



(12) 发明专利申请

(10) 申请公布号 CN 112800998 A

(43) 申请公布日 2021.05.14

(21) 申请号 202110159085.8

(22) 申请日 2021.02.05

(71) 申请人 南京邮电大学

地址 210012 江苏省南京市雨花台区软件大道186号

(72) 发明人 卢官明 朱清扬 卢峻禾

(74) 专利代理机构 南京苏高专利商标事务所 (普通合伙) 32204

代理人 孟红梅

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

G06F 17/16 (2006.01)

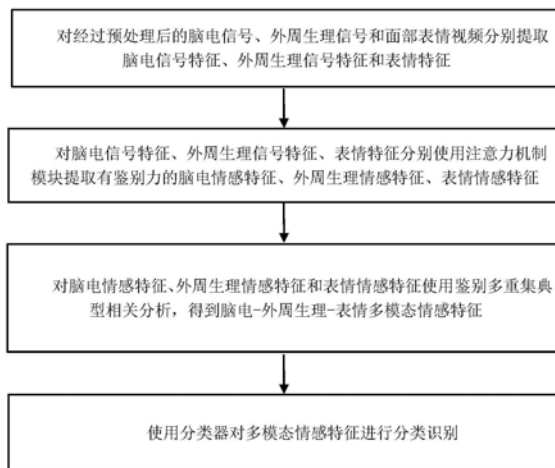
权利要求书5页 说明书12页 附图1页

(54) 发明名称

融合注意力机制和DMCCA的多模态情感识别方法及系统

(57) 摘要

本发明公开了一种融合注意力机制和鉴别多重集典型相关分析(DMCCA)的多模态情感识别方法及系统。该方法包括:对经过预处理后的脑电信号、外周生理信号和面部表情视频分别提取脑电信号特征、外周生理信号特征和表情特征;使用注意力机制分别提取有鉴别力的脑电情感特征、外周生理情感特征、表情情感特征;对脑电情感特征、外周生理情感特征和表情情感特征使用DMCCA方法,得到脑电-外周生理-表情多模态情感特征;使用分类器对多模态情感特征进行分类识别。本发明采用注意力机制选择性地重点关注各模态中更具情感鉴别力的特征,并结合DMCCA充分利用不同模态情感特征之间的相关性和互补性,可以有效提高情感识别的准确率和鲁棒性。



1. 融合注意力机制和DMCCA的多模态情感识别方法,其特征在于,所述方法包括以下步骤:

(1) 对经过预处理后的脑电信号和面部表情视频分别使用各自训练好的神经网络模型提取脑电信号特征向量和表情特征向量,对预处理后的外周生理信号,通过抽取信号波形描述符及其统计特征,提取外周生理信号特征向量;

(2) 对所述的脑电信号特征向量、外周生理信号特征向量、表情特征向量分别通过线性变换矩阵映射成若干组特征向量,并分别使用注意力机制模块确定不同特征向量组的重要性权重,通过加权融合形成维数相同的有鉴别力的脑电情感特征向量、外周生理情感特征向量、表情情感特征向量;

(3) 对所述的脑电情感特征向量、外周生理情感特征向量和表情情感特征向量,使用鉴别多重集典型相关分析(DMCCA)方法,通过最大化同一类别样本的不同模态情感特征之间的相关性,确定各情感特征向量的投影矩阵,并将各情感特征向量投影到一个公共子空间,相加融合后得到脑电-外周生理-表情多模态情感特征向量;

(4) 使用分类器对多模态情感特征向量进行分类识别,得到情感类别。

2. 根据权利要求1所述的融合注意力机制和DMCCA的多模态情感识别方法,其特征在于,步骤(2)包括以下子步骤:

(2.1) 将步骤(1)提取到的脑电信号特征以矩阵形式表示成

$\mathbf{F}^{(1)} = [\mathbf{f}_1^{(1)}, \mathbf{f}_2^{(1)}, \dots, \mathbf{f}_N^{(1)}] \in \mathbb{R}^{M_1 \times N}$, 并通过线性变换矩阵 $\mathbf{W}^{(1)}$ 映射成 M_1 组特征向量 $(\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_{M_1}^{(1)})$, $4 \leq M_1 \leq 16$, 每组特征向量的维数为 N , $16 \leq N \leq 64$, 令

$\mathbf{W}^{(1)} = [\mathbf{W}_1^{(1)}, \mathbf{W}_2^{(1)}, \dots, \mathbf{W}_{M_1}^{(1)}] \in \mathbb{R}^{M_1 \times M_1}$, $\mathbf{E}^{(1)} = [\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_{M_1}^{(1)}] \in \mathbb{R}^{N \times M_1}$, 其线性变换表达式为:

$$\mathbf{E}^{(1)} = (\mathbf{F}^{(1)})^T \mathbf{W}^{(1)}$$

其中,上标(1)代表脑电模态,T表示转置符号;

使用第一个注意力机制模块来确定不同特征向量组的重要性权重,通过加权融合形成有鉴别力的脑电情感特征向量,其中第 r 组脑电信号特征向量的权重 $a_r^{(1)}$ 以及脑电情感特征向量 $\mathbf{x}^{(1)}$ 表示为:

$$a_r^{(1)} = \frac{\exp((\mathbf{f}_r^{(1)})^T \mathbf{W}_r^{(1)})}{\sum_{r=1}^{M_1} \exp((\mathbf{f}_r^{(1)})^T \mathbf{W}_r^{(1)})}$$

$$\mathbf{x}^{(1)} = \sum_{r=1}^{M_1} a_r^{(1)} \mathbf{e}_r^{(1)}$$

其中, $r=1, 2, \dots, M_1$, $\mathbf{e}_r^{(1)} \in \mathbb{R}^{N \times 1}$ 表示第 r 组脑电信号特征向量, $\mathbf{W}_r^{(1)} \in \mathbb{R}^{M_1 \times 1}$ 为可训练的线性变换参数向量, $\exp(\cdot)$ 表示以自然常数 e 为底的指数函数;

(2.2) 将步骤(1)提取到的外周生理信号特征以矩阵形式表示成 $\mathbf{F}^{(2)} = [\mathbf{f}_1^{(2)}, \mathbf{f}_2^{(2)}, \dots, \mathbf{f}_N^{(2)}] \in \mathbb{R}^{M_2 \times N}$, 并通过线性变换矩阵 $\mathbf{W}^{(2)}$ 映射成 M_2 组特征向量

$$\left(\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \dots, \mathbf{e}_{M_2}^{(2)} \right), 4 \leq M_2 \leq 16, \text{ 令 } \mathbf{W}^{(2)} = \left[\mathbf{W}_1^{(2)}, \mathbf{W}_2^{(2)}, \dots, \mathbf{W}_{M_2}^{(2)} \right] \in \mathbb{R}^{M_2 \times M_2},$$

$$\mathbf{E}^{(2)} = \left[\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \dots, \mathbf{e}_{M_2}^{(2)} \right] \in \mathbb{R}^{N \times M_2}, \text{ 其线性变换表达式为:}$$

$$\mathbf{E}^{(2)} = (\mathbf{F}^{(2)})^T \mathbf{W}^{(2)}$$

其中,上标(2)代表外周生理模态;

使用第二个注意力机制模块来确定不同特征向量组的重要性权重,通过加权融合形成有鉴别力的外周生理情感特征向量,其中第s组外周生理信号特征向量的权重 $a_s^{(2)}$ 以及外周生理情感特征向量 $\mathbf{x}^{(2)}$ 表示为:

$$a_s^{(2)} = \frac{\exp\left(\left(\mathbf{f}_s^{(2)}\right)^T \mathbf{W}_s^{(2)}\right)}{\sum_{s=1}^{M_2} \exp\left(\left(\mathbf{f}_s^{(2)}\right)^T \mathbf{W}_s^{(2)}\right)}$$

$$\mathbf{x}^{(2)} = \sum_{s=1}^{M_2} a_s^{(2)} \mathbf{e}_s^{(2)}$$

其中, $s=1, 2, \dots, M_2$, $\mathbf{e}_s^{(2)} \in \mathbb{R}^{N \times 1}$ 表示第s组外周生理信号特征向量, $\mathbf{W}_s^{(2)} \in \mathbb{R}^{M_2 \times 1}$ 为可训练的线性变换参数向量;

(2.3) 将步骤(1)提取到的表情特征以矩阵形式表示成

$$\mathbf{F}^{(3)} = \left[\mathbf{f}_1^{(3)}, \mathbf{f}_2^{(3)}, \dots, \mathbf{f}_N^{(3)} \right] \in \mathbb{R}^{M_3 \times N}, \text{ 并通过线性变换矩阵 } \mathbf{W}^{(3)} \text{ 映射成 } M_3 \text{ 组特征向量}$$

$$\left(\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \dots, \mathbf{e}_{M_3}^{(3)} \right), 4 \leq M_3 \leq 16, \text{ 令 } \mathbf{W}^{(3)} = \left[\mathbf{W}_1^{(3)}, \mathbf{W}_2^{(3)}, \dots, \mathbf{W}_{M_3}^{(3)} \right] \in \mathbb{R}^{M_3 \times M_3},$$

$$\mathbf{E}^{(3)} = \left[\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \dots, \mathbf{e}_{M_3}^{(3)} \right] \in \mathbb{R}^{N \times M_3}, \text{ 其线性变换表达式为:}$$

$$\mathbf{E}^{(3)} = (\mathbf{F}^{(3)})^T \mathbf{W}^{(3)}$$

其中,上标(3)代表表情模态;

