



(12) 发明专利申请

(10) 申请公布号 CN 113947127 A

(43) 申请公布日 2022. 01. 18

(21) 申请号 202111079583.8

A61B 5/00 (2006.01)

(22) 申请日 2021.09.15

A61B 5/16 (2006.01)

(71) 申请人 复旦大学

A61B 5/369 (2021.01)

地址 200433 上海市杨浦区邯郸路220号

A61B 5/372 (2021.01)

(72) 发明人 张立华 黄帅 杨鼎康 王顺利

G06V 10/82 (2022.01)

邝昊鹏

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(74) 专利代理机构 上海科盛知识产权代理有限

公司 31225

代理人 赵继明

(51) Int. Cl.

G06K 9/62 (2006.01)

G06V 40/16 (2022.01)

G10L 25/18 (2013.01)

G10L 25/45 (2013.01)

G10L 25/63 (2013.01)

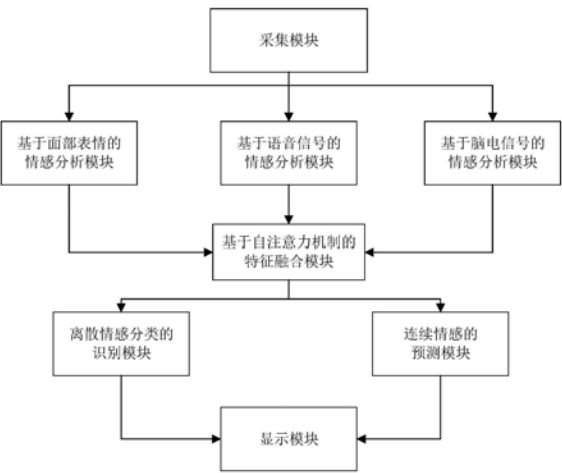
权利要求书3页 说明书9页 附图7页

(54) 发明名称

一种用于陪伴机器人的多模态情感识别方法和系统

(57) 摘要

本发明涉及一种用于陪伴机器人的多模态情感识别方法和系统,方法包括分别采集面部表情图片、语音信号和脑电信号;提取面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量;获取权重矩阵,将各特征向量与权重矩阵相乘,得到融合特征;通过支持向量机实现对于高兴、悲伤、平静和厌恶四类常见情感的分类;通过将情感演化为愉快度-紧张度-激动度-确定度这四个维度,采取多元非线性回归的方式进行情感评分的预测。与现有技术相比,本发明通过信息融合,具有更接近人类的情感识别能力;利用动态更新权重参数的形式,实现自主进化和不断调整情感判别的能力;离散和连续的情感识别,能够实现更加科学、深入的描绘情感变化。



1. 一种用于陪伴机器人的多模态情感识别方法,其特征在于,包括以下步骤:

分别采集面部表情图片、语音信号和脑电信号;

根据所述面部表情图片提取面部表情的情感特征向量,根据所述语音信号提取语音的情感特征向量,根据所述脑电信号提取脑电信号的特征向量;

获取权重矩阵,将所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量与该权重矩阵相乘,得到融合特征;

将所述融合特征载入预先构建并训练好的分类模型进行分类识别,得到多个离散的情感标签识别结果,所述分类模型的训练过程中还用于训练所述权重矩阵;

根据所述融合特征进行情感预测,该情感预测用于对融合特征进行数据拟合训练,得到连续的情感强度数值,该情感强度数值划分为多个情感维度,该多个情感维度包括愉快度、紧张度、激动度和确定度。

2. 根据权利要求1所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述提取面部表情的情感特征向量具体包括:

首先利用Adaboost算法提取面部表情图片中的Haar特征,构建Haar特征图,随后通过直方图均衡化对Haar特征图进行预处理,之后采用均匀模式LBP算法提取面部表情的情感特征向量。

3. 根据权利要求2所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述均匀模式LBP算法的提取过程包括:

构造大小3\*3的纹理区域,阈值为纹理区域的中心像素值,将周围的8个像素值与阈值进行比较,若值大于阈值像素的值,该像素所在区域处被置为1;若值小于阈值像素的值,该像素所在区域处则被置为0;在3\*3的纹理区域内,将其相邻的8个像素点产生的值按照顺时针的方向组成8位二进制数,统计8位二进制数中0到1或者1到0跳变的次数,若跳变数在两次以内,则该二进制数对应的十进制数就是该3\*3邻域中心的LBP值;若跳变的次数大于2,取 $P=8$ ,则将该区域中心的LBP值置为 $P+1=9$ ;遍历所有的像素点,得到整幅图像的LBP值,将所有的LBP值按顺序连接成一个特征向量,即为面部表情的情感特征向量。

4. 根据权利要求1所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述提取语音的情感特征向量具体包括:

首先对语音信号进行加窗处理,使用汉明窗进行平滑处理,将时域信号转化到频域进行后续的频谱分析;随后设计高通滤波器消除声带发音的噪声干扰,进行MFCC特征提取;将经过傅里叶变换的语谱图输入到预先构建并训练好的卷积神经网络层里,提取语谱图特征,得到语音的情感特征向量。

5. 根据权利要求1所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述提取脑电信号的特征向量具体包括:

首先对脑电信号进行预处理去噪,然后分别提取分型维数特征和多尺度熵特征,构建脑电信号的特征向量。

6. 根据权利要求5所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述预处理去噪包括:

通过固定的采样频率收集主体的脑电信号,随后选取db5小波进行多层小波分解,之后使用软阈值法将噪声产生的小波包系数置零,最后完成脑电信号的重构。

7. 根据权利要求6所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述分型维数特征的提取过程包括:

对原始序列进行均匀采样,获得K个序列,计算K个序列中各元素的变化量,构造新的序列,对该新的序列进行拟合,得到斜率,取该斜率的相反数作为FD初始特征;利用预处理去噪后的脑电信号做窗口处理,然后用窗口将数据分为多段,根据每段数据分别提取分型维数特征;

所述多尺度熵特征的提取过程包括:

计算脑电信号的多尺度熵并求在被试主体高兴情绪和悲伤情绪的平均多尺度熵值,随后选取前一个或多个尺度下的多尺度熵作为脑电信号的特征向量。

8. 根据权利要求1所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,所述分类识别前还包括:分别对所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量进行数据归一化后,得到融合特征,进行分类识别;所述分类模型采用SVM分类模型,该SVM分类模型的核函数为RBF核函数;所述情感标签包括高兴、悲伤、平静和厌恶。

9. 根据权利要求1所述的一种用于陪伴机器人的多模态情感识别方法,其特征在于,通过自注意力机制计算模态注意力,从而构建所述权重矩阵,得到融合权重,所述模态注意力的计算表达式为:

$$A = \text{softmax}(\frac{\Theta \cdot \Phi^T}{\sqrt{d}})$$

式中,A为模态注意力,  $(\cdot)$  为矩阵乘法,  $\Theta$  为查询矩阵,  $\Phi$  为键矩阵, T为转置符号, d为嵌入维数;

