及音频分贝差值后,则可以确定需要目标用户进行调整的音频范围,从而可以调取对应的初始调整音频,其中,初始调整音频主要包括基准音频段以及穿插于基准音频段中的夹角填充槽位以及分贝填充槽位,例如,初始调整音频可以为“尊敬的用户,您好,请将您的音频位置调整至「槽位A」、以及将音频分贝调整至「槽位B」”,槽位A即代表夹角填充槽位,槽位B即代表分贝填充槽位,通过将音频夹角差值以及音频分贝差值填充至对应的槽位中,从而可以实现对于音频的定制化处理,得到当前调整音频;

[0103] 最后,通过经过定制化的处理来获取到对应的当前调整音频后则可以将该当前调整音频进行基于预设音频采集面的输出,并且在输出过程中对应的音量可基于基准音频分贝,以供目标用户进行参考,从而便于目标用户进行相应的测试调整,以快速获取相应的音频采集数据。

[0104] 另外,对于上述提到的处理芯片可以基于以下的方式进行配置:

[0105] (1) 硬件配置

[0106] 选择合适的麦克风和声卡,确保采集设备能够高质量捕获用户的语音信号。

[0107] 麦克风:选择高灵敏度的麦克风,如电容麦克风或动态麦克风。

[0108] 声卡:选择高性能的声卡,确保低延迟和高精度的模数转换。

[0109] (2) 软件准备

[0110] 安装和配置相关的音频采集库,例如Python中的PyAudio库,提供对音频设备的接

口。

[0111] 安装PyAudio:

[0112] 在Python环境中安装PyAudio库

[0113] (3) 实现实时音频采集

[0114] 初始化音频流:

[0115] 使用PyAudio库初始化音频流,设置采样率、通道数和缓冲区大小等参数。

[0116] (4) 数据存储和管理

[0117] 实时存储与处理:

[0118] 实时存储采集到的音频数据,并进行简单的预处理,如降噪、分帧和加窗。

[0119] 保存到文件:

[0120] 将采集到的音频数据保存为音频文件,供后续处理和分析。

[0121] 在步骤S104中,包括以下内容:

[0122] 对所述音频采集数据进行预处理,得到音频优化数据。

[0123] 例如,在本实施例中,由于需要基于采集得到的音频采集数据来进行相应的情感分析,因此,需要对音频采集数据进行预处理操作,从而得到相应的音频优化数据。

[0124] 例如,在本实施例中,上述的“对所述音频采集数据进行预处理,得到音频优化数据”,可以包括以下内容:

[0125] 基于预设频域滤波策略对所述音频采集数据进行噪声去除;

[0126] 对经过噪声去除后的音频采集数据进行归一化处理;

[0127] 响应于完成对所述音频采集数据的归一化处理,对完成归一化处理后的音频采集数据进行特征提取,得到与所述音频采集数据对应的音频优化数据。

[0128] 下面,针对上述的预处理过程可作如下的详细介绍:

[0129] 首先,针对上述的“基于预设频域滤波策略对所述音频采集数据进行噪声去除”,可采用以下方式进行:

[0130] 基于语音活动检测对所述音频采集数据进行噪声类型的确定,得到与所述音频采集数据对应的噪声特性;

[0131] 对所述音频采集数据分别进行包括数据分割、信号加窗的信号预处理,得到对应时域的音频采集数据;

[0132] 对经过信号预处理的音频采集数据进行频域转换,得到对应所述频域的音频采集数据;

[0133] 对对应所述频域的音频采集数据进行滤波处理,并将经过所述滤波处理后的音频采集数据进行时域转换,得到对应所述时域的音频采集数据;

[0134] 对对应所述时域的音频采集数据进行性能评估,得到域所述音频采集数据对应的感知质量,并响应于所述感知质量满足预设输出条件,完成对所述音频采集数据的噪声去除。

[0135] 例如,上述的过程可通过以下方法步骤来实现:

[0136] (1) 噪声类型的识别

[0137] 在处理语音信号时,不同类型的噪声会对信号的质量产生不同的影响。常见的噪声包括:

[0138] 环境噪声:如背景讲话声、交通噪声等。

[0139] 设备噪声:如录音设备自身产生的噪声。

[0140] 电源噪声:如50Hz或60Hz的电源噪声(工频噪声)。

[0141] 瞬态噪声:如咳嗽声、键盘敲击声等。

[0142] 识别具体的噪声类型有助于选择合适的去噪方法。同时,也可以通过噪声估计方法,如基于语音活动检测(VAD)的噪声估计技术,来获取噪声特性。

[0143] (2) 信号预处理

[0144] 在进行频域处理之前,通常需要对语音信号进行一些基本的预处理操作,如分帧和加窗:

[0145] 分帧:将长时间的语音信号分割成若干帧,每帧大小通常为20-40 ms,帧移为10-20 ms。这样可以减少动态范围并便于短时傅里叶变换(STFT)的计算。

[0146] 加窗:在每一帧信号上应用一个窗函数(如Hamming窗或Hanning窗),减小频谱泄漏现象。窗口函数的公式为:

[0147] $[w[n] = 0.54 - 0.46 \cos(\frac{2\pi n}{N-1})]$;

[0148] (3) 频域变换

[0149] 将时域信号转换到频域,以便更有效地识别和处理噪声:

[0150] 短时傅里叶变换(STFT):