使用第三个注意力机制模块来确定不同特征向量组的重要性权重,通过加权融合形成有鉴别力的表情情感特征向量,其中第t组表情特征向量的权重 $a_t^{(3)}$ 以及表情情感特征向量 $\mathbf{x}^{(3)}$ 表示为:

$$a_t^{(3)} = \frac{\exp\left(\left(\mathbf{f}_t^{(3)}\right)^T \mathbf{W}_t^{(3)}\right)}{\sum_{t=1}^{M_3} \exp\left(\left(\mathbf{f}_t^{(3)}\right)^T \mathbf{W}_t^{(3)}\right)}$$

$$\mathbf{x}^{(3)} = \sum_{t=1}^{M_3} a_t^{(3)} \mathbf{e}_t^{(3)}$$

其中, $t=1, 2, \dots, M_3$, $\mathbf{e}_t^{(3)} \in \mathbb{R}^{N \times 1}$ 表示第t组表情特征向量, $\mathbf{W}_t^{(3)} \in \mathbb{R}^{M_3 \times 1}$ 为可训练的线性变换参数向量。

3. 根据权利要求2所述的融合注意力机制和DMCCA的多模态情感识别方法,其特征在

于,步骤(3)包括以下子步骤:

(3.1) 获取通过训练得到的分别对应于脑电情感特征、外周生理情感特征和表情情感特征的DMCCA投影矩阵 $\Omega = [\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_d^{(1)}] \in \mathbb{R}^{N \times d}$ 、

$\Phi = [\mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}, \dots, \mathbf{w}_d^{(2)}] \in \mathbb{R}^{N \times d}$ 和 $\Psi = [\mathbf{w}_1^{(3)}, \mathbf{w}_2^{(3)}, \dots, \mathbf{w}_d^{(3)}] \in \mathbb{R}^{N \times d}$, $32 \leq d \leq 128$;

(3.2) 分别使用投影矩阵 Ω 、 Φ 和 Ψ 将步骤(2)提取到的脑电情感特征向量 $\mathbf{x}^{(1)}$ 、外周生理情感特征向量 $\mathbf{x}^{(2)}$ 和表情情感特征向量 $\mathbf{x}^{(3)}$ 投影到一个d维的公共子空间,其中脑电情感特征向量 $\mathbf{x}^{(1)}$ 到d维公共子空间的投影为 $\Omega^T \mathbf{x}^{(1)}$,外周生理情感特征向量 $\mathbf{x}^{(2)}$ 到d维公共子空间的投影为 $\Phi^T \mathbf{x}^{(2)}$,表情情感特征向量 $\mathbf{x}^{(3)}$ 到d维公共子空间的投影为 $\Psi^T \mathbf{x}^{(3)}$;

(3.3) 将 $\Omega^T \mathbf{x}^{(1)}$ 、 $\Phi^T \mathbf{x}^{(2)}$ 和 $\Psi^T \mathbf{x}^{(3)}$ 进行融合,得到脑电-外周生理-表情多模态情感特征向量为 $\Omega^T \mathbf{x}^{(1)} + \Phi^T \mathbf{x}^{(2)} + \Psi^T \mathbf{x}^{(3)}$ 。

4. 根据权利要求3所述的融合注意力机制和DMCCA的多模态情感识别方法,其特征在于,步骤(3.1)中的投影矩阵 Ω 、 Φ 和 Ψ 通过以下步骤的训练得到:

(3.1.1) 从训练样本集中分别抽取各情感类别的训练样本生成3组情感特征向量 $\{\mathbf{X}^{(i)} \in \mathbb{R}^{N \times M}\}_{i=1}^3$, 其中 $\mathbf{X}^{(i)} = [\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_M^{(i)}]$, M为训练样本数, $i=1, 2, 3$, $m=1, 2, \dots, M$; 令 $i=1$ 代表脑电模态, $i=2$ 代表外周生理模态, $i=3$ 代表表情模态, $\mathbf{x}_m^{(1)}$ 代表脑电情感特征向量, $\mathbf{x}_m^{(2)}$ 代表外周生理情感特征向量, $\mathbf{x}_m^{(3)}$ 代表表情情感特征向量;

(3.1.2) 计算 $\mathbf{X}^{(i)}$ 中各列向量的均值,对 $\mathbf{X}^{(i)}$ 进行中心化操作;

(3.1.3) 基于鉴别多重集典型相关分析(DMCCA)的思想求得一组投影矩阵 Ω 、 Φ 和 Ψ , 使得同类样本在公共投影子空间的线性相关性最大,同时最大化模态内数据的类间散布与最小化模态内数据的类内散布,令 $\mathbf{X}^{(i)}$ 的投影向量为 $\mathbf{w}^{(i)} \in \mathbb{R}^{N \times 1}$, $i=1, 2, 3$, DMCCA的目标函数为:

$$J(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}) = \frac{\sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)}}{\sqrt{\sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)}}}$$

其中, \mathbf{S}_i^w 表示 $\mathbf{X}^{(i)}$ 的类内散布矩阵, $\mathbf{S}_{ij} = \begin{cases} \mathbf{S}_i^b, & i=j \\ \text{cov}(\mathbf{S}_i^b, \mathbf{S}_j^b), & i \neq j \end{cases}$, \mathbf{S}_i^b 表示 $\mathbf{X}^{(i)}$ 的类间散布

矩阵, $\text{cov}(\cdot, \cdot)$ 表示协方差, $i, j \in \{1, 2, 3\}$; 构建如下优化模型并求解得到投影矩阵 Ω 、 Φ 和 Ψ :

$$\begin{cases} \max_{\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}} \sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)} \\ \text{s.t.} \quad \sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)} = 1 \end{cases} .$$

5. 根据权利要求4所述的融合注意力机制和DMCCA的多模态情感识别方法,其特征在于,使用拉格朗日乘子法求解所构建的DMCCA目标函数的优化模型,具体为:将优化模型表示为如下拉格朗日函数:

$$L(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}) = \sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)} - \lambda \left(\sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)} - 1 \right)$$

其中, λ 是拉格朗日乘子,再分别求 $L(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)})$ 对 $\mathbf{w}^{(1)}$ 、 $\mathbf{w}^{(2)}$ 和 $\mathbf{w}^{(3)}$ 的偏导数并令其为零,即令

$$\frac{\partial L}{\partial \mathbf{w}^{(i)}} = \sum_{j=1}^3 \mathbf{S}_{ij} \mathbf{w}^{(j)} - \lambda \mathbf{S}_i^w \mathbf{w}^{(i)} = 0, \quad i = 1, 2, 3$$

得到

$$\sum_{j=1}^3 \mathbf{S}_{ij} \mathbf{w}^{(j)} = \lambda \mathbf{S}_i^w \mathbf{w}^{(i)}, \quad i = 1, 2, 3$$

进一步对上式作简化处理,则可获得如下的广义特征值问题:

$$\begin{bmatrix} \mathbf{S}_1^b & \mathbf{S}_{12} & \mathbf{S}_{13} \\ \mathbf{S}_{21} & \mathbf{S}_2^b & \mathbf{S}_{23} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_3^b \end{bmatrix} \begin{bmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \\ \mathbf{w}^{(3)} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{S}_1^w & 0 & 0 \\ 0 & \mathbf{S}_2^w & 0 \\ 0 & 0 & \mathbf{S}_3^w \end{bmatrix} \begin{bmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \\ \mathbf{w}^{(3)} \end{bmatrix}$$

通过求解上式中的广义特征值问题,选取前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 对应的特征向量,即可得到投影矩阵 $\mathbf{\Omega} = [\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_d^{(1)}] \in \mathbb{R}^{N \times d}$ 、

$\mathbf{\Phi} = [\mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}, \dots, \mathbf{w}_d^{(2)}] \in \mathbb{R}^{N \times d}$ 和 $\mathbf{\Psi} = [\mathbf{w}_1^{(3)}, \mathbf{w}_2^{(3)}, \dots, \mathbf{w}_d^{(3)}] \in \mathbb{R}^{N \times d}$ 。

6. 融合注意力机制和DMCCA的多模态情感识别系统,其特征在于,包括:

特征初步提取模块,用于对经过预处理后的脑电信号和面部表情视频分别使用各自训练好的神经网络模型提取脑电信号特征向量和表情特征向量,对预处理后的外周生理信号,通过抽取信号波形描述符及其统计特征,提取外周生理信号特征向量;

特征鉴别增强模块,用于对所述的脑电信号特征向量、外周生理信号特征向量、表情特征向量分别通过线性变换矩阵映射成若干组特征向量,并分别使用注意力机制模块确定不同特征向量组的重要性权重,通过加权融合形成维数相同的有鉴别力的脑电情感特征向量、外周生理情感特征向量、表情情感特征向量;