所述查询矩阵  $\Theta$  的构建过程为:将各模态的特征向量通过第一全连接层连接,该第一全连接层的连接算式为  $y_1 = w_1 x + b_1$ ,最后通过激活函数输出,得到特征量矩阵,构成查询矩阵  $\Theta$ ;该查询矩阵  $\Theta$  用于表示当前模态本身对其他模态的影响;

所述键矩阵  $\Phi$  的构建过程为:将各模态的特征向量通过第二全连接层连接,该第二全连接层的连接算式为  $y_2 = w_2 x + b_2$ ,最后通过激活函数输出,得到特征量矩阵,构成键矩阵  $\Phi$ ;该键矩阵  $\Phi$  用于表示其他模态对当前模态本身的影响;

将模态注意力A中每行的元素相加,得到模态i的权重,对应的计算公式为:

$$\Psi_i = \sum_k a_{ki}$$

式中,  $\Psi_i$  为模态i的权重,  $a_{ki}$  为模态注意力A中第k行第i列的元素;

所有模态的权重高的和为1,即满足:

$$\sum_i \Psi_i = 1$$

模态注意力A随分类模型的训练过程共同进行训练,以调整第一全连接层和第二全连接层中的参数  $w_1$ 、 $b_1$ 、 $w_2$  和  $b_2$ 。

10. 一种采用如权利要求1-9任一所述的一种用于陪伴机器人的多模态情感识别方法的系统,其特征在于,包括:

多模态采集模块,用于分别采集面部表情图片、语音信号和脑电信号;

基于面部表情的情感分析模块,用于根据所述面部表情图片提取面部表情的情感特征向量;

基于语音信号的情感分析模块,用于根据所述语音信号提取语音的情感特征向量;

基于脑电信号的情感分析模块,用于根据所述脑电信号提取脑电信号的特征向量;

基于自注意力机制的特征融合模块,用于获取权重矩阵,将所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量与该权重矩阵相乘,得到融合特征;

基于离散情感分类的识别模块,用于将所述融合特征载入预先构建并训练好的分类模型进行分类识别,得到多个离散的情感标签识别结果,所述分类模型的训练过程中还用于训练所述权重矩阵;

基于连续情感的预测模块,用于根据所述融合特征进行情感预测,该情感预测用于对融合特征进行数据拟合训练,得到连续的情感强度数值,该情感强度数值划分为多个情感维度,该多个情感维度包括愉快度、紧张度、激动度和确定度。

## 一种用于陪伴机器人的多模态情感识别方法和系统

### 技术领域

[0001] 本发明涉及情感识别技术领域,尤其是涉及一种用于陪伴机器人的多模态情感识别方法和系统。

### 背景技术

[0002] 情感交互在人机自然交互的研究中受到了很大的重视,而情感识别是人机情感交互的关键,其研究目的是让机器感知人类的情感状态,提高机器的人性化水平。多模态情感识别技术在陪伴机器人领域具有广泛的应用前景和研究价值。通过利用机器人所携带的多样传感器,从中获取诸如人类面部表情,行为,语音以及生理信号等蕴含潜在情感特征的多模式信号,基于深度学习方法实现特征提取和融合,并对人类的情感进行分析和预测,使得陪伴机器人具有更强的情感识别和情感理解能力。

[0003] 目前适用于陪伴机器人搭载情感识别装置,通常经由芯片、视频、音频等系统对人类感情变动所产生的生理特点、姿势、手势与语调等波形的变动展开详细分析与辨识,继而深刻理解人的情感并列出清晰适时的回应。基于面部表情的情感识别通常使用二维图像进行分析和研究,即基于脸部各器官和脸部凸起的位置特征的几何法、基于脸部纹理特征的像素法以及两者结合的混合法进行情感特征表示。基于语音的情感识别方法通常提取和归纳语音信号中的韵律信息和音质特征,包括梅尔频率倒谱系数(MFCC)、Teager能量算子等,通过支持向量机或者长短期记忆网络实现情感分类和识别。在生理信号方面,利用与情绪最相关的信号频段和脑区、脑电的时间稳定性特性等,基于传统的机器学习方式和脉冲神经网络进行情感理解。目前部分研究在情绪识别框架中整合来自各种行为表现和生理表现的特征进行情感识别。例如从联合头部运动和面部表情来推断行为相应的精神状态,继而推理人类的情感表达。同时还有通过结合表情和语音信号,利用共享模态之间表达子空间的方式实现离散的情感分类和识别。

[0004] 现有的大部分陪伴机器人都缺乏情感识别的功能。部分搭载特定传感器的机器人仅能实现基于单一模态的简易情感识别功能。仅仅基于面部表情或者语音信号的情感识别技术,没有考虑到不同模态获取的情感表达之间的互补性和促进性作用,在相对应的情感信息受到干扰或者信息获取不充分的情况,往往情感识别效率低下,无法满足情感交互的应用需求;同时,现有的技术手段大多针对于人类外在的情感表现特征进行识别,没有考虑到生理信号的监控和检测对于情感识别的重要意义。脑电和神经信号可以准确、客观和实时的反映异常情感和心理状态变化,有助于陪伴机器人针对不同的陪伴人群进行情感分析和达到精准的情感安慰的目的。

[0005] 近年来,陪伴机器人搭载的情感识别功能模块,往往只能对获取的模态信息进行简单的预处理,因此经常出现数据丢失和错误的问题;在此基础上,对于多模态的数据融合方式一般都采取数据集融合,即在不能保证数据完整性的情况下进行复杂繁琐的数据融合,造成了极大地数据资源浪费。此外,传统的方法大多数都采用离散的情感识别策略,没有充分的考虑到人类情感变化的连续性和异构性,因此情感识别的性能表现通常欠佳。

[0006] 综上所述,开发基于多模态特征获取和表达的方法,利用面部表情,语音以及生理信号的多模态情感表征数据,充分表达异构特征的判别力,克服现有的单模态数据研究困难,构建适用于陪伴机器人的多模态情感识别系统成为了本研究领域技术人员亟待解决的问题。

## 发明内容

[0007] 本发明的目的就是为了解决上述现有技术存在的缺陷而提供一种利用面部表情,语音以及生理信号的多模态情感表征数据,充分表达异构特征的判别力的用于陪伴机器人的多模态情感识别方法和系统。

[0008] 本发明的目的可以通过以下技术方案来实现:

[0009] 一种用于陪伴机器人的多模态情感识别方法,包括以下步骤:

[0010] 分别采集面部表情图片、语音信号和脑电信号;

[0011] 根据所述面部表情图片提取面部表情的情感特征向量,根据所述语音信号提取语音的情感特征向量,根据所述脑电信号提取脑电信号的特征向量;

[0012] 获取权重矩阵,将所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量与该权重矩阵相乘,得到融合特征;

[0013] 将所述融合特征载入预先构建并训练好的分类模型进行分类识别,得到多个离散的情感标签识别结果,所述分类模型的训练过程中还用于训练所述权重矩阵;

[0014] 根据所述融合特征进行情感预测,该情感预测用于对融合特征进行数据拟合训练,得到连续的情感强度数值,该情感强度数值划分为多个情感维度,该多个情感维度包括愉快度、紧张度、激动度和确定度。