[0151] $[X(k, l) = \sum_{n=0}^{N-1} x(n + lH)w(n) e^{-j(2\pi kn/N)}]$;

[0152] 其中, $(x(n))$ 为输入信号, $(w(n))$ 为窗函数, (N) 为FFT点数, (H) 为帧移量, (k) 和 (l) 分别为频率和时间索引。

[0153] (4) 滤波器设计及应用

[0154] 根据噪声的频率特性,设计并应用适当的滤波器来去除噪声:

[0155] 带通滤波器:用于保留语音信号的主要频段(通常为300Hz至3400Hz),并去除低频(<300Hz)和低频(>3400Hz)噪声。带通滤波器的传递函数为:

[0156]
$$[H(f) = \begin{cases} 1, & f_{\text{low}} \leq f \leq f_{\text{high}} \\ 0, & \text{otherwise} \end{cases}];$$

[0157] 其中, (f_{low}) 和 (f_{high}) 分别为带通滤波器的低频和高频截止频率。

[0158] 自适应滤波器:如维纳滤波器,根据噪声和信号的统计特性自适应调整滤波器的系数。维纳滤波器的频域表达式为:

[0159] $[H_w(f) = \frac{S_x(f)}{S_x(f) + S_n(f)}];$

[0160] 其中, $(S_x(f))$ 和 $(S_n(f))$ 分别为信号和噪声的功率谱密度。

[0161] (5) 时域还原

[0162] 将经过滤波处理的频域信号转换回时域信号:

[0163] 逆短时傅里叶变换 (ISTFT):

$$[0164] \quad [x(n) = \text{ISTFT}(X(k, l)) = \sum_l \sum_k \\ = 0]^{N-1} X(k, l) W(n - lH) e^{j(2\pi k n / N)}];$$

[0165] 加窗和重叠相加:在逆变换过程中,每帧加窗,重叠部分相加,以还原完整的时域信号。

[0166] (6) 性能评估

[0167] 对去噪后的信号进行性能评估,以确保信号的质量:

[0168] 信噪比 (SNR):衡量去噪效果的一种常用量度。定义为去噪后信号的功率与残留噪声的功率之比:

$$[0169] \quad [\text{SNR} = 10 \log_{10} \left(\frac{\sum x_{\text{clean}}^2}{\sum (x_{\text{noisy}} - x_{\text{denoised}})^2} \right)];$$

[0170] 语音质量评估 (如PESQ):客观语音质量评估法,用于评估语音信号的感知质量。

[0171] 用户听觉评估:通过主观听觉测试,确保去噪后的语音信号满足用户的感知质量。

[0172] 通过上述几个步骤,可以有效地减少语音信号中的各种类型的噪声,确保语音信号的清晰度和准确性,为后续的特征提取和情感识别提供高质量的输入数据。

[0173] 其次,针对上述的“对经过噪声去除后的音频采集数据进行归一化处理”,可采用以下方式进行:

[0174] 对经过噪声去除后的音频采集数据进行分帧处理,并将经过分帧处理后的音频采集数据进行加窗处理;

[0175] 对经过加窗处理后的音频采集数据进行均值、标准差的计算,并将得到的对应所述音频采集数据的均值、标准差代入以下公式对所述音频采集数据进行归一化处理:

$$[0176] \quad [X_{\text{norm}} = \frac{X - \mu}{\sigma}];$$

[0177] 其中, (X) 为经过加窗处理的音频采集数据, (μ) 为均值, (σ) 为标准差, (X_{norm}) 为经过归一化处理后的音频采集数据。

[0178] 例如,上述的过程可通过以下方法步骤来实现:

[0179] (1) 数据准备

[0180] 在进行归一化处理之前,需要确保输入的数据是适合处理和分析的。具体步骤如下:

[0181] 分帧:将长时间的语音信号分割成若干帧。每帧的长度通常为20-40毫秒,帧移为10-20毫秒。这一步可以避免处理过长的信号片段。

[0182] 加窗:在每一帧信号上应用窗函数(如Hamming窗或Hanning窗),减少频谱泄漏现象。窗口函数的公式为:

$$[0183] \quad [w[n] = 0.54 - 0.46 \cos \left(\frac{2\pi n}{N-1} \right)];$$

[0184] (2) 均值和标准差计算

[0185] 计算语音信号的均值和标准差是进行归一化处理的基础步骤。

[0186] 计算均值:均值是信号样本的平均值,表示信号的集体中心位置。计算公式为:

[0187]
$$\mu = \frac{1}{N} \sum_{i=1}^N X_i;$$

[0188] 其中, (X_i) 为第 i 个样本, (N) 为样本总数。

[0189] 计算标准差: 标准差表示信号样本的离散程度, 计算公式为:

[0190]
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2};$$

[0191] (3) 归一化公式应用

[0192] 使用先前计算的均值和标准差, 将每个语音样本进行归一化处理。归一化将信号的振幅调整到一个标准范围, 通常为零均值和单位方差。

[0193] 归一化处理: 针对每一个样本应用如下公式进行归一化处理:

[0194]
$$X_{\text{norm}} = \frac{X - \mu}{\sigma};$$

[0195] (X) 为原始语音信号样本。

[0196] (μ) 为均值。

[0197] (σ) 为标准差。

[0198] (X_{norm}) 为归一化后的样本。

[0199] 归一化处理后的信号将具有零均值和单位方差, 这有助于后续特征的提取和模型的高效训练。

[0200] (4) 验证和性能评估

[0201] 在归一化处理后, 需要对处理结果进行验证和性能评估, 以确保信号质量满足要求。

[0202] 信号还原验证: 检查归一化后的信号是否能通过反归一化处理还原到原始信号。

[0203]
$$X_{\text{original}} = X_{\text{norm}} \cdot \sigma + \mu;$$

[0204] 性能评估指标: 使用偏度和峰度等统计指标评估信号的分布特性是否符合预期。

[0205] 偏度 (Skewness): 评估信号分布的对称性, 计算公式为:

[0206]
$$\text{Skewness} = \frac{\sum_{i=1}^N (X_i - \mu)^3}{(N-1)\sigma^3};$$

[0207] 峰度 (Kurtosis): 评估信号分布的平坦度, 计算公式为:

[0208]
$$\text{Kurtosis} = \frac{\sum_{i=1}^N (X_i - \mu)^4}{(N-1)\sigma^4} - 3;$$

[0209] 归一化能够提高信号处理和分析的稳定性和一致性, 确保语音特征在特征提取和模型训练中具有良好的表现。

[0210] 最后, 针对上述的“响应于完成对所述音频采集数据的归一化处理, 对完成归一化处理后的音频采集数据进行特征提取, 得到与所述音频采集数据对应的音频优化数据”, 可采用以下方式进行:

[0211] 响应于完成对所述音频采集数据的归一化处理, 对经过归一化处理后的音频采集数据进行分帧处理, 并将经过分帧处理后的音频采集数据进行加窗处理;

[0212] 对经过所述加窗处理后的音频采集数据进行快速傅立叶变换, 以将对应时域的音

频采集数据转换为对应频域的音频采集数据;

[0213] 对对应频域的音频采集数据进行特征提取,得到对应多维特征矩阵的音频优化数据。

[0214] 例如,上述的过程可通过以下方法步骤来实现:

[0215] 提取MFCC(梅尔频率倒谱系数)和其他有效的语音特征。

[0216] 可以分为以下几个关键点:

[0217] 分帧和加窗(Framing and Windowing)

[0218] 快速傅里叶变换(FFT) (Fast Fourier Transform)

[0219] 梅尔频率倒谱系数(MFCC) (Mel-Frequency Cepstral Coefficients)

[0220] Chroma特征 (Chroma Features)

[0221] 调谐频率振荡系数(Tonnetz) (Tonnetz Features)

[0222] 其他特征提取(如ZCR、STFT等)

[0223] 特征提取详细介绍

[0224] (1)分帧和加窗(Framing and Windowing)

[0225] 分帧和加窗是特征提取的第一步,用于将长时间的语音信号分割成小段,以便进一步处理。

[0226] 分帧:将语音信号分割成若干帧,每帧的长度通常为20-40毫秒,帧移为10-20毫秒。每帧信号包含足够的信息,但又不会太长,避免处理复杂。

[0227] $[x[n] = s[n - iM], \text{quad } 0 \leq i \leq N - 1];$

[0228] 其中,(s[n])是原始信号,(M)是帧移量,(N)是窗口长度。

[0229] 加窗:在每帧信号上应用窗函数(如Hamming窗或Hanning窗),减少频谱泄漏现象。窗口函数的公式为:

[0230] $[w[n] = 0.54 - 0.46 \cos(\frac{2\pi n}{N - 1})];$

[0231] (2)快速傅里叶变换(FFT) (Fast Fourier Transform)

[0232] FFT将时域信号转换为频域信号,用于分析信号的频率成分。

[0233] 计算FFT:对每帧应用快速傅里叶变换,得出频谱。

[0234] $[X(k) = \sum_{n=0}^{N-1} x[n] e^{-j \frac{2\pi k n}{N}}];$

[0235] 其中,(X(k))是频谱,(x[n])是输入信号,(N)是FFT点数。

[0236] 频谱计算:得到的频谱用于进一步的特征提取。

[0237] 3. 梅尔频率倒谱系数(MFCC) (Mel-Frequency Cepstral Coefficients)

[0238] MFCC是捕捉人类听觉特性的主要特征,步骤如下:

[0239] Mel频率变换:将频谱转换为Mel尺度,Mel尺度更符合人耳的听觉特性。

[0240] $[\text{Mel}(f) = 2595 \log_{10}(1 + \frac{f}{700})];$

[0241] Mel滤波器组:将Mel频率范围内的频谱能量通过一组三角滤波器进行加权求和,得到滤波后的频谱能量。

[0242] 对数压缩:对滤波后的频谱能量取对数压缩,增强小信号,并减少动态范围。

[0243] $[E_m = \log(S_m)];$

[0244] 其中, (S_m) 是滤波后的频谱能量, (E_m) 是对数压缩后的能量。

[0245] 离散余弦变换 (DCT): 对对数压缩后的能量应用 DCT, 得到 MFCC 特征。

$$[c_n = \sum_{m=0}^{M-1} E_m \cos\left(\frac{\pi n}{M}\left(m + \frac{1}{2}\right)\right)]$$

[0246] $= 0\}^{M-1} E_m \cos\left(\frac{\pi n}{M}\left(m + \frac{1}{2}\right)\right)$;

[0247] (4) Chroma 特征 (Chroma Features)