投影矩阵确定模块,用于使用鉴别多重集典型相关分析(DMCCA)方法,通过最大化同一类别样本的不同模态情感特征之间的相关性,确定各情感特征向量的投影矩阵;

特征融合模块,用于对所述的脑电情感特征向量、外周生理情感特征向量和表情情感特征向量,通过各自对应的投影矩阵投影到一个公共子空间,相加融合后得到脑电-外周生理-表情多模态情感特征向量;

以及,分类识别模块,用于使用分类器对多模态情感特征向量进行分类识别,得到情感类别。

7. 融合注意力机制和DMCCA的多模态情感识别系统,其特征在于,包括至少一台计算设备,所述计算设备包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程

序,所述计算机程序被加载至处理器时实现根据权利要求1-5任一项所述的融合注意力机制和DMCCA的多模态情感识别方法。

融合注意力机制和DMCCA的多模态情感识别方法及系统

技术领域

[0001] 本发明涉及情感识别和人工智能技术领域,尤其涉及一种融合注意力机制和鉴别多重集典型相关分析(DMCCA)的多模态情感识别方法及系统。

背景技术

[0002] 人类情感是伴随着人的意识过程产生的心理和生理状态,在人际交流中起着重要作用。随着人工智能等技术的不断进步,获得更加智能化、人性化的人机交互(Human-Computer Interactions,HCIs)体验越来越受到人们的关注。人们对机器智能化的要求越来越高,期望机器能够具有感知、理解甚至表达情感的能力,实现人性化的人机交互,更好地为人类服务。情感识别作为情感计算的一个分支,是实现人-机情感交互的基础和核心技术,已经成为计算机科学、认知科学和人工智能等领域的研究热点,受到了学术界和工业界的广泛关注。例如,在临床医疗护理中,如果能够知道患者,尤其是有表达障碍的患者的情感状态,就可以采取不同的护理措施,以提高护理质量。此外,在对精神障碍患者的心理行为监控、情感机器人的人-机友好交互等方面也得到了越来越多的关注。

[0003] 以往对情感识别的研究大多集中在通过单一模态的信息来识别人类情感状态,比如基于语音的情感识别、基于面部表情的情感识别等。由于单一的语音或表情信息所表达的情感信息是不完整的,且容易受到外界各种因素的影响,例如面部表情识别容易受遮挡和光照变化的影响,而基于语音的情感识别容易受环境噪音干扰和不同受试者的声音差异的影响,此外,有时候人们为了掩盖自己的真实情感而强颜欢笑、装腔作势或沉默不语,此时,面部表情或身体姿态具有一定的欺骗性,而当人们沉默不语时基于语音的情感识别方法就会失效,所以,单模态情感识别存在一定的局限性。因此,越来越多的研究人员将目光转向基于多模态信息融合的情感识别研究,期望能够利用各个模态信息之间的互补性,来构建鲁棒的情感识别模型,以达到更高的情感识别准确率。

[0004] 目前,在进行多模态情感识别研究中,比较常用的信息融合策略有决策层融合和特征层融合。决策层融合通常基于各模态单独识别的结果,再依据相关规则,如均值(Mean)规则、求和(Sum)规则、最大值(Max)规则、少数服从多数的投票表决机制等,做出决策判断,得到最终的识别结果。决策层融合技术根据不同模态信息对于情感识别的贡献不同,比较全面地考虑了不同模态信息之间的差异性,但却忽略了不同模态信息之间的相关性。基于决策层融合的多模态情感识别性能不仅与单个模态的情感识别率有关,还依赖于决策层融合算法的性能。特征层融合是指将多个模态的情感特征联合起来,形成一个融合特征向量。特征层融合方法利用了不同模态情感特征的互补性,但如何确定不同模态情感特征的权值,以体现不同特征在情感分类识别中的差异性,是进行多模态特征融合的关键,目前仍然是面临挑战的开放课题。

发明内容

[0005] 发明目的:针对单模态情感识别准确率低、鲁棒性差以及现有多模态情感特征融

合方法存在的缺点,本发明的目的是提供一种融合注意力机制和鉴别多重集典型相关分析(DMCCA)的多模态情感识别方法及系统,通过引入注意力机制选择性地重点关注各模态中有鉴别力的情感特征,并结合DMCCA充分利用不同模态情感特征之间的相关性和互补性,可以有效提高多模态情感识别的准确率和鲁棒性。

[0006] 技术方案:本发明为实现上述发明目的采用以下技术方案:

[0007] 一种融合注意力机制和DMCCA的多模态情感识别方法,包括以下步骤:

[0008] (1)对经过预处理后的脑电信号和面部表情视频分别使用各自训练好的神经网络模型提取脑电信号特征向量和表情特征向量,对预处理后的外周生理信号,通过抽取信号波形描述符及其统计特征,提取外周生理信号特征向量;

[0009] (2)对所述的脑电信号特征向量、外周生理信号特征向量、表情特征向量分别通过线性变换矩阵映射成若干组特征向量,并分别使用注意力机制模块确定不同特征向量组的重要性权重,通过加权融合形成维数相同的有鉴别力的脑电情感特征向量、外周生理情感特征向量、表情情感特征向量;

[0010] (3)对所述的脑电情感特征向量、外周生理情感特征向量和表情情感特征向量,使用鉴别多重集典型相关分析(DMCCA)方法,通过最大化同一类别样本的不同模态情感特征之间的相关性,确定各情感特征向量的投影矩阵,并将各情感特征向量投影到一个公共子空间,相加融合后得到脑电-外周生理-表情多模态情感特征向量;

[0011] (4)使用分类器对多模态情感特征向量进行分类识别,得到情感类别。

[0012] 进一步地,步骤(2)中使用注意力机制模块提取有鉴别力的脑电情感特征、外周生理情感特征、表情情感特征的具体步骤包括:

[0013] (2.1)将步骤(1)提取到的脑电信号特征以矩阵形式表示成

$\mathbf{F}^{(1)} = [\mathbf{f}_1^{(1)}, \mathbf{f}_2^{(1)}, \dots, \mathbf{f}_N^{(1)}] \in \mathbb{R}^{M_1 \times N}$,并通过线性变换矩阵 $\mathbf{W}^{(1)}$ 映射成 M_1 组特征向量 $(\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_{M_1}^{(1)})$, $4 \leq M_1 \leq 16$,每组特征向量的维数为 N , $16 \leq N \leq 64$,令

$\mathbf{W}^{(1)} = [\mathbf{W}_1^{(1)}, \mathbf{W}_2^{(1)}, \dots, \mathbf{W}_{M_1}^{(1)}] \in \mathbb{R}^{M_1 \times M_1}$, $\mathbf{E}^{(1)} = [\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_{M_1}^{(1)}] \in \mathbb{R}^{N \times M_1}$,其线性变换表达式为:

[0014] $\mathbf{E}^{(1)} = (\mathbf{F}^{(1)})^T \mathbf{W}^{(1)}$

[0015] 其中,上标(1)代表脑电模态,T表示转置符号;

[0016] 使用第一个注意力机制模块来确定不同特征向量组的重要性权重,通过加权融合形成有鉴别力的脑电情感特征向量,其中第 r 组脑电信号特征向量的权重 $a_r^{(1)}$ 以及脑电情感特征向量 $\mathbf{x}^{(1)}$ 表示为:

[0017]
$$a_r^{(1)} = \frac{\exp((\mathbf{f}_r^{(1)})^T \mathbf{W}_r^{(1)})}{\sum_{r=1}^{M_1} \exp((\mathbf{f}_r^{(1)})^T \mathbf{W}_r^{(1)})}$$

[0018]
$$\mathbf{x}^{(1)} = \sum_{r=1}^{M_1} a_r^{(1)} \mathbf{e}_r^{(1)}$$

[0019] 其中, $r=1, 2, \dots, M_1$, $\mathbf{e}_r^{(1)} \in \mathbb{R}^{N \times 1}$ 表示第 r 组脑电信号特征向量, $\mathbf{W}_r^{(1)} \in \mathbb{R}^{M_1 \times 1}$ 为可训练的线性变换参数向量, $\exp(\cdot)$ 表示以自然常数 e 为底的指数函数;

[0020] (2.2) 将步骤(1)提取到的外周生理信号特征以矩阵形式表示成 $\mathbf{F}^{(2)} = [\mathbf{f}_1^{(2)}, \mathbf{f}_2^{(2)}, \dots, \mathbf{f}_N^{(2)}] \in \mathbb{R}^{M_2 \times N}$, 并通过线性变换矩阵 $\mathbf{W}^{(2)}$ 映射成 M_2 组特征向量 $(\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \dots, \mathbf{e}_{M_2}^{(2)})$, $4 \leq M_2 \leq 16$, 令 $\mathbf{W}^{(2)} = [\mathbf{W}_1^{(2)}, \mathbf{W}_2^{(2)}, \dots, \mathbf{W}_{M_2}^{(2)}] \in \mathbb{R}^{M_2 \times M_2}$, $\mathbf{E}^{(2)} = [\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \dots, \mathbf{e}_{M_2}^{(2)}] \in \mathbb{R}^{N \times M_2}$, 其线性变换表达式为:

$$[0021] \quad \mathbf{E}^{(2)} = (\mathbf{F}^{(2)})^T \mathbf{W}^{(2)}$$

[0022] 其中, 上标(2)代表外周生理模态;