[0015] 进一步地,所述提取面部表情的情感特征向量具体包括:

[0016] 首先利用Adaboost算法提取面部表情图片中的Haar特征,构建Haar特征图,随后通过直方图均衡化对Haar特征图进行预处理,之后采用均匀模式LBP算法提取面部表情的情感特征向量。

[0017] 进一步地,所述均匀模式LBP算法的提取过程包括:

[0018] 构造大小3\*3的纹理区域,阈值为纹理区域的中心像素值,将周围的8个像素值与阈值进行比较,若值大于阈值像素的值,该像素所在区域处被置为1;若值小于阈值像素的值,该像素所在区域处则被置为0;在3\*3的纹理区域内,将其相邻的8个像素点产生的值按照顺时针的方向组成8位二进制数,统计8位二进制数中0到1或者1到0跳变的次数,若跳变数在两次以内,则该二进制数对应的十进制数就是该3\*3邻域中心的LBP值;若跳变的次数大于2,取 $P=8$ ,则将该区域中心的LBP值置为 $P+1=9$ ;遍历所有的像素点,得到整幅图像的LBP值,将所有的LBP值按顺序连接成一个特征向量,即为面部表情的情感特征向量。

[0019] 进一步地,所述提取语音的情感特征向量具体包括:

[0020] 首先对语音信号进行加窗处理,使用汉明窗进行平滑处理,将时域信号转化到频域进行后续的频谱分析;随后设计高通滤波器消除声带发音的噪声干扰,进行MFCC特征提取;将经过傅里叶变换的语谱图输入到预先构建并训练好的卷积神经网络层里,提取语谱图特征,得到语音的情感特征向量。

[0021] 进一步地,所述提取脑电信号的特征向量具体包括:

[0022] 首先对脑电信号进行预处理去噪,然后分别提取分型维数特征和多尺度熵特征,构建脑电信号的特征向量。

[0023] 进一步地,所述预处理去噪包括:

[0024] 通过固定的采样频率收集主体的脑电信号,随后选取db5小波进行多层小波分解,之后使用软阈值法将噪声产生的小波包系数置零,最后完成脑电信号的重构。

[0025] 进一步地,所述分型维数特征的提取过程包括:

[0026] 对原始序列进行均匀采样,获得K个序列,计算K个序列中各元素的变化量,构造新的序列,对该新的序列进行拟合,得到斜率,取该斜率的相反数作为FD初始特征;利用预处理去噪后的脑电信号做窗口处理,然后用窗口将数据分为多段,根据每段数据分别提取分型维数特征;

[0027] 所述多尺度熵特征的提取过程包括:

[0028] 计算脑电信号的多尺度熵并求在被试主体高兴情绪和悲伤情绪的平均多尺度熵值,随后选取前一个或多个尺度下的多尺度熵作为脑电信号的特征向量。

[0029] 进一步地,所述分类识别前还包括:分别对所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量进行数据归一化后,得到融合特征,进行分类识别;所述分类模型采用SVM分类模型,该SVM分类模型的核函数为RBF核函数;所述情感标签包括高兴、悲伤、平静和厌恶。

[0030] 进一步地,通过自注意力机制计算模态注意力,从而构建所述权重矩阵,得到融合权重,所述模态注意力的计算表达式为:

$$[0031] \quad A = \text{softmax}\left(\frac{\Theta \cdot \Phi^T}{\sqrt{d}}\right)$$

[0032] 式中,A为模态注意力,  $(\cdot)$  为矩阵乘法,  $\Theta$  为查询矩阵,  $\Phi$  为键矩阵, T为转置符号, d为嵌入维数;

[0033] 所述查询矩阵  $\Theta$  的构建过程为:将各模态的特征向量通过第一全连接层连接,该第一全连接层的连接算式为  $y_1 = w_1 x + b_1$ ,最后通过激活函数输出,得到特征量矩阵,构成查询矩阵  $\Theta$ ;该查询矩阵  $\Theta$  用于表示当前模态本身对其他模态的影响;

[0034] 所述键矩阵  $\Phi$  的构建过程为:将各模态的特征向量通过第二全连接层连接,该第二全连接层的连接算式为  $y_2 = w_2 x + b_2$ ,最后通过激活函数输出,得到特征量矩阵,构成键矩阵  $\Phi$ ;该键矩阵  $\Phi$  用于表示其他模态对当前模态本身的影响;

[0035] 将模态注意力A中每行的元素相加,得到模态i的权重,对应的计算公式为:

$$[0036] \quad \Psi_i = \sum_k a_{ki}$$

[0037] 式中,  $\Psi_i$  为模态i的权重,  $a_{ki}$  为模态注意力A中第k行第i列的元素;

[0038] 所有模态的权重高的和为1,即满足:

$$[0039] \quad \sum_i \Psi_i = 1。$$

[0040] 模态注意力A随分类模型的训练过程共同进行训练,以调整第一全连接层和第二全连接层中的参数  $w_1$ 、 $b_1$ 、 $w_2$  和  $b_2$ 。

[0041] 本发明还提供一种用于陪伴机器人的多模态情感识别系统,包括:

[0042] 多模态采集模块,用于分别采集面部表情图片、语音信号和脑电信号;

[0043] 基于面部表情的情感分析模块,用于根据所述面部表情图片提取面部表情的情感

特征向量；

[0044] 基于语音信号的情感分析模块,用于根据所述语音信号提取语音的情感特征向量；

[0045] 基于脑电信号的情感分析模块,用于根据所述脑电信号提取脑电信号的特征向量；

[0046] 基于自注意力机制的特征融合模块,用于获取权重矩阵,将所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量与该权重矩阵相乘,得到融合特征；

[0047] 基于离散情感分类的识别模块,用于将所述融合特征载入预先构建并训练好的分类模型进行分类识别,得到多个离散的情感标签识别结果,所述分类模型的训练过程中还用于训练所述权重矩阵；

[0048] 基于连续情感的预测模块,用于根据所述融合特征进行情感预测,该情感预测用于对融合特征进行数据拟合训练,得到连续的情感强度数值,该情感强度数值划分为多个情感维度,该多个情感维度包括愉快度、紧张度、激动度和确定度。

[0049] 与现有技术相比,本发明具有以下优点：

[0050] 不同于以往的单一模态情感识别系统,本发明充分结合面部表情、语音信号以及脑电信号进行情感分析和判别,通过多模态的信息融合,能够增强情感特征的变大能力,完善情感表征空间的映射能力,使得机器人展现出更接近人类的情感识别能力。

[0051] 同时,本发明设计的基于自注意力机制的多模态融合模式,具有自适应性和灵活性,通过结合不同异构模态的情感表达优势,利用动态更新权重参数的形式,使得机器人具备自主进化和不断调整情感判别的能力,极大的为人机情感交互提供了新范式。

[0052] 此外,基于离散情感分类和连续情感维度预测的情感识别模式的联合处理,有效的刻画了多模态情感状态,能够更加科学,深入的描绘情感变化,对于机器人广泛的理解人类情感具有明显的优势。该系统在面对复杂的非线性多模态信息处理时减少了大量信息的损耗,且在处理模态跨度较大的数据时表现较好。