[0248] Chroma 特征反映了音符的能量分布, 有助于捕捉和声信息。

[0249] 计算 Chroma 频谱: 将频谱转换为 12 个等音高类能量值。

$$[C(p) = \sum_{k \in K(p)} |X(k)|^2];$$

[0251] 其中, $(C(p))$ 是第 p 个音高类的能量, $(X(k))$ 是频谱, $(K(p))$ 是属于第 p 个音高类的频率索引集合。

[0252] 归一化: 将 12 维 Chroma 特征归一化, 使得所有音高类的能量总和为 1。

[0253] (5) 调谐频率振荡系数 (Tonnetz) (Tonnetz Features)

[0254] Tonnetz 特征通过捕捉调谐频率关系, 用于分析音调感知。

[0255] 计算简谐振荡频率: 通过频谱分析得到简谐振荡频率。

[0256] 映射到时间序列: 将简谐振荡频率映射到时间序列特征, 用于捕捉音调变化信息。

[0257] (6) 其他特征提取 (如 ZCR、STFT 等)

[0258] 过零率 (ZCR, Zero-Crossing Rate): 用于描述信号变化的速率, 计算信号在零水平线上的过零次数。

$$[\text{ZCR} = \frac{1}{2N} \sum_{n=1}^N |\text{sign}(x[n]) - \text{sign}(x[n-1])|];$$

[0260] 短时傅里叶变换 (STFT, Short-Time Fourier Transform): 用于分析信号的时频特性。

$$[STFT\{x[n]\}(m, \omega) = \sum_{n=0}^{N-1} x[n] w[n-m] e^{-j\omega n}];$$

[0262] 在步骤 S106 中, 包括以下内容:

[0263] 基于预设情感分析策略对所述音频优化数据进行情感分析, 得到与所述音频优化数据对应的情感信息。

[0264] 例如, 在本实施例中, 在完成对于音频采集数据的预处理而得到对应的音频优化数据后, 则可以基于预存的预设情感分析策略来对应音频优化数据进行相应的情感分析, 从而得到与音频优化数据对应的情感信息。

[0265] 进一步的, 在本实施例中, 上述的“基于预设情感分析策略对所述音频优化数据进行情感分析, 得到与所述音频优化数据对应的情感信息”, 可以包括以下步骤:

[0266] 基于循环神经网络以及卷积神经网络构建初始情感分析模型;

[0267] 获取训练数据集, 其中, 所述训练数据集中包括各标注了不同情感类别的语音数据;

[0268] 基于所述训练数据集对所述初始情感分析模型进行训练, 得到训练后的当前情感

分析模型；

[0269] 将所述音频优化数据输入至所述当前情感分析模型中进行情感分析,得到对应不同情感类别的各置信度；

[0270] 将对应置信度最高的情感状态确定为与所述音频优化数据对应的情感信息。

[0271] 例如,上述的过程可通过以下方法步骤来实现：

[0272] 基于上述内容,在提取了MFCC、Chroma、Tonnetz等特征后,我们得到的是一个多维的特征矩阵。这个特征矩阵通常包含以下维度：

[0273] 时间帧数:信号被分割成的帧数。

[0274] 特征维度:每个特征向量的维度,如MFCC的维度。

[0275] 假设我们提取了 N 帧,每帧的特征向量维度为 D ,那么输入到神经网络模型的特征矩阵的形状为 (N, D) 。

[0276] 接下来,我们将特征矩阵输入到构建好的深度学习模型中。我们这里以一个包含CNN和RNN的混合模型为例。

[0277] 循环神经网络层(LSTM Layer)

[0278] 输入:池化后的特征图。

[0279] LSTM操作:处理时间序列特征,捕获长期依赖关系。

[0280] 输出:时间序列的隐藏状态向量,形状为 (num_lstm_units) 。

[0281] (1) 模型处理过程

[0282] 卷积层(Conv1D Layer)

[0283] 输入:特征矩阵 (N, D) 。

[0284] 卷积操作:使用多个卷积核扫描输入特征矩阵,提取局部特征。卷积核的大小和数量可以调整。

[0285] 激活函数:对卷积结果应用ReLU激活函数,增加非线性。

[0286] 输出:卷积后的特征图,形状为 $(N, \text{num_filters})$ 。

[0287] 池化层(Pooling Layer)

[0288] 输入:卷积后的特征图。

[0289] 池化操作:下采样操作通常用最大池化(MaxPooling),减小特征图尺寸,提高计算效率。

[0290] 输出:池化后的特征图,形状为 $(N//2, \text{num_filters})$ (如果池化窗口大小为2)。

[0291] 全连接层(Dense Layer)

[0292] 输入:LSTM的输出向量。

[0293] 全连接操作:将输入特征映射到高维空间。

[0294] 激活函数:应用ReLU激活函数。

[0295] 输出:全连接层的输出向量。

[0296] 输出层(Output Layer)

[0297] 输入:全连接层的输出向量。

[0298] Softmax函数:将输出映射到情感类别的概率分布。

[0299] 输出:每个情感类别的概率,形状为 (num_classes) 。

[0300] (2) 输出与情感识别结果

[0301] 概率分布与情感类别

[0302] 模型的输出是一个概率分布,表示每个情感类别的概率。通过选择概率最大的类别来确定当前的情感状态。

[0303] 置信度计算

[0304] 置信度是最高概率值,表示模型对预测结果的信心。

[0305] (3) 情感判断逻辑

[0306] 通过输出的情感类别和置信度,可以判断用户的情感状态。例如下示的表1 (置信度确定表):