[0023] 使用第二个注意力机制模块来确定不同特征向量组的重要性权重, 通过加权融合形成有鉴别力的外周生理情感特征向量, 其中第s组外周生理信号特征向量的权重 $a_s^{(2)}$ 以及外周生理情感特征向量 $\mathbf{x}^{(2)}$ 表示为:

$$[0024] \quad a_s^{(2)} = \frac{\exp((\mathbf{f}_s^{(2)})^T \mathbf{W}_s^{(2)})}{\sum_{s=1}^{M_2} \exp((\mathbf{f}_s^{(2)})^T \mathbf{W}_s^{(2)})}$$

$$[0025] \quad \mathbf{x}^{(2)} = \sum_{s=1}^{M_2} a_s^{(2)} \mathbf{e}_s^{(2)}$$

[0026] 其中, $s=1, 2, \dots, M_2$, $\mathbf{e}_s^{(2)} \in \mathbb{R}^{N \times 1}$ 表示第s组外周生理信号特征向量, $\mathbf{W}_s^{(2)} \in \mathbb{R}^{M_2 \times 1}$ 为可训练的线性变换参数向量;

[0027] (2.3) 将步骤(1)提取到的表情特征以矩阵形式表示成 $\mathbf{F}^{(3)} = [\mathbf{f}_1^{(3)}, \mathbf{f}_2^{(3)}, \dots, \mathbf{f}_N^{(3)}] \in \mathbb{R}^{M_3 \times N}$, 并通过线性变换矩阵 $\mathbf{W}^{(3)}$ 映射成 M_3 组特征向量 $(\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \dots, \mathbf{e}_{M_3}^{(3)})$, $4 \leq M_3 \leq 16$, 令 $\mathbf{W}^{(3)} = [\mathbf{W}_1^{(3)}, \mathbf{W}_2^{(3)}, \dots, \mathbf{W}_{M_3}^{(3)}] \in \mathbb{R}^{M_3 \times M_3}$, $\mathbf{E}^{(3)} = [\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \dots, \mathbf{e}_{M_3}^{(3)}] \in \mathbb{R}^{N \times M_3}$, 其线性变换表达式为:

$$[0028] \quad \mathbf{E}^{(3)} = (\mathbf{F}^{(3)})^T \mathbf{W}^{(3)}$$

[0029] 其中, 上标(3)代表表情模态;

[0030] 使用第三个注意力机制模块来确定不同特征向量组的重要性权重, 通过加权融合形成有鉴别力的表情情感特征向量, 其中第t组表情特征向量的权重 $a_t^{(3)}$ 以及表情情感特征向量 $\mathbf{x}^{(3)}$ 表示为:

$$[0031] \quad a_t^{(3)} = \frac{\exp((\mathbf{f}_t^{(3)})^T \mathbf{W}_t^{(3)})}{\sum_{t=1}^{M_3} \exp((\mathbf{f}_t^{(3)})^T \mathbf{W}_t^{(3)})}$$

$$[0032] \quad \mathbf{x}^{(3)} = \sum_{t=1}^{M_3} a_t^{(3)} \mathbf{e}_t^{(3)}$$

[0033] 其中, $t=1, 2, \dots, M_3$, $\mathbf{e}_t^{(3)} \in \mathbb{R}^{N \times 1}$ 表示第t组表情特征向量, $\mathbf{W}_t^{(3)} \in \mathbb{R}^{M_3 \times 1}$ 为可训练的线性变换参数向量。

[0034] 进一步地, 步骤(3)具体包括以下子步骤:

[0035] (3.1) 获取通过训练得到的分别对应于脑电情感特征、外周生理情感特征和表情情感特征的DMCCA投影矩阵 $\Omega = [\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_d^{(1)}] \in \mathbb{R}^{N \times d}$ 、 $\Phi = [\mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}, \dots, \mathbf{w}_d^{(2)}] \in \mathbb{R}^{N \times d}$ 和 $\Psi = [\mathbf{w}_1^{(3)}, \mathbf{w}_2^{(3)}, \dots, \mathbf{w}_d^{(3)}] \in \mathbb{R}^{N \times d}$, $32 \leq d \leq 128$;

[0036] (3.2) 分别使用投影矩阵 Ω 、 Φ 和 Ψ 将步骤 (2) 提取到的脑电情感特征向量 $\mathbf{x}^{(1)}$ 、外周生理情感特征向量 $\mathbf{x}^{(2)}$ 和表情情感特征向量 $\mathbf{x}^{(3)}$ 投影到一个 d 维的公共子空间, 其中脑电情感特征向量 $\mathbf{x}^{(1)}$ 到 d 维公共子空间的投影为 $\Omega^T \mathbf{x}^{(1)}$, 外周生理情感特征向量 $\mathbf{x}^{(2)}$ 到 d 维公共子空间的投影为 $\Phi^T \mathbf{x}^{(2)}$, 表情情感特征向量 $\mathbf{x}^{(3)}$ 到 d 维公共子空间的投影为 $\Psi^T \mathbf{x}^{(3)}$;

[0037] (3.3) 将 $\Omega^T \mathbf{x}^{(1)}$ 、 $\Phi^T \mathbf{x}^{(2)}$ 和 $\Psi^T \mathbf{x}^{(3)}$ 进行融合, 得到脑电-外周生理-表情多模态情感特征向量为 $\Omega^T \mathbf{x}^{(1)} + \Phi^T \mathbf{x}^{(2)} + \Psi^T \mathbf{x}^{(3)}$ 。

[0038] 进一步地, 步骤 (3.1) 中的投影矩阵 Ω 、 Φ 和 Ψ 通过以下步骤的训练得到:

[0039] (3.1.1) 从训练样本集中分别抽取各情感类别的训练样本生成 3 组情感特征向量 $\{\mathbf{X}^{(i)} \in \mathbb{R}^{N \times M}\}_{i=1}^3$, 其中 $\mathbf{X}^{(i)} = [\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_M^{(i)}]$, M 为训练样本数, N 为 $\mathbf{x}_m^{(i)}$ 的维数, $i=1, 2, 3, m=1, 2, \dots, M$; 令 $i=1$ 代表脑电模态, $i=2$ 代表外周生理模态, $i=3$ 代表表情模态, $\mathbf{x}_m^{(1)}$ 代表脑电情感特征向量, $\mathbf{x}_m^{(2)}$ 代表外周生理情感特征向量, $\mathbf{x}_m^{(3)}$ 代表表情情感特征向量;

[0040] (3.1.2) 计算 $\mathbf{X}^{(i)}$ 中各列向量的均值, 对 $\mathbf{X}^{(i)}$ 进行中心化操作;

[0041] (3.1.3) 基于鉴别多重集典型相关分析 (DMCCA) 的思想求得一组投影矩阵 Ω 、 Φ 和 Ψ , 使得同类样本在公共投影子空间的线性相关性最大, 同时最大化模态内数据的类间散布与最小化模态内数据的类内散布, 令 $\mathbf{X}^{(i)}$ 的投影向量为 $\mathbf{w}^{(i)} \in \mathbb{R}^{N \times 1}$, $i=1, 2, 3$, DMCCA 的目标函数为:

$$[0042] \quad J(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}) = \frac{\sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)}}{\sqrt{\sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)}}}$$

[0043] 其中, \mathbf{S}_i^w 表示 $\mathbf{X}^{(i)}$ 的类内散布矩阵, $\mathbf{S}_{ij} = \begin{cases} \mathbf{S}_i^b, & i=j \\ \text{cov}(\mathbf{S}_i^b, \mathbf{S}_j^b), & i \neq j \end{cases}$, \mathbf{S}_i^b 表示 $\mathbf{X}^{(i)}$ 的类间散布矩阵, $\text{cov}(\cdot, \cdot)$ 表示协方差, $i, j \in \{1, 2, 3\}$;

[0044] 构建如下优化模型并求解得到投影矩阵 Ω 、 Φ 和 Ψ :

$$[0045] \quad \begin{cases} \max_{\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}} \sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)} \\ \text{s.t.} \quad \sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)} = 1 \end{cases}$$

[0046] 进一步地, 使用拉格朗日乘子法 (Lagrange multiplier) 求解 DMCCA 目标函数的优化模型, 可得到如下拉格朗日 (Lagrange) 函数:

$$[0047] \quad L(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}) = \sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)} - \lambda \left(\sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)} - 1 \right)$$

[0048] 其中, λ 是拉格朗日乘子, 再分别求 $L(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)})$ 对 $\mathbf{w}^{(1)}$ 、 $\mathbf{w}^{(2)}$ 和 $\mathbf{w}^{(3)}$ 的偏导数并令

其为零,即令

$$[0049] \quad \frac{\partial L}{\partial \mathbf{w}^{(i)}} = \sum_{j=1}^3 \mathbf{S}_{ij} \mathbf{w}^{(j)} - \lambda \mathbf{S}_i^w \mathbf{w}^{(i)} = 0, \quad i=1,2,3$$