## 附图说明

[0053] 图1为本发明实施例中提供的一种用于陪伴机器人的多模态情感识别系统的原理框图；

[0054] 图2为本发明实施例中提供的一种基于面部表情的情感分析模块的原理框图；

[0055] 图3为本发明实施例中提供的一种基于语音信号的情感分析模块的原理框图；

[0056] 图4为本发明实施例中提供的一种基于脑电信号的情感分析模块的原理框图；

[0057] 图5为本发明实施例中提供的一种基于自注意力机制的特征融合模块的原理框图；

[0058] 图6为本发明实施例中提供的一种基于离散情感分类的识别模块的原理框图；

[0059] 图7为本发明实施例中提供的一种基于连续情感的预测模块的原理框图。

## 具体实施方式

[0060] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是



本发明一部分实施例,而不是全部的实施例。通常在此处附图中描述和示出的本发明实施例的组件可以以各种不同的配置来布置和设计。

[0061] 因此,以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围,而是仅仅表示本发明的选定实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0062] 实施例1

[0063] 如图1所示,本实施例提供一种用于陪伴机器人的多模态情感识别系统,包括:

[0064] 多模态采集模块,用于分别采集面部表情图片、语音信号和脑电信号;

[0065] 基于面部表情的情感分析模块,用于根据所述面部表情图片提取面部表情的情感特征向量;

[0066] 基于语音信号的情感分析模块,用于根据所述语音信号提取语音的情感特征向量;

[0067] 基于脑电信号的情感分析模块,用于根据所述脑电信号提取脑电信号的特征向量;

[0068] 基于自注意力机制的特征融合模块,用于获取权重矩阵,将所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量与该权重矩阵相乘,得到融合特征;

[0069] 基于离散情感分类的识别模块,用于将所述融合特征载入预先构建并训练好的分类模型进行分类识别,得到多个离散的情感标签识别结果,所述分类模型的训练过程中还用于训练所述权重矩阵;

[0070] 基于连续情感的预测模块,用于根据所述融合特征进行情感预测,该情感预测用于对融合特征进行数据拟合训练,得到连续的情感强度数值,该情感强度数值划分为多个情感维度,该多个情感维度包括愉快度、紧张度、激动度和确定度;

[0071] 显示模块,用于对离散情感分类和连续情感预测的结果进行实时显示。

[0072] 可总结为,多模态信息来源于陪伴机器人通过采集模块获取的面部表情、语音信号以及脑电信号。在识别模块通过不同模态的识别单元实现对多模态信息的预处理、数据特征提取以及模态输出等操作。在融合模块通过引入自注意力机制为不同模态计算注意力权重系数,从而实现特征级别的模态融合,通过情感分类和预测模块后,将多模态情感识别结果利用显示模块进行结果展示。

[0073] 下面对各模块分别进行具体描述。

[0074] 1、多模态采集模块

[0075] 多模态采集模块包括面部表情的情感采集装置、语音信号的情感采集装置、脑电信号的情感采集装置以及生理信号的情感采集装置等,通过多传感器实现对陪伴机器人针对的陪伴主体的多模态数据的收集操作。

[0076] 2、基于面部表情的情感分析模块

[0077] 图2为本发明提供的基于面部表情的情感分析模块的原理框图,该模块包含人脸检测、图像与处理、特征提取以及模态输出四个单元,基于不同的单元,其具体步骤如下:

[0078] (1) 首先利用Adaboost结合Haar特征实现目标检测。其中的Adaboost算法提取面部的Haar-like特征,包括输入图像的矩形特征。Haar特征是一种反映图像的灰度变化的一

种特征,用黑白两种矩形组合成特征模板,通过用黑色矩形区域的像素之和减去白色矩形区域像素之和计算模板的Haar特征值。常用特征有边缘特征、线特征、中心环绕特征以及对角线特征,人脸的五官轮廓与脸部有颜色差,可以用Haar特征描述人脸灰度变化。为了实现快速计算,使用积分图的方式。积分图可以快速计算出图像中任意一个矩形区域的像素之和,从而快速计算出图像的Haar特征。

[0079] (2) 图像预处理可以将图像中的有用信息恢复出来,降低图像中无关的信息,这里利用直方图均衡化将图像的直方图分布变成近似均匀分布,从而增强图像的对比度。

[0080] (3) 在特征提取单元,考虑到适用于陪伴机器人的系统低复杂度要求,采用均匀模式LBP算法提取人脸表情特征,通过构造3\*3的纹理区域,阈值为纹理区域的中心像素值,将周围的8个像素值与阈值进行比较,若值大于阈值像素的值,该邻域处被置为1;若值小于阈值像素的值,该邻域处则被置为0。在3\*3的区域内,将其相邻的8个像素点产生的值按照顺时针的方向组成8位二进制数。统计8位二进制数中0到1或者1到0跳变的次数,若跳变数在两次以内,则该二进制数对应的十进制数就是该3\*3邻域中心的LBP值;若跳变的次数大于2,取 $P=8$ ,则将该区域中心的LBP值置为 $P+1=9$ ;遍历所有的像素点,得到整幅图像的LBP值,将所有的LBP值按顺序连接成一个特征向量,即为整幅图的均匀模式LBP特征。

[0081] (4) 得到面部表情的情感特征向量后,通过模态输出单元进行模态预存储和输出。

[0082] 3、基于语音信号的情感分析模块

[0083] 图3为本发明提供的基于语音信号的情感分析模块的原理框图,其主要包括数据预处理、MFCC特征提取、语谱图特征提取以及模态输出单元。所述单元的基本步骤如下:

[0084] (1) 为使得获取的原始信号更加平滑,对信号进行加窗处理,使用汉明窗进行平滑处理,将时域信号转化到频域进行后续的频谱分析。

[0085] (2) 对于MFCC特征提取阶段,将语音信号转化到梅尔频率,赫兹频率,随后进行倒频分析,该方法的思路是人在说话的过程中,由于嘴唇和声带产生的其他干扰,为了解决这种干扰,步长语音信号受到发音系统所压抑的高频部分,并且能凸显出高频的共振峰,为此增加一个高通滤波器。通过在频域上增加系数,且和频率成正相关,进而提升高频的幅值。

[0086] (3) 语谱图特征提取单元中,将经过傅里叶变换的语谱图输入到卷积神经网络层里,用来提取语谱图特征,该过程包含输入层、池化层、卷积层和全连接层。语谱图特征提取的CNN结构,其中包含输入层、2个卷积层、2个池化层、一个全连接层。输入的图像大小为128\*128像素,第一层的卷积层是由5\*5的64个卷积核组成,经过卷积后引入非线性单元ReLU激活函数,经过卷积后连接由2\*2的卷积核构成pooling层,连接此层的目的是为了降低计算复杂度并提取主要特征。第二层卷积是由5\*5的128个卷积核组成。然后连接ReLU激活函数,在第二层卷积层后连接2\*2的pooling层,经过第二个池化层后输出得到128个尺寸为32\*32的像素特征图,最后接入由512个神经元组成的全连接层,最后则得到512维的特征向量。