[0307] 表1 置信度确定表

[0308]	情感类别	置信度范围	情感状态描述	;
	0	0.0-1.0	快乐	
	1	0.0-1.0	悲伤	
	2	0.0-1.0	愤怒	
	3	0.0-1.0	惊讶	

[0309] 通过置信度的高低,可以了解模型对当前预测结果的信心,置信度高则说明模型对当前情感判断较为确定。

[0310] 进一步,针对对于初始情感分析模型的构建以及训练过程可如下述:

[0311] 针对循环神经网络来说,包括:

[0312] (1) RNN的基本结构

[0313] 循环神经网络 (RNN) 是一种能够处理序列数据的神经网络,因为它们保留了先前输入的信息并使用这些信息来影响当前输入的处理结果。其基本结构包括:

[0314] 输入层:接收输入序列数据。

[0315] 隐藏层:RNN的核心部件,包含多个神经元,每个神经元都具有内部状态(记忆),并使用重复连接在时间步上共享参数。

[0316] 输出层:生成当前时间步的输出。

[0317] RNN的核心部分是其隐藏层,隐藏层的状态通过以下公式进行更新:

[0318]
$$[h_t = \phi(W_{xh}x_t + W_{hh}h_{t-1} + b_h)];$$

[0319] 其中:

[0320] (h_t) 是当前时间步的隐藏状态,

[0321] (x_t) 是当前时间步的输入,

[0322] (h_{t-1}) 是前一时间步的隐藏状态,

[0323] (W_{xh}) 和 (W_{hh}) 是需要学习的权重矩阵,

[0324] (b_h) 是偏置,

[0325] (ϕ) 是激活函数,如tanh或ReLU。

[0326] (2) 长短期记忆 (LSTM) 单元

[0327] LSTM单元是RNN的扩展版本,能够更好地捕捉长时间的依赖关系,通过引入门机制来控制信息流。

[0328] LSTM的结构包括以下几个主要部件:

[0329] 遗忘门(Forget Gate):控制前一状态信息的保留或丢弃。

[0330] $[f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)];$

[0331] 输入门(Input Gate):控制当前输入信息的写入。

[0332] $[i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)];$

[0333] 候选状态(Cell State Candidate):生成新的候选记忆细胞状态。

[0334] $[\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)];$

[0335] 输出门(Output Gate):控制输出到隐藏状态的信息。

[0336] $[o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)];$

[0337] LSTM单元的更新公式为:

[0338] 更新细胞状态:

[0339] $[C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t];$

[0340] 更新隐藏状态:

[0341] $[h_t = o_t \odot \tanh(C_t)];$

[0342] (3) 门控循环单元(GRU)

[0343] GRU是另一种改进的RNN单元,与LSTM类似但结构更简单,包含两个门:

[0344] 更新门(Update Gate):控制信息的更新。

[0345] $[z_t = \sigma(W_z \cdot [h_{t-1}, x_t])];$

[0346] 重置门(Reset Gate):控制前一状态信息的重置程度。

[0347] $[r_t = \sigma(W_r \cdot [h_{t-1}, x_t])];$

[0348] GRU的更新公式为:

[0349] 当前隐状态的候选版本:

[0350] $[\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t])];$

[0351] 更新隐藏状态:

[0352] $[h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t];$

[0353] (4) 数据输入和处理步骤

[0354] RNN部分处理的数据是一系列的序列数据,每个时间步输入一个数据点。

[0355] 输入数据:通常是由预处理模块提取的特征,如MFCC。每个时间步的输入表示为 (x_t) 。

[0356] 初始化隐藏状态:RNN需要初始化隐藏状态 (h_0) ,通常初始化为全零矢量。

[0357] 前向传播:对每个时间步,计算隐藏状态 (h_t) 和输出 (y_t) 。

[0358] 具体步骤如下:

[0359] 初始化:初始隐藏状态 ($h_0 = 0$)。

[0360] 时间步1:计算 ($h_1 = \text{RNN}(x_1, h_0)$),得到输出 (y_1)。

[0361] 时间步2:计算 ($h_2 = \text{RNN}(x_2, h_1)$),得到输出 (y_2)。

...

[0362] 时间步T:计算 ($h_T = \text{RNN}(x_T, h_{T-1})$),得到最终输出 (y_T)。

[0363] (5) 模型训练和优化过程

[0364] 损失函数:使用交叉熵损失函数衡量模型预测与实际标签之间的差异。

[0365] $[L = -\sum_{i=1}^N y_i \log(p_i)]$;

[0366] 其中, (y_i) 是实际标签 (one-hot编码), (p_i) 是模型预测概率。

[0367] 梯度计算:使用反向传播算法,通过时间反向传播 (Backpropagation Through Time, BPTT) 计算损失对各参数的梯度。

[0368]
$$[\frac{\partial L}{\partial \theta_t} = \sum_{t=1}^T \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial \theta_t}]$$

[0369] 优化算法:常用如Adam或RMSProp优化算法来更新模型参数。

[0370] $[\theta_{t+1} = \theta_t - \alpha \frac{\partial L}{\partial \theta_t}]$;

[0371] 其中, (α) 是学习率, (θ_t) 是在第t步的参数。

[0372] 正则化技术:为了防止过拟合,常用Dropout等正则化技术。

[0373] $[h_t = \text{Dropout}(h_t, p)]$;

[0374] 其中, (p) 是保留激活单元的概率。

[0375] 针对卷积神经网络,包括:

[0376] (1) CNN的基本概念与结构

[0377] 卷积神经网络 (CNN) 是一种专为处理数据的空间结构而设计的神经网络,主要包括卷积层 (Convolutional Layer)、池化层 (Pooling Layer) 和全连接层 (Fully Connected Layer)。

[0378] 卷积层:通过卷积操作提取局部特征,权重共享和稀疏连接使得卷积层适合处理高维数据。

[0379] 池化层:减少特征维度,提高计算效率,同时具备一定的抗噪能力。

[0380] 全连接层:将卷积和池化后的特征映射到目标空间,如情感分类任务中的情感标签。

[0381] (2) CNN各部分的详细流程

[0382] 卷积层 (Convolutional Layer)

[0383] 卷积层通过卷积运算提取输入数据的局部特征,其公式为:

[0384] $[y_{i,j} = \sum_{m,n} x_{i+m,j+n} \cdot w_{m,n} + b]$;

[0385] 其中:

[0386] $(y_{\{i,j\}})$ 是卷积输出。

[0387] $(x_{\{i+m,j+n\}})$ 是输入特征。

[0388] $(x_{\{i+m,j+n\}})$ 是输入特征。

[0389] (b) 是偏置。

[0390] 特征图的生成过程包括:

[0391] 卷积操作:使用多个卷积核扫描输入数据,生成一组特征图 (Feature Maps)。

[0392] 激活函数:对卷积结果应用激活函数 (如ReLU) 引入非线性:

[0393] $[a_{\{i,j\}} = \max(0, y_{\{i,j\}})]$;

[0394] 池化层 (Pooling Layer)

[0395] 池化层用于下采样特征图,提高特征表示的计算效率和抗噪能力。

[0396] 常见的池化操作有最大池化 (Max Pooling) 和平均池化 (Average Pooling)。

[0397] 最大池化:

[0398] $[p_{\{i,j\}} = \max(y_{\{i+m,j+n\}})]$;

[0399] 其中, $(y_{\{i+m,j+n\}})$ 是池化窗口内的元素。

[0400] 平均池化:

[0401] $[p_{\{i,j\}} = \frac{1}{Z} \sum y_{\{i+m,j+n\}}]$;

[0402] 其中, (Z) 是池化窗口的元素个数。

[0403] 全连接层 (Fully Connected Layer)

[0404] 全连接层将卷积和池化后的特征映射到目标空间。

[0405] 线性变换:

[0406] $[z_i = \sum_j w_{\{ij\}} \cdot a_j + b_i]$;

[0407] 其中, $(w_{\{ij\}})$ 是权重, (a_j) 是池化输出, (b_i) 是偏置。

[0408] 激活函数:

[0409] 对线性变换后的结果应用激活函数 (如ReLU或Softmax)。

[0410] (3) 数据输入和处理步骤

[0411] 为了使CNN能够处理语音数据,输入数据通常需要预处理和格式转换。

[0412] 输入数据:通常是由预处理和特征提取模块得到的特征,如MFCC、Chroma等。这些特征通常以二维矩阵表示 (如时间和频率维度)。

[0413] 输入格式转换:将输入特征转化为适合卷积处理的数据格式。对于语音数据,输入维度通常是 (通道数, 高度, 宽度), 如单通道二维矩阵。

[0414] 具体步骤如下:

[0415] 数据预处理:将原始语音信号转换为特征矩阵。

[0416] 卷积层处理:对输入特征矩阵进行卷积操作,提取局部特征。

[0417] 池化层处理:对卷积后的特征进行下采样,提高计算效率。

[0418] 全连接层输出:将池化后的特征映射到情感分类输出。

[0419] (4) 模型训练和优化过程

[0420] 损失函数

[0421] 常用的损失函数为交叉熵损失函数,用于衡量预测结果与实际标签之间的差异:

[0422] $[L = -\sum_{i=1}^N y_i \log(p_i)];$

[0423] 其中, (y_i) 是实际标签 (one-hot 编码), (p_i) 是模型预测概率。

[0424] 梯度计算和反向传播

[0425] CNN模型训练依赖于反向传播算法,通过计算损失对各参数的梯度来更新模型参数。

[0426] 前向传播:计算输入经过每一层的输出。

[0427] 反向传播:计算损失对每一层参数的梯度,并更新参数。

$$[\begin{aligned} & \frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial z_i} \cdot \frac{\partial z_i}{\partial w_{ij}} \cdot \frac{\partial L}{\partial a_j} \\ & = \sum_i \frac{\partial L}{\partial z_i} \cdot w_{ij} \end{aligned}]$$

[0429] 优化算法

[0430] 常用的优化算法包括SGD (随机梯度下降)、Adam等,通过调整学习率来更新模型参数。

[0431] SGD:

[0432] $[\theta_{t+1} = \theta_t - \alpha \frac{\partial L}{\partial \theta_t}];$

[0433] 其中, (α) 是学习率, (α) 是第t步的参数。

[0434] Adam:自适应学习率优化算法:

```

[
    \begin{align*}
        m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \nabla L \backslash \\
        v_t &= \beta_2 v_{t-1} + (1 - \beta_2) (\nabla L)^2 \backslash \\
        \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \backslash \\
[0435] \quad \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \backslash \quad ; \\
        \theta_{t+1} &= \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t}} \backslash \\
        &+ \epsilon \backslash \\
    \end{align*}
]