[0050] 得到

$$[0051] \quad \sum_{j=1}^3 \mathbf{S}_{ij} \mathbf{w}^{(j)} = \lambda \mathbf{S}_i^w \mathbf{w}^{(i)}, \quad i=1,2,3$$

[0052] 进一步对上式作简化处理,则可获得如下的广义特征值问题:

$$[0053] \quad \begin{bmatrix} \mathbf{S}_1^b & \mathbf{S}_{12} & \mathbf{S}_{13} \\ \mathbf{S}_{21} & \mathbf{S}_2^b & \mathbf{S}_{23} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_3^b \end{bmatrix} \begin{bmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \\ \mathbf{w}^{(3)} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{S}_1^w & 0 & 0 \\ 0 & \mathbf{S}_2^w & 0 \\ 0 & 0 & \mathbf{S}_3^w \end{bmatrix} \begin{bmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \\ \mathbf{w}^{(3)} \end{bmatrix}$$

[0054] 通过求解上式中的广义特征值问题,选取前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 对应的特征向量,即可得到投影矩阵 $\mathbf{\Omega} = [\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_d^{(1)}] \in \mathbb{R}^{N \times d}$ 、 $\mathbf{\Phi} = [\mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}, \dots, \mathbf{w}_d^{(2)}] \in \mathbb{R}^{N \times d}$ 和 $\mathbf{\Psi} = [\mathbf{w}_1^{(3)}, \mathbf{w}_2^{(3)}, \dots, \mathbf{w}_d^{(3)}] \in \mathbb{R}^{N \times d}$ 。

[0055] 基于相同的发明构思,本发明提供的融合注意力机制和DMCCA的多模态情感识别系统,包括:

[0056] 特征初步提取模块,用于对经过预处理后的脑电信号和面部表情视频分别使用各自训练好的神经网络模型提取脑电信号特征向量和表情特征向量,对预处理后的外周生理信号,通过抽取信号波形描述符及其统计特征,提取外周生理信号特征向量;

[0057] 特征鉴别增强模块,用于对所述的脑电信号特征向量、外周生理信号特征向量、表情特征向量分别通过线性变换矩阵映射成若干组特征向量,并分别使用注意力机制模块确定不同特征向量组的重要性权重,通过加权融合形成维数相同的有鉴别力的脑电情感特征向量、外周生理情感特征向量、表情情感特征向量;

[0058] 投影矩阵确定模块,用于使用鉴别多重集典型相关分析(DMCCA)方法,通过最大化同一类别样本的不同模态情感特征之间的相关性,确定各情感特征向量的投影矩阵;

[0059] 特征融合模块,用于对所述的脑电情感特征向量、外周生理情感特征向量和表情情感特征向量,通过各自对应的投影矩阵投影到一个公共子空间,相加融合后得到脑电-外周生理-表情多模态情感特征向量;

[0060] 以及,分类识别模块,用于使用分类器对多模态情感特征向量进行分类识别,得到情感类别。

[0061] 基于相同的发明构思,本发明提供的融合注意力机制和DMCCA的多模态情感识别系统,包括至少一台计算设备,所述计算设备包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述计算机程序被加载至处理器时实现所述的融合注意力机制和DMCCA的多模态情感识别方法。

[0062] 有益效果:与现有技术相比,本发明具有以下技术效果:

[0063] (1) 本发明采用注意力机制选择性地重点关注各模态中对情感识别起关键作用的显著性特征,自适应地学习具有情感鉴别能力的特征,可以有效提高多模态情感识别的准确率和鲁棒性。

[0064] (2) 本发明采用鉴别多重集典型相关分析方法,引入了样本的类别信息,通过最大化同一类别样本不同模态情感特征之间的相关性,以及最大化同一模态情感特征的类间散布与最小化同一模态情感特征的类内散布,能够挖掘不同模态之间非线性的相关关系,充分利用脑电情感特征、外周生理情感特征和表情情感特征之间的相关性和互补性,同时又消除了一些无效的冗余特征,可以有效提高特征表示的鉴别力与鲁棒性。

[0065] (3) 与单模态情感识别方法相比,本发明综合利用了情感表达过程中的多种模态信息,能够结合不同模态的特点并充分利用其互补性来挖掘多模态情感特征,可以有效提高情感识别的准确率和鲁棒性。

附图说明

[0066] 图1是本发明实施例的方法流程图;

[0067] 图2是本发明实施例的结构图。

具体实施方式

[0068] 为了更加详细了解本发明,下面结合说明书附图和具体实施例对本发明做进一步的说明。

[0069] 如图1和图2所示,本发明实施例提供一种融合注意力机制和DMCCA的多模态情感识别方法,主要包括如下步骤:

[0070] (1) 对经过预处理后的脑电信号和面部表情视频分别使用各自训练好的神经网络模型提取脑电信号特征向量和表情特征向量,对预处理后的外周生理信号,通过抽取信号波形描述符及其统计特征,提取外周生理信号特征向量。

[0071] 本实施例中采用DEAP (Database for Emotion Analysis using Physiological Signals) 情感数据库,在实际中也可以采用其他的包含脑电、外周生理信号、面部表情视频的情感数据库。本实施例中使用的DEAP数据库是由英国伦敦玛丽皇后大学的Koelstra等人采集并到公开的多模态情感数据库。该数据库包含32名受试者在观看40个时长均为1分钟的不同种类音乐视频片段诱发刺激下产生的生理信号、外周生理信号以及前22名受试者在观看音乐视频片段时的面部表情视频。每个受试者需要进行40次实验,并且在每次实验结束后都要进行及时的自我评估 (Self-assessment Manikins, SAM),需要在SAM问卷调查表上进行40次自我评估。SAM问卷调查表包含受试者对视频的唤醒度 (Arousal)、效价度 (Valence)、支配度 (Dominance) 和喜好度 (Liking) 的心理量表。唤醒度表示人的状态兴奋程度,变化范围由平静状态逐渐过渡到兴奋状态,用数字1到9的分值来衡量;效价度也称愉悦度,表示人的心情愉悦程度,变化范围由消极 (Negative) 状态逐渐过渡到积极 (Positive) 状态,也用数字1到9的分值来衡量;支配度的变化范围从顺从 (或“无控制”) 到支配 (或“有控制”);喜好度表示受试者对视频的个人喜好。每位受试者需要在每次实验后选择代表情感状态的分值,用作后面的情感分类的类别和识别分析。

[0072] 在DEAP数据库中,生理信号采用512Hz采样、128Hz复采样 (官方提供了经过预处理的复采样数据),每个受试者的生理信号矩阵为 $40 \times 40 \times 8064$ (40个不同种类音乐视频片段,40个生理信号通道,8064个采样点)。在采集的40个生理信号通道中,前32个通道采集的是脑电信号,后8个通道采集的是外周生理信号。8064个采样数据则是在128Hz采样率下时

长为63s的数据,每一段信号记录前,都有3s静默时间。

[0073] 在本发明实施例中,我们采用同时具有脑电信号、外周生理信号和面部表情的880个样本作为训练样本,在唤醒度、效价度、支配度和喜好度的4个维度上分别做二分类识别。

[0074] 用于提取脑电信号特征的神经网络模型可以采用长短时记忆(Long Short-Term Memory,LSTM)网络或卷积神经网络(Convolutional Neural Network,CNN),用于提取表情特征的神经网络模型可以采用3D卷积神经网络、CNN-LSTM等。在本实施例中,使用训练好的卷积神经网络(CNN)模型对预处理后的脑电信号进行特征提取,得到256维的脑电信号特征向量;对预处理后的心电、呼吸、眼电、肌电等外周生理信号,通过抽取信号波形的低层描述符(Low Level Descriptor,LLD)及其统计特征(包括平均值、标准偏差、功率谱、中值、最大值和最小值),提取了128维的外周生理信号特征向量;对预处理后的面部表情视频使用训练好的CNN-LSTM模型提取256维的表情特征向量。

[0075] (2)对脑电信号特征向量、外周生理信号特征向量、表情特征向量分别使用注意力机制模块提取有鉴别力的脑电情感特征向量、外周生理情感特征向量、表情情感特征向量。

[0076] (3)对脑电情感特征向量、外周生理情感特征向量和表情情感特征向量,使用鉴别多重集典型相关分析(DMCCA)方法,得到脑电-外周生理-表情多模态情感特征向量。

[0077] (4)使用分类器对多模态情感特征向量进行分类识别,得到情感类别。

[0078] 进一步地,步骤(2)中使用注意力机制模块提取有鉴别力的脑电情感特征、外周生理情感特征、表情情感特征的具体步骤包括:

[0079] (2.1)将步骤(1)提取到的脑电信号特征以矩阵形式表示成

$$\mathbf{F}^{(1)} = [\mathbf{f}_1^{(1)}, \mathbf{f}_2^{(1)}, \dots, \mathbf{f}_N^{(1)}] \in \mathbb{R}^{M_1 \times N}, \text{ 并通过线性变换矩阵 } \mathbf{W}^{(1)} \text{ 映射成 } M_1 \text{ 组特征向量 } (\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_{M_1}^{(1)}),$$