[0087] (4) 得到语音的情感特征向量后,通过模态输出单元进行模态预存储和输出。

[0088] 4、基于脑电信号的情感分析模块

[0089] 图4为本发明提供的基于脑电信号的情感分析模块的原理框图,所述模块主要包括预处理去噪、分型维数特征提取、多尺度熵特征提取以及模态输出单元。具体步骤如下:

[0090] (1) 首先对脑电信号进行预处理去噪,信号来自采集模块去除陪伴者的FP1通道的

脑电信号,长度为采样率为128Hz,信号采集的时间长度为63s,去除前3s的基线时间,信号时间的60s,故采样点总数为7680。系统选取db5小波进行5层小波分解,之后使用软阈值法将噪声产生的小波包系数置零,最后完成脑电信号的重构。经过预处理前后的EEG信号可以更加平滑,适合进一步的处理。

[0091] (2) 在分型维数特征提取部分,将脑电信号的长度固定为N,随后均匀采样出不同的K值,之后构造出新的K个序列,再从K个序列中选取相邻两个数之间的间隔,构造出新的序列,然后计算所有新序列的长度,再使用最小二乘法拟合得到斜率,最后去斜率的相反数即为其所要求的FD初始特征。接下来将小波阈值去噪后的脑电信号做窗口处理,用长度为256点的窗口将数据分段,可以将数据分为30段,每段数据提取分型维数特征,可以得到30维特征向量。

[0092] (3) 在多尺度熵特征提取单元中,多尺度熵从不同时间尺度分析时间序列的复杂性。为了计算不同时间尺度下的样本熵,需要对原始信号进行粗粒化处理。粗粒化就是使用非重叠长度为i的窗口将原始信号分段。不同的尺度下得到的样本熵值也会不同,所构成的多尺度熵特征的维数也不同。本系统选取陪伴检测者FP1通道的脑电信号,提取他们所有实验得到的脑电数据,将自我评价中Valence维度大于6的样本判定为高兴情绪,小于4则判定为悲伤情绪。计算脑

[0093] 电信号的多尺度熵并求在实验中高兴情绪和悲伤情绪的平均多尺度熵值,随后选取前几个尺度下的多尺度熵作为脑电信号的特征向量,一共可以得到15维的多尺度熵特征。

[0094] (4) 得到分型维数特征和多尺度熵特征后,通过模态输出单元进行模态预存储和输出。

[0095] 5、基于自注意力机制的特征融合模块

[0096] 图5为本发明提供的基于自注意力机制的特征融合模块的原理框图,所述的注意力计算机制主要过程为信息输入、计算注意力分布以及计算输入信息的加权平均。通过提取到的面部表情、语音以及脑电信号等多模态特征后,进行有效融合,首先将一个权重矩阵初始化,用来表示各个模态特征的权重值,在特征融合的过程中将各个模态的权重值和对应的特征向量进行相乘,相乘之后在进行级联。且在整个系统模型的训练过程中,此权重矩阵会跟随着系统模型训练,根据训练对应的值也会不断变换调整。相较于人工固定的权重值,效果也就更好。

[0097] 特征融合模块包含两种类型的输入之间的交互: $\alpha_{ii}$ 分别是模态i的自我注意交互。 $\alpha_{ij}$ 是模态之间注意交互作用,反映了模态i对模态j的影响。模态注意力的计算如下所示:

$$[0098] \quad A = \text{softmax}\left(\frac{\Theta \cdot \Phi^T}{\sqrt{d}}\right)$$

[0099] 式中,A为模态注意力, $(\cdot)$ 为矩阵乘法, $\Theta$ 为查询矩阵, $\Phi$ 为键矩阵,T为转置符号,d为嵌入维数;

[0100] 查询矩阵 $\Theta$ 的构建过程为:将各模态的特征向量通过第一全连接层连接,该第一全连接层的连接算式为 $y_1 = w_1 x + b_1$ ,最后通过激活函数输出,得到特征量矩阵,构成查询矩阵 $\Theta$ ;

[0101] 键矩阵 $\Phi$ 的构建过程为:将各模态的特征向量通过第二全连接层连接,该第二全

连接层的连接算式为 $y_2 = w_2 x + b_2$ ,最后通过激活函数输出,得到特征量矩阵,构成键矩阵 $\Phi$ ;

[0102] 将模态注意力A中每行的元素相加,得到模态i的权重,对应的计算公式为:

[0103]  $\Psi_i = \sum_k a_{ki}$

[0104] 式中, $\Psi_i$ 为模态i的权重, $a_{ki}$ 为模态注意力A中第k行第i列的元素;

[0105] 所有模态的权重高的和为1,即满足:

[0106]  $\sum_i \Psi_i = 1$ 。

[0107] 模态注意力A随SVM分类模型的训练过程,共同进行训练,调整第一全连接层和第二全连接层中的参数 $w_1$ 、 $b_1$ 、 $w_2$ 和 $b_2$ 。

[0108] 6、基于离散情感分类的识别模块

[0109] 图6为本发明提供的基于离散情感分类的识别模块的原理框图,计算出多模态的特征向量之后,通常需要对特征向量进行数据归一化。在使用数据进行SVM分类时,归一化后的数据跟原始数据相比,不仅训练时间缩短,而且测试正确率提高,数据归一化使数据更加紧凑,有利于最优分类超平面的得出。系统使用svm-scale对数据进行相应的缩放,将数据大小缩放到[0,1]或者[-1,1],缩放的目的是防止某个特征过大或过小,加快计算速度,有利于训练模型。系统选择RBF核函数作为SVM分类算法的核函数,RBF作为一种对应于非线性映射的一种特例,可以处理非线性可分问题,适合处理多维的向量。通过模型训练后,该识别模块将正确输出对应的高兴、悲伤、平静和厌恶四种情感标签中的一种。

[0110] 7、基于连续情感的预测模块

[0111] 图7为本发明提供的基于连续情感的预测模块的原理框图,所述的模块中首先对连续维度的情感预测定义为愉快度-紧张度-激动度-确定度,随后通过标准量化的方式将情感预测定义为0-10的取值范围,紧接着利用多元非线性回归的方法对多模态的情感特征进行数据拟合训练,最终模块将输出四个不同维度所对应的情感强度数值。

[0112] 本实施例展示的用于陪伴机器人的多模态情感识别系统很好的解决了陪伴机器人的情感交互和识别问题,通过充分采集的面部表情、语音信号,难点信号等多模态信息,结合不同识别模块实现异构数据的预处理和特征提取,进一步基于自注意力机制实现特征层的多模态信息融合。不同于传统的机器人搭载情感识别系统,本发明充分结合了离散情绪分类和连续维度的情感预测,完整描绘了主体的情感特征空间,可以综合的从系统中获得识别的情感反馈,预测主体的情感变换走势,极大的提高了情感识别的精度。