```

[0436] 正则化技术

[0437] 为了防止过拟合,常用的正则化技术包括Dropout、L2正则化等。

[0438] Dropout:在训练过程中随机丢弃一部分神经元,防止过拟合。

[0439] $[h_t = \text{Dropout}(h_t, p)]$;

[0440] 其中, (p) 是保留激活单元的概率。

[0441] L2正则化:在损失函数中加入权重的L2范数,以防止模型过拟合。

[0442] $[L_{\text{reg}} = L + \lambda \sum_{i,j} w_{ij}^2]$;

[0443] 其中, (λ) 是正则化系数。

[0444] 在步骤S108中,包括以下内容:

[0445] 基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。

[0446] 例如,在本实施例中,在获取上述的情感信息后,则可以该情感信息来对目标用户的情感状态进行确定,其中,情感状态可以包括正面情感、负面情感以及中性情感,正面情感可以包括快乐,负面情感可以包括悲伤、愤怒,而中性情感则可以包括平静;一般来说,当目标用户的情感状态为负面状态时,则需要基于该不良状态对目标用户的情感进行情感干预,以帮助改善目标用户的情感状态。

[0447] 综上所述,根据本发明的方案,本发明可以通过获取目标用户的音频采集数据,并进一步对音频采集数据进行后续的预处理,以根据得到的音频优化数据来进行对目标用户的情感分析,从而对目标用户的情感状态进行确定,同时,根据获取的情感状态还能够确定是否需要进行相应的情感干预,以帮助目标用户能够处于正面情绪,提高对于目标用户的心理纠正效果;另外,对于客户服务方面来说,本发明能够实时监控客户情绪,优化客服沟通策略,提升客户满意度;对于心理健康方面来说,本发明能够监控和分析用户的情绪变化,辅助心理健康诊断和干预;对于教育领域来说,本发明能够通过分析学生的情感状态,优化教学方法,提高教学效果;对于智能家居来说,本发明能够通过理解和响应用户的情感,提高人机交互的自然性和亲和力;而对于市场调研来说,本发明则可以通过分析消费者的情感反应,改进产品和服务。

[0448] 本发明的另一个实施例提供了一种基于人工智能的语音情感分析系统,图3为其对应系统框图,该系统包括:

[0449] 音频采集模块,被配置为响应于在任一触发时刻接收到采集触发指令,建立音频采集任务,以基于所述音频采集任务进行对应目标用户的音频采集,得到音频采集数据;

[0450] 数据优化模块,被配置为对所述音频采集数据进行预处理,得到音频优化数据;

[0451] 情感分析模块,被配置为基于预设情感分析策略对所述音频优化数据进行情感分析,得到与所述音频优化数据对应的情感信息;

[0452] 情感干预模块,被配置为基于所述情感信息确定与所述目标用户对应的情感状态,并基于所述情感状态确定是否需要与所述目标用户进行情感干预。

[0453] 在此处所提供的说明书中,算法和显示不与任何特定计算机、虚拟系统或者其它设备固有相关。各种通用系统也可以与本发明的示例一起使用。根据上面的描述,构造这类系统所要求的结构是显而易见的。此外,本发明也不针对任何特定编程语言。应当明白,可以利用各种编程语言实现在此描述的本发明的内容,并且上面对特定语言所做的描述是为了披露本发明的较佳实施方式。

[0454] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下被实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0455] 类似地,应当理解,为了精简本公开并帮助理解各个发明方面中的一个或多个,在上面对本发明的示例性实施例的描述中,本发明的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。

[0456] 本领域那些技术人员应当理解在本文所公开的示例中的设备的模块或单元或组件可以布置在如该实施例中所描述的设备中,或者可替换地可以定位在与该示例中的设备不同的一个或多个设备中。前述示例中的模块可以组合为一个模块或者此外可以分成多个子模块。

[0457] 本领域那些技术人员可以理解,可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在与该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件,以及此外可以把它分成多个子模块或子单元或子组件。

[0458] 此外,本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。

[0459] 此外,所述实施例中的一些在此被描述成可以由计算机系统的处理器或者由执行所述功能的其它装置实施的方法或方法元素的组合。因此,具有用于实施所述方法或方法元素的必要指令的处理器形成用于实施该方法或方法元素的装置。此外,装置实施例的在此所述的元素是如下装置的例子:该装置用于实施由为了实施该发明的目的的元素所执行的功能。

[0460] 如在此所使用的那样,除非另行规定,使用序数词“第一”、“第二”、“第三”等等来描述普通对象仅仅表示涉及类似对象的不同实例,并且并不意图暗示这样被描述的对象必须具有时间上、空间上、排序方面或者以任意其它方式的给定顺序。

[0461] 尽管根据有限数量的实施例描述了本发明,但是受益于上面的描述,本技术领域内的技术人员明白,在由此描述的本发明的范围内,可以设想其它实施例。此外,应当注意,本说明书中使用的语言主要是为了可读性和教导的目的而选择的,而不是为了解释或者限定本发明的主题而选择的。