$4 \leq M_1 \leq 16$, 每组特征向量的维数为 N , $16 \leq N \leq 64$, 令

$$\mathbf{W}^{(1)} = [\mathbf{W}_1^{(1)}, \mathbf{W}_2^{(1)}, \dots, \mathbf{W}_{M_1}^{(1)}] \in \mathbb{R}^{M_1 \times M_1}, \quad \mathbf{E}^{(1)} = [\mathbf{e}_1^{(1)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_{M_1}^{(1)}] \in \mathbb{R}^{N \times M_1}, \text{ 其线性变换表达式为:}$$

$$[0080] \quad \mathbf{E}^{(1)} = (\mathbf{F}^{(1)})^T \mathbf{W}^{(1)}$$

[0081] 其中,上标(1)代表脑电模态,T表示转置符号。

[0082] 使用第一个注意力机制模块来确定不同特征向量组的重要性权重,通过加权融合形成有鉴别力的脑电情感特征向量,其中第 r 组脑电信号特征向量的权重 $a_r^{(1)}$ 以及脑电情感特征向量 $\mathbf{x}^{(1)}$ 表示为:

$$[0083] \quad a_r^{(1)} = \frac{\exp((\mathbf{f}_r^{(1)})^T \mathbf{W}_r^{(1)})}{\sum_{r=1}^{M_1} \exp((\mathbf{f}_r^{(1)})^T \mathbf{W}_r^{(1)})}$$

$$[0084] \quad \mathbf{x}^{(1)} = \sum_{r=1}^{M_1} a_r^{(1)} \mathbf{e}_r^{(1)}$$

[0085] 其中, $r=1,2,\dots,M_1$, $\mathbf{e}_r^{(1)} \in \mathbb{R}^{N \times 1}$ 表示第 r 组脑电信号特征向量, $\mathbf{W}_r^{(1)} \in \mathbb{R}^{M_1 \times 1}$ 为可训练的线性变换参数向量, $\exp(\cdot)$ 表示以自然常数 e 为底的指数函数。在本实施例中, $M_1=8$, $N=32$ 。

[0086] 为了训练线性变换矩阵 $\mathbf{W}^{(1)}$ 的参数,需要在第一个注意力机制模块之后连接一个softmax分类器,将第一个注意力机制模块输出的脑电情感特征向量 $\mathbf{x}^{(1)}$ 连接到softmax分

类器的C个输出节点,经过softmax函数之后输出一个概率分布向量 $\hat{y}_c^{(1)}$,其中 $c \in [1, C]$,C为情感类别数。

[0087] 进一步地,由下式所示的交叉熵损失函数来训练线性变换矩阵 $W^{(1)}$ 的参数。

$$[0088] \quad \hat{y}_c^{(1)} = \text{softmax}(\mathbf{x}^{(1)})$$

$$[0089] \quad \text{Loss}^{(1)} = -\frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C y_{m,c}^{(1)} \log(\hat{y}_{m,c}^{(1)})$$

[0090] 其中, $\mathbf{x}^{(1)}$ 为32维的脑电情感特征向量; $\hat{y}_c^{(1)}$ 表示softmax分类模型预测情感类别的概率分布向量; $y_{m,c}^{(1)}$ 表示第m个脑电样本的真实情感类别标签,当采用one-hot编码时,若第m个脑电样本的真实情感类别标签为c,则 $y_{m,c}^{(1)}=1$,否则 $y_{m,c}^{(1)}=0$; $\hat{y}_{m,c}^{(1)}$ 表示softmax分类模型将第m个脑电样本预测为类别c的概率;Loss⁽¹⁾表示线性变换矩阵 $W^{(1)}$ 在训练过程中的损失函数;在本实施例中,C=2,M=880。

[0091] 通过误差反向传播算法不断迭代训练,直至模型参数达到最优。之后,就可对新输入的测试样本的脑电信号提取脑电情感特征向量 $\mathbf{x}^{(1)}$ 。

[0092] (2.2)将步骤(1)提取到的外周生理信号特征以矩阵形式表示成

$\mathbf{F}^{(2)} = [\mathbf{f}_1^{(2)}, \mathbf{f}_2^{(2)}, \dots, \mathbf{f}_N^{(2)}] \in \mathbb{R}^{M_2 \times N}$,并通过线性变换矩阵 $W^{(2)}$ 映射成 M_2 组特征向量

$(\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \dots, \mathbf{e}_{M_2}^{(2)})$, $4 \leq M_2 \leq 16$,令 $\mathbf{W}^{(2)} = [\mathbf{W}_1^{(2)}, \mathbf{W}_2^{(2)}, \dots, \mathbf{W}_{M_2}^{(2)}] \in \mathbb{R}^{M_2 \times M_2}$,

$\mathbf{E}^{(2)} = [\mathbf{e}_1^{(2)}, \mathbf{e}_2^{(2)}, \dots, \mathbf{e}_{M_2}^{(2)}] \in \mathbb{R}^{N \times M_2}$,其线性变换表达式为:

$$[0093] \quad \mathbf{E}^{(2)} = (\mathbf{F}^{(2)})^T \mathbf{W}^{(2)}$$

[0094] 其中,上标(2)代表外周生理模态。

[0095] 使用第二个注意力机制模块来确定不同特征向量组的重要性权重,通过加权融合形成有鉴别力的外周生理情感特征向量,其中第s组外周生理信号特征向量的权重 $a_s^{(2)}$ 以及外周生理情感特征向量 $\mathbf{x}^{(2)}$ 表示为:

$$[0096] \quad a_s^{(2)} = \frac{\exp((\mathbf{f}_s^{(2)})^T \mathbf{W}_s^{(2)})}{\sum_{s=1}^{M_2} \exp((\mathbf{f}_s^{(2)})^T \mathbf{W}_s^{(2)})}$$

$$[0097] \quad \mathbf{x}^{(2)} = \sum_{s=1}^{M_2} a_s^{(2)} \mathbf{e}_s^{(2)}$$

[0098] 其中, $s=1, 2, \dots, M_2$, $\mathbf{e}_s^{(2)} \in \mathbb{R}^{N \times 1}$ 表示第s组外周生理信号特征向量, $\mathbf{W}_s^{(2)} \in \mathbb{R}^{M_2 \times 1}$ 为可训练的线性变换参数向量。在本实施例中, $M_2=4$ 。

[0099] 为了训练线性变换矩阵 $W^{(2)}$ 的参数,需要在第二个注意力机制模块之后连接一个softmax分类器,将第二个注意力机制模块输出的外周生理情感特征向量 $\mathbf{x}^{(2)}$ 连接到softmax分类器的C个输出节点,经过softmax函数之后输出一个概率分布向量 $\hat{y}_c^{(2)}$ 。

[0100] 进一步地,由下式所示的交叉熵损失函数来训练线性变换矩阵 $W^{(2)}$ 的参数。

$$[0101] \quad \hat{\mathbf{y}}_c^{(2)} = \text{softmax}(\mathbf{x}^{(2)})$$

$$[0102] \quad \text{Loss}^{(2)} = -\frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C y_{m,c}^{(2)} \log(\hat{y}_{m,c}^{(2)})$$

[0103] 其中, $\mathbf{x}^{(2)}$ 为32维的外周生理情感特征向量; $\hat{\mathbf{y}}_c^{(2)}$ 表示softmax分类模型预测情感类别的概率分布向量; $y_{m,c}^{(2)}$ 表示第m个外周生理信号样本的真实情感类别标签, 当采用one-hot编码时, 若第m个外周生理信号样本的真实情感类别标签为c, 则 $y_{m,c}^{(2)}=1$, 否则 $y_{m,c}^{(2)}=0$; $\hat{y}_{m,c}^{(2)}$ 表示softmax分类模型将第m个外周生理信号样本预测为类别c的概率; $\text{Loss}^{(2)}$ 表示线性变换矩阵 $\mathbf{W}^{(2)}$ 在训练过程中的损失函数; 在本实施例中, $C=2, M=880$ 。

[0104] 通过误差反向传播算法不断迭代训练, 直至模型参数达到最优。之后, 就可对新输入的测试样本的外周生理信号提取外周生理情感特征向量 $\mathbf{x}^{(2)}$ 。

[0105] (2.3) 将步骤(1)提取到的表情特征以矩阵形式表示成 $\mathbf{F}^{(3)} = [\mathbf{f}_1^{(3)}, \mathbf{f}_2^{(3)}, \dots, \mathbf{f}_N^{(3)}] \in \mathbb{R}^{M_3 \times N}$, 并通过线性变换矩阵 $\mathbf{W}^{(3)}$ 映射成 M_3 组特征向量 $(\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \dots, \mathbf{e}_{M_3}^{(3)})$, $4 \leq M_3 \leq 16$, 令 $\mathbf{W}^{(3)} = [\mathbf{W}_1^{(3)}, \mathbf{W}_2^{(3)}, \dots, \mathbf{W}_{M_3}^{(3)}] \in \mathbb{R}^{M_3 \times M_3}$,