[0113] 实施例2

[0114] 本实施例提供一种用于陪伴机器人的多模态情感识别方法,该方法与实施例1中多模态情感识别系统内各模块的处理过程相对应,方法具体包括以下步骤:

[0115] 分别采集面部表情图片、语音信号和脑电信号;

[0116] 根据所述面部表情图片提取面部表情的情感特征向量,根据所述语音信号提取语音的情感特征向量,根据所述脑电信号提取脑电信号的特征向量;

[0117] 获取权重矩阵,将所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量与该权重矩阵相乘,得到融合特征;

[0118] 将所述融合特征载入预先构建并训练好的分类模型进行分类识别,得到多个离散的情感标签识别结果,所述分类模型的训练过程中还用于训练所述权重矩阵;

[0119] 根据所述融合特征进行情感预测,该情感预测用于对融合特征进行数据拟合训

练,得到连续的情感强度数值,该情感强度数值划分为多个情感维度,该多个情感维度包括愉快度、紧张度、激动度和确定度。

[0120] 所述提取面部表情的情感特征向量具体包括:

[0121] 首先利用Adaboost算法提取面部表情图片中的Haar特征,构建Haar特征图,随后通过直方图均衡化对Haar特征图进行预处理,之后采用均匀模式LBP算法提取面部表情的情感特征向量。

[0122] 所述均匀模式LBP算法的提取过程包括:

[0123] 构造大小3\*3的纹理区域,阈值为纹理区域的中心像素值,将周围的8个像素值与阈值进行比较,若值大于阈值像素的值,该邻域处被置为1;若值小于阈值像素的值,该邻域处则被置为0;在3\*3的纹理区域内,将其相邻的8个像素点产生的值按照顺时针的方向组成8位二进制数,统计8位二进制数中0到1或者1到0跳变的次数,若跳变数在两次以内,则该二进制数对应的十进制数就是该3\*3邻域中心的LBP值;若跳变的次数大于2,取 $P=8$ ,则将该区域中心的LBP值置为 $P+1=9$ ;遍历所有的像素点,得到整幅图像的LBP值,将所有的LBP值按顺序连接成一个特征向量,即为面部表情的情感特征向量。

[0124] 所述提取语音的情感特征向量具体包括:

[0125] 首先对语音信号进行加窗处理,使用汉明窗进行平滑处理,将时域信号转化到频域进行后续的频谱分析。随后设计高通滤波器消除声带发音的噪声干扰,进行MFCC特征提取;将经过傅里叶变换的语谱图输入到预先构建并训练好的卷积神经网络层里,提取语谱图特征,得到语音的情感特征向量。

[0126] 所述提取脑电信号的特征向量具体包括:

[0127] 首先对脑电信号进行预处理去噪,然后分别提取分型维数特征和多尺度熵特征,构建脑电信号的特征向量。

[0128] 所述预处理去噪包括:

[0129] 通过固定的采样频率收集主体的脑电信号,随后选取db5小波进行多层小波分解,之后使用软阈值法将噪声产生的小波包系数置零,最后完成脑电信号的重构。

[0130] 所述分型维数特征的提取过程包括:

[0131] 对原始序列进行分析后,采样获取新的序列信号,利用预处理去噪后的脑电信号做窗口处理,然后用窗口将数据分为多段,根据每段数据分别提取分型维数特征。

[0132] 所述多尺度熵特征的提取过程包括:

[0133] 计算脑电信号的多尺度熵并求在被试主体高兴情绪和悲伤情绪的平均多尺度熵值,随后选取前一个或多个尺度下的多尺度熵作为脑电信号的特征向量。

[0134] 所述分类识别前还包括:分别对所述面部表情的情感特征向量、语音的情感特征向量和脑电信号的特征向量进行数据归一化后,得到融合特征,进行分类识别;所述分类模型采用SVM分类模型,该SVM分类模型的核函数为RBF核函数;所述情感标签包括高兴、悲伤、平静和厌恶。

[0135] 以上详细描述了本发明的较佳具体实施例。应当理解,本领域的普通技术人员无需创造性劳动就可以根据本发明的构思做出诸多修改和变化。因此,凡本技术领域中技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案,皆应在由权利要求书所确定的保护范围内。