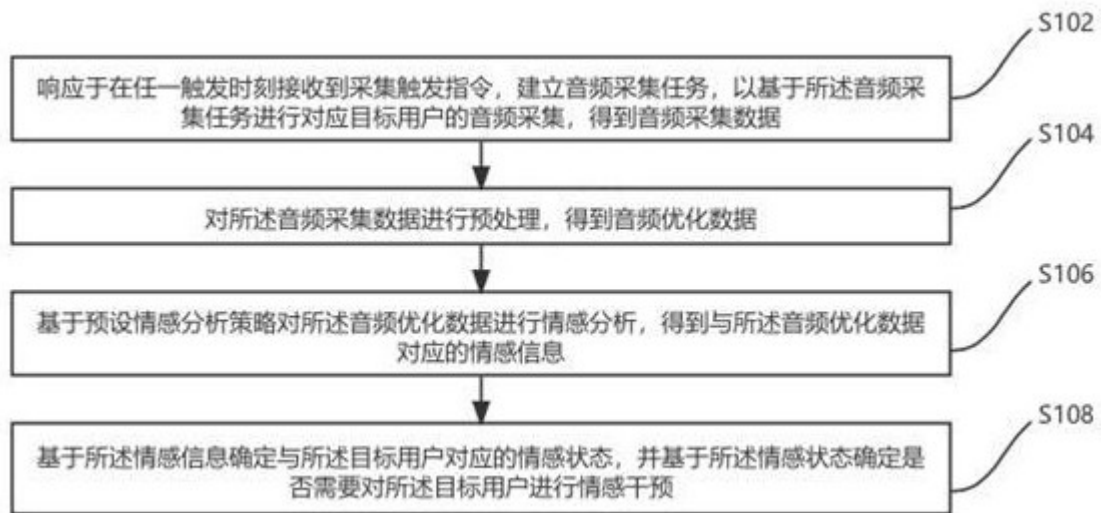


图 1



图 2

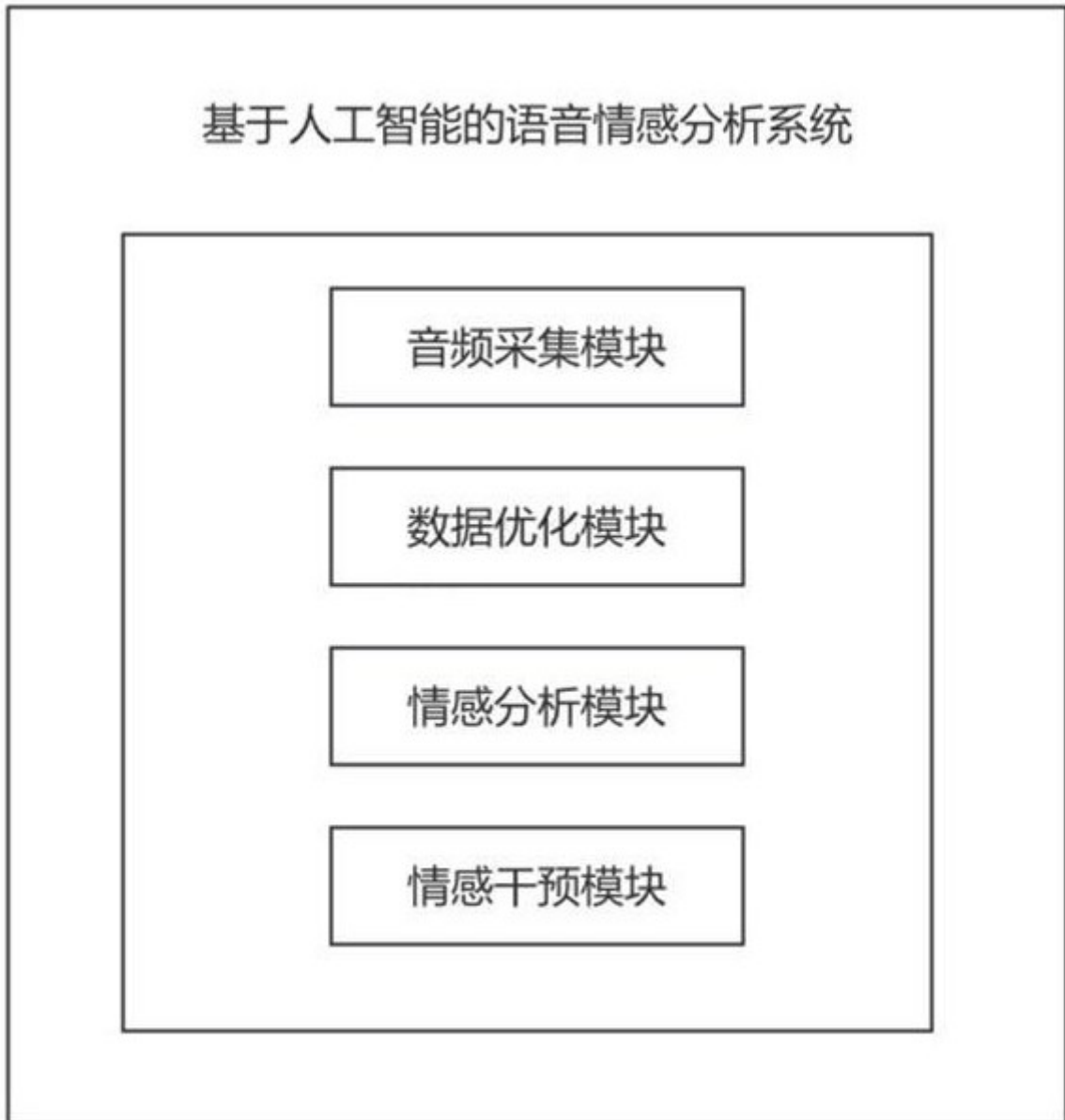


图 3