$\mathbf{E}^{(3)} = [\mathbf{e}_1^{(3)}, \mathbf{e}_2^{(3)}, \dots, \mathbf{e}_{M_3}^{(3)}] \in \mathbb{R}^{N \times M_3}$, 其线性变换表达式为:

$$[0106] \quad \mathbf{E}^{(3)} = (\mathbf{F}^{(3)})^T \mathbf{W}^{(3)}$$

[0107] 其中, 上标(3)代表表情模态。

[0108] 使用第三个注意力机制模块来确定不同特征向量组的重要性权重, 通过加权融合形成有鉴别力的表情情感特征向量, 其中第t组表情特征向量的权重 $a_t^{(3)}$ 以及表情情感特征向量 $\mathbf{x}^{(3)}$ 表示为:

$$[0109] \quad a_t^{(3)} = \frac{\exp((\mathbf{f}_t^{(3)})^T \mathbf{W}_t^{(3)})}{\sum_{t=1}^{M_3} \exp((\mathbf{f}_t^{(3)})^T \mathbf{W}_t^{(3)})}$$

$$[0110] \quad \mathbf{x}^{(3)} = \sum_{t=1}^{M_3} a_t^{(3)} \mathbf{e}_t^{(3)}$$

[0111] 其中, $t=1, 2, \dots, M_3$, $\mathbf{e}_t^{(3)} \in \mathbb{R}^{N \times 1}$ 表示第t组表情特征向量, $\mathbf{W}_t^{(3)} \in \mathbb{R}^{M_3 \times 1}$ 为可训练的线性变换参数向量。在本实施例中, $M_3=8$ 。

[0112] 为了训练线性变换矩阵 $\mathbf{W}^{(3)}$ 的参数, 需要在第三个注意力机制模块之后连接一个softmax分类器, 将第三个注意力机制模块输出的表情情感特征向量 $\mathbf{x}^{(3)}$ 连接到softmax分类器的C个输出节点, 经过softmax函数之后输出一个概率分布向量 $\hat{\mathbf{y}}_c^{(3)}$ 。

[0113] 进一步地, 由下式所示的交叉熵损失函数来训练线性变换矩阵 $\mathbf{W}^{(3)}$ 的参数。

$$[0114] \quad \hat{\mathbf{y}}_c^{(3)} = \text{softmax}(\mathbf{x}^{(3)})$$

$$[0115] \quad Loss^{(3)} = -\frac{1}{M} \sum_{m=1}^M \sum_{c=1}^C y_{m,c}^{(3)} \log(\hat{y}_{m,c}^{(3)})$$

[0116] 其中, $x^{(3)}$ 为32维的表情情感特征向量; $\hat{y}_c^{(3)}$ 表示softmax分类模型预测情感类别的概率分布向量; $y_{m,c}^{(3)}$ 表示第m个表情视频样本的真实情感类别标签, 当采用one-hot编码时, 若第m个表情视频样本的真实情感类别标签为c, 则 $y_{m,c}^{(3)}=1$, 否则 $y_{m,c}^{(3)}=0$; $\hat{y}_{m,c}^{(3)}$ 表示softmax分类模型将第m个表情视频样本预测为类别c的概率; $Loss^{(3)}$ 表示线性变换矩阵 $W^{(3)}$ 在训练过程中的损失函数; 在本实施例中, $C=2, M=880$ 。

[0117] 通过误差反向传播算法不断迭代训练, 直至模型参数达到最优。之后, 就可对新输入的测试样本的表情视频提取表情情感特征向量 $x^{(3)}$ 。

[0118] 进一步地, 步骤(3)具体包括以下子步骤:

[0119] (3.1) 获取通过训练得到的分别对应于脑电情感特征、外周生理情感特征和表情情感特征的DMCCA投影矩阵 $\Omega = [\omega_1^{(1)}, \omega_2^{(1)}, \dots, \omega_d^{(1)}] \in \mathbb{R}^{N \times d}$ 、 $\Phi = [\omega_1^{(2)}, \omega_2^{(2)}, \dots, \omega_d^{(2)}] \in \mathbb{R}^{N \times d}$ 和 $\Psi = [\omega_1^{(3)}, \omega_2^{(3)}, \dots, \omega_d^{(3)}] \in \mathbb{R}^{N \times d}$, $32 \leq d \leq 128$ 。在本实施例中, $d=40$ 。

[0120] (3.2) 分别使用投影矩阵 Ω 、 Φ 和 Ψ 将步骤(2)提取到的脑电情感特征向量 $x^{(1)}$ 、外周生理情感特征向量 $x^{(2)}$ 和表情情感特征向量 $x^{(3)}$ 投影到一个d维的公共子空间, 其中脑电情感特征向量 $x^{(1)}$ 到d维公共子空间的投影为 $\Omega^T x^{(1)}$, 外周生理情感特征向量 $x^{(2)}$ 到d维公共子空间的投影为 $\Phi^T x^{(2)}$, 表情情感特征向量 $x^{(3)}$ 到d维公共子空间的投影为 $\Psi^T x^{(3)}$ 。

[0121] (3.3) 将 $\Omega^T x^{(1)}$ 、 $\Phi^T x^{(2)}$ 和 $\Psi^T x^{(3)}$ 进行融合, 得到脑电-外周生理-表情多模态情感特征向量为 $\Omega^T x^{(1)} + \Phi^T x^{(2)} + \Psi^T x^{(3)}$ 。

[0122] 进一步地, 步骤(3.1)中的投影矩阵 Ω 、 Φ 和 Ψ 通过以下步骤的训练得到:

[0123] (3.1.1) 对于训练样本集中的C类情感类别的样本生成3组情感特征向量 $\{X^{(i)} \in \mathbb{R}^{N \times M}\}_{i=1}^3$, 其中 $X^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_M^{(i)}]$, M为训练样本数(本例中样本集中数据量不大, 所有样本参与计算, 数据量大的样本集可随机抽取各情感类别的样本), $i=1, 2, 3, m=1, 2, \dots, M$; 令 $i=1$ 代表脑电模态, $i=2$ 代表外周生理模态, $i=3$ 代表表情模态, $x_m^{(1)}$ 代表脑电情感特征向量, $x_m^{(2)}$ 代表外周生理情感特征向量, $x_m^{(3)}$ 代表表情情感特征向量; 在本实施例中, $C=2, M=880, N=32$ 。

[0124] (3.1.2) 计算 $X^{(i)}$ 中各列向量的均值 $\mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_M^{(i)}$, 对 $X^{(i)}$ 进行中心化操作, 得到 $\bar{X}^{(i)} = (x_1^{(i)} - \mu_1^{(i)}, x_2^{(i)} - \mu_2^{(i)}, \dots, x_M^{(i)} - \mu_M^{(i)})$, $i=1, 2, 3$, 为了便于描述, 下面将中心化后的 $\bar{X}^{(i)}$ 仍记为 $X^{(i)}$, 即假设 $\{X^{(i)}\}_{i=1}^3$ 均已被中心化。

[0125] (3.1.3) 鉴别多重集典型相关分析(DMCCA)的思想旨在求得一组投影矩阵 Ω 、 Φ 和 Ψ , 使得同类样本在公共投影子空间的线性相关性最大, 同时还最大化了模态内数据的类间散布与最小化了模态内数据的类内散布, 令 $X^{(i)}$ 的投影向量为 $w^{(i)} \in \mathbb{R}^{N \times 1}$, $i=1, 2, 3$, DMCCA的目标函数为:

$$[0126] \quad J(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}) = \frac{\sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)}}{\sqrt{\sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)}}}$$

[0127] 其中, \mathbf{S}_i^w 表示 $X^{(i)}$ 的类内散布矩阵, $\mathbf{S}_{ij} = \begin{cases} \mathbf{S}_i^b, & i=j \\ \text{cov}(\mathbf{S}_i^b, \mathbf{S}_j^b), & i \neq j \end{cases}$, \mathbf{S}_i^b 表示 $X^{(i)}$ 的类间散布矩阵, $\text{cov}(\cdot, \cdot)$ 表示协方差, $i, j \in \{1, 2, 3\}$ 。

[0128] 对DMCCA目标函数的求解可以表示为如下的优化模型:

$$[0129] \quad \begin{cases} \max_{\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}} \sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)} \\ \text{s.t.} \quad \sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)} = 1 \end{cases}$$

[0130] (3.1.4) 使用拉格朗日乘子法 (Lagrange multiplier) 求解DMCCA目标函数的优化模型, 可得到如下拉格朗日 (Lagrange) 函数:

$$[0131] \quad L(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}) = \sum_{i=1}^3 \sum_{j=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_{ij} \mathbf{w}^{(j)} - \lambda \left(\sum_{i=1}^3 (\mathbf{w}^{(i)})^T \mathbf{S}_i^w \mathbf{w}^{(i)} - 1 \right)$$

[0132] 其中, λ 是拉格朗日乘子, 再分别求 $L(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)})$ 对 $\mathbf{w}^{(1)}$ 、 $\mathbf{w}^{(2)}$ 和 $\mathbf{w}^{(3)}$ 的偏导数并令其为零, 即令

$$[0133] \quad \frac{\partial L}{\partial \mathbf{w}^{(i)}} = \sum_{j=1}^3 \mathbf{S}_{ij} \mathbf{w}^{(j)} - \lambda \mathbf{S}_i^w \mathbf{w}^{(i)} = 0, \quad i=1, 2, 3$$