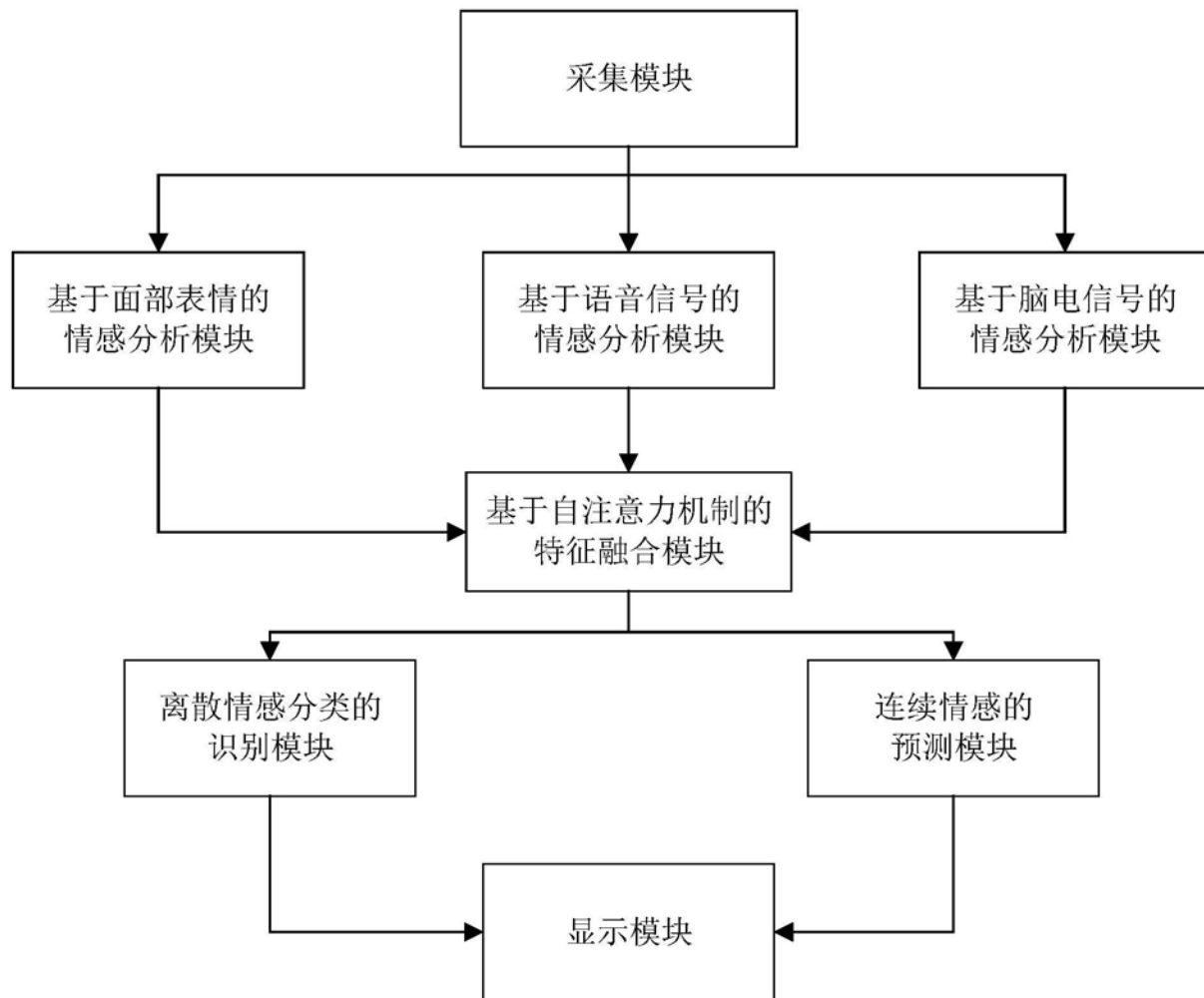


图1

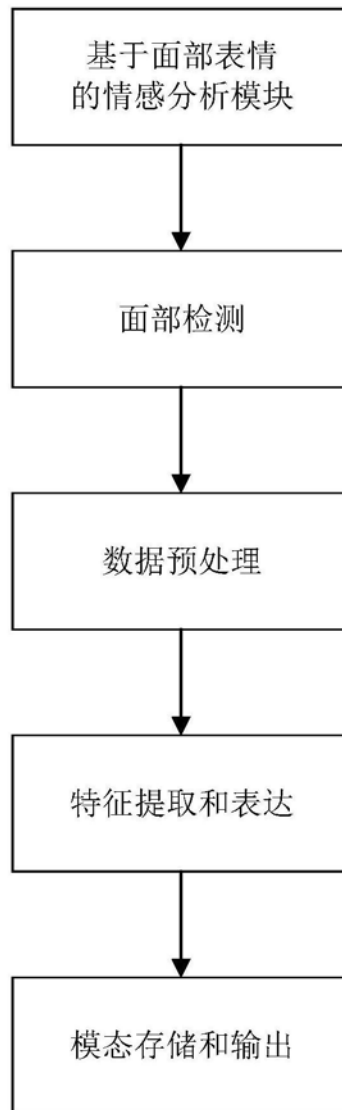


图2

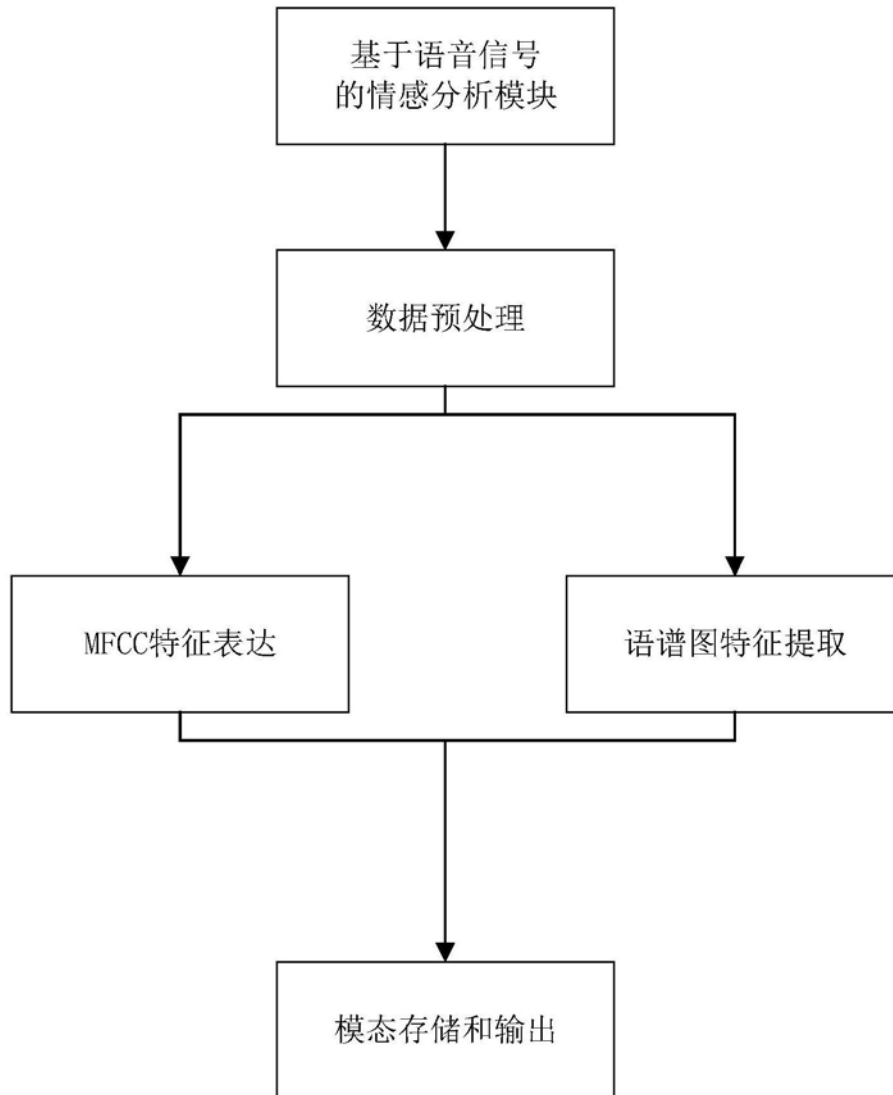


图3