[0134] 得到

$$[0135] \quad \sum_{j=1}^3 \mathbf{S}_{ij} \mathbf{w}^{(j)} = \lambda \mathbf{S}_i^w \mathbf{w}^{(i)}, \quad i=1, 2, 3$$

[0136] 进一步对上式作简化处理, 则可获得如下的广义特征值问题:

$$[0137] \quad \begin{bmatrix} \mathbf{S}_1^b & \mathbf{S}_{12} & \mathbf{S}_{13} \\ \mathbf{S}_{21} & \mathbf{S}_2^b & \mathbf{S}_{23} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_3^b \end{bmatrix} \begin{bmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \\ \mathbf{w}^{(3)} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{S}_1^w & 0 & 0 \\ 0 & \mathbf{S}_2^w & 0 \\ 0 & 0 & \mathbf{S}_3^w \end{bmatrix} \begin{bmatrix} \mathbf{w}^{(1)} \\ \mathbf{w}^{(2)} \\ \mathbf{w}^{(3)} \end{bmatrix}$$

[0138] 通过求解上式中的广义特征值问题, 选取前 d 个最大特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 对应的特征向量, 即可得到投影矩阵 $\mathbf{\Omega} = [\mathbf{w}_1^{(1)}, \mathbf{w}_2^{(1)}, \dots, \mathbf{w}_d^{(1)}] \in \mathbb{R}^{N \times d}$ 、 $\mathbf{\Phi} = [\mathbf{w}_1^{(2)}, \mathbf{w}_2^{(2)}, \dots, \mathbf{w}_d^{(2)}] \in \mathbb{R}^{N \times d}$ 和 $\mathbf{\Psi} = [\mathbf{w}_1^{(3)}, \mathbf{w}_2^{(3)}, \dots, \mathbf{w}_d^{(3)}] \in \mathbb{R}^{N \times d}$ 。在本实施例中, $d=40$ 。

[0139] 基于相同的发明构思, 本发明实施例提供的融合注意力机制和DMCCA的多模态情感识别系统, 包括:

[0140] 特征初步提取模块, 用于对经过预处理后的脑电信号和面部表情视频分别使用各自训练好的神经网络模型提取脑电信号特征向量和表情特征向量, 对预处理后的外周生理信号, 通过抽取信号波形描述符及其统计特征, 提取外周生理信号特征向量;

[0141] 特征鉴别增强模块, 用于对的脑电信号特征向量、外周生理信号特征向量、表情特

征向量分别通过线性变换矩阵映射成若干组特征向量,并分别使用注意力机制模块确定不同特征向量组的重要性权重,通过加权融合形成维数相同的有鉴别力的脑电情感特征向量、外周生理情感特征向量、表情情感特征向量;

[0142] 投影矩阵确定模块,用于使用DMCCA方法,通过最大化同一类别样本的不同模态情感特征之间的相关性,确定各情感特征向量的投影矩阵;

[0143] 特征融合模块,用于对的脑电情感特征向量、外周生理情感特征向量和表情情感特征向量,通过各自对应的投影矩阵投影到一个公共子空间,相加融合后得到脑电-外周生理-表情多模态情感特征向量;

[0144] 以及,分类识别模块,用于使用分类器对多模态情感特征向量进行分类识别,得到情感类别。

[0145] 各模块的具体实现参考上述方法实施例,不再赘述。本领域技术人员可以理解,可以对实施例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系统中。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,以及此外可以把它们分成多个子模块或子单元或子组件。

[0146] 基于相同的发明构思,本发明实施例提供的融合注意力机制和DMCCA的多模态情感识别系统,包括至少一台计算设备,该计算设备包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,该计算机程序被加载至处理器时实现上述的融合注意力机制和DMCCA的多模态情感识别方法。

[0147] 本发明所公开的技术方案既包含了上述实施方案中涉及的技术方法,也包括由以上技术方法任意组合成的技术方案。本技术领域的普通技术人员,在不脱离本发明的原理的前提下,可以作出一定的改善和修饰,这些改善和修饰也被认为本发明的保护范围。

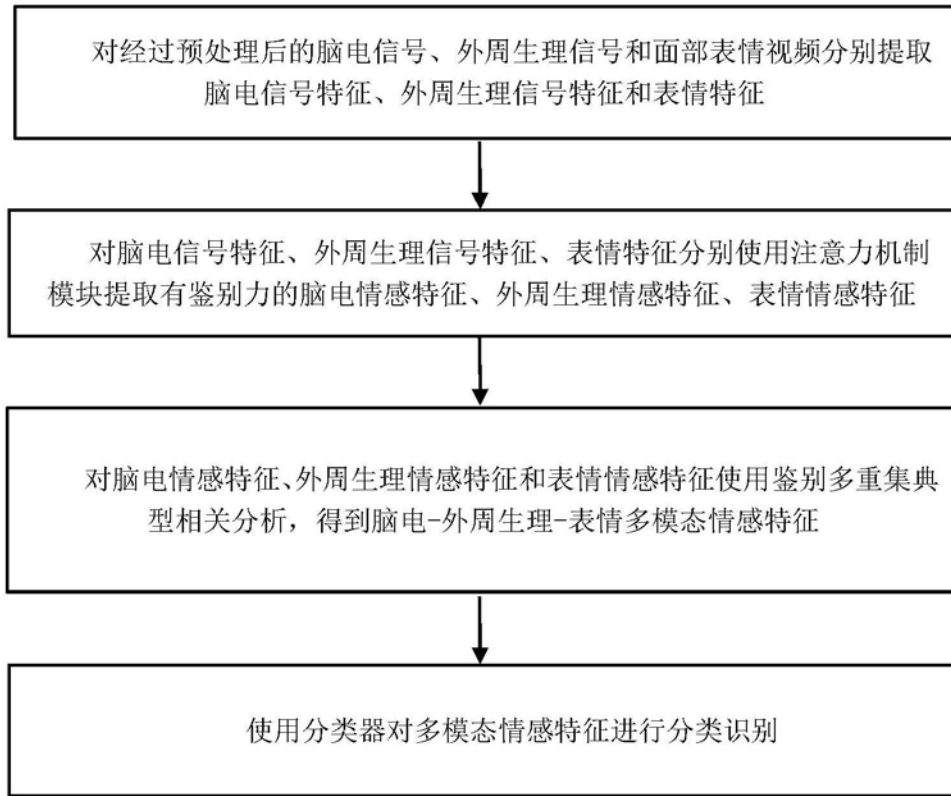


图1

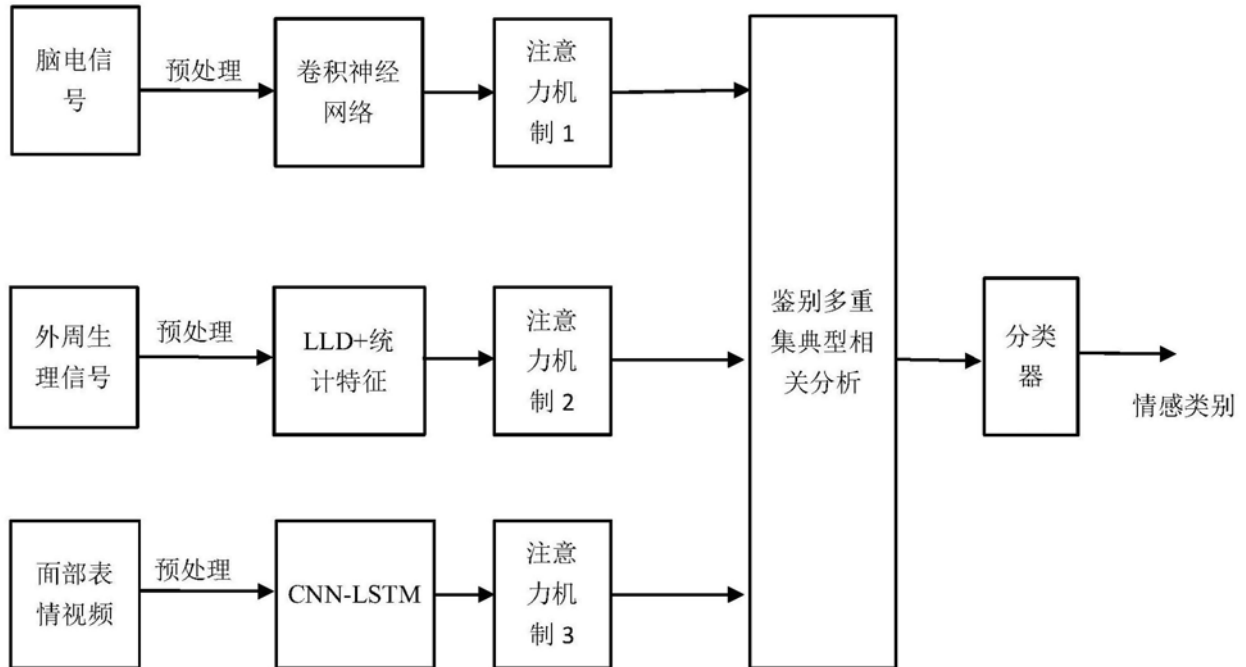


图2