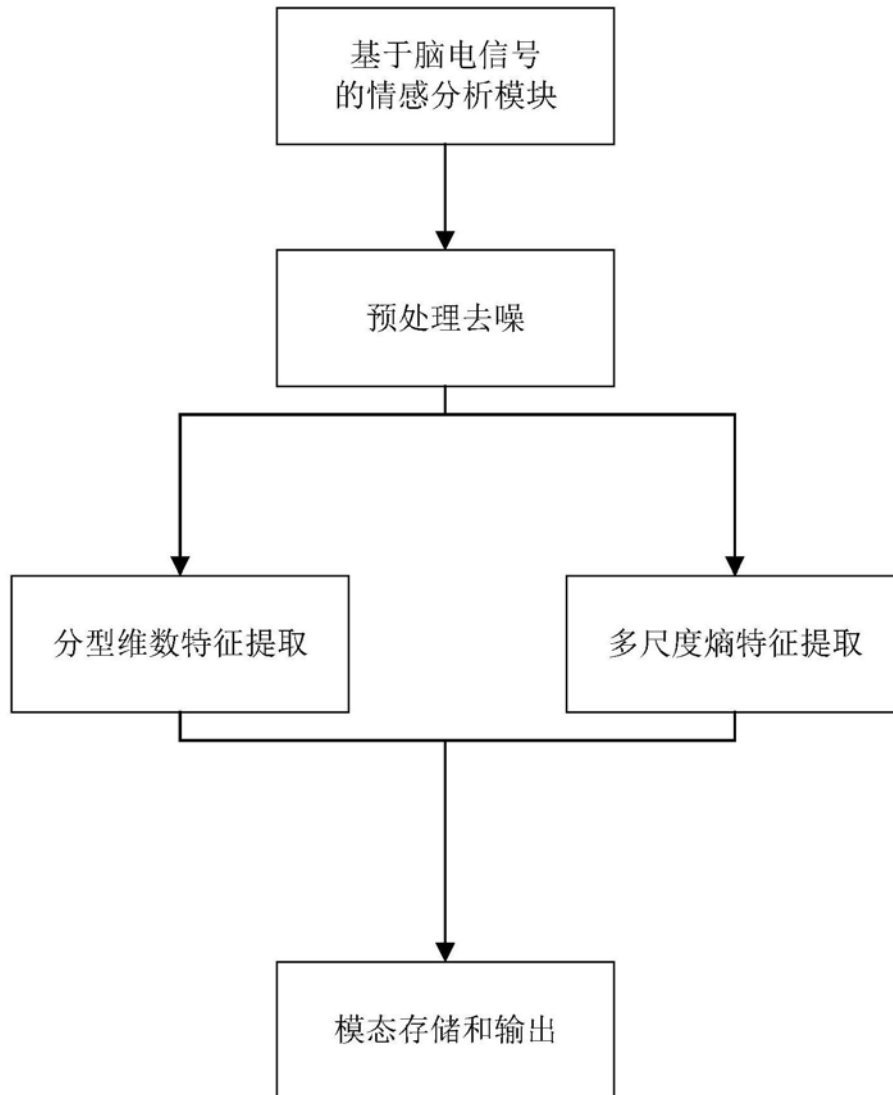


图4

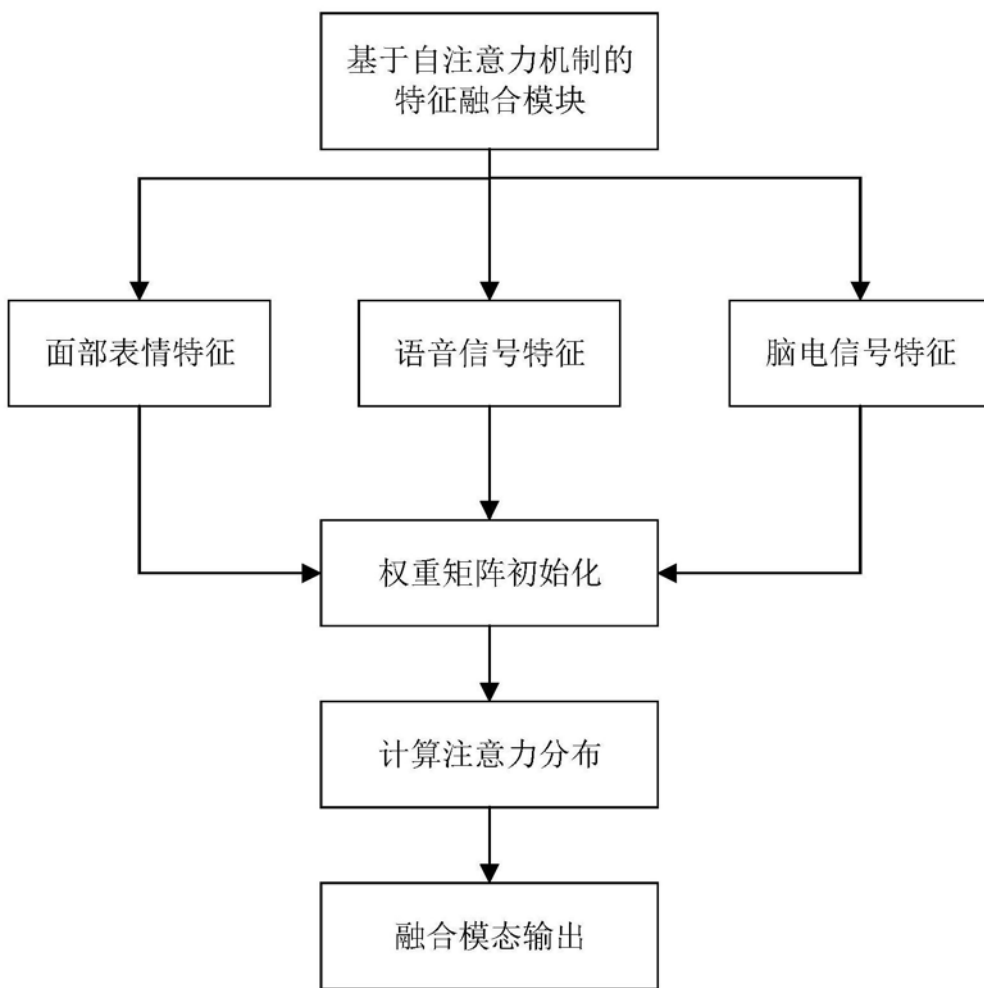


图5

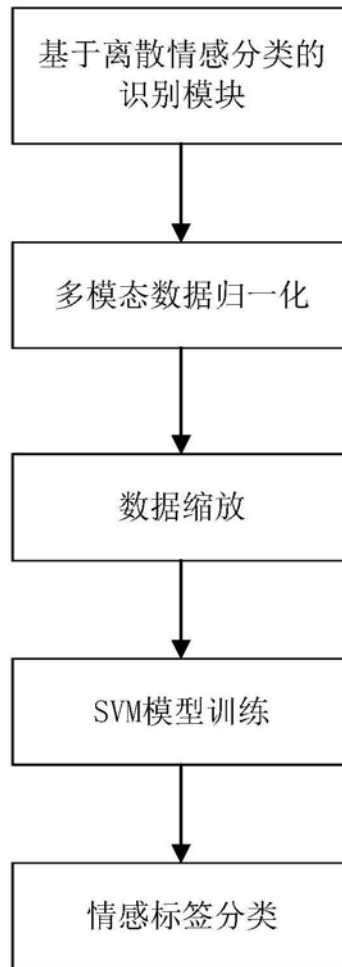


图6

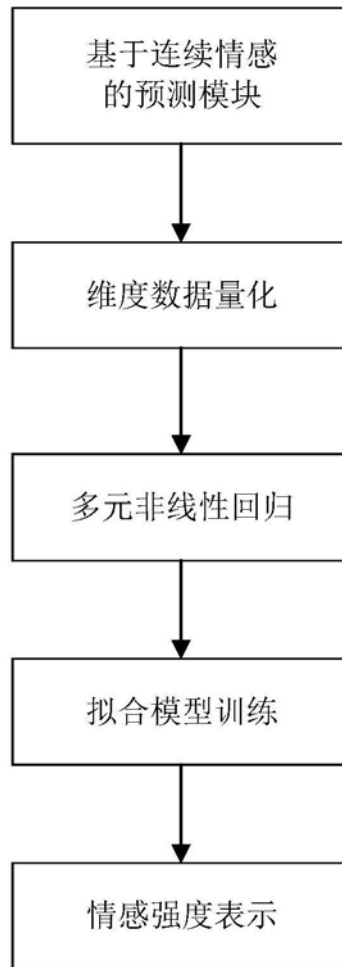


图7