



(12) 发明专利申请

(10) 申请公布号 CN 114743235 A

(43) 申请公布日 2022. 07. 12

(21) 申请号 202210196086.4

(22) 申请日 2022.03.01

(71) 申请人 东南大学

地址 211102 江苏省南京市江宁区东南大学路2号

(72) 发明人 郑文明 魏梦婷 江星洵 宗源 常洪丽

(74) 专利代理机构 南京苏高专利商标事务所 (普通合伙) 32204

专利代理人 冯艳芬

(51) Int. Cl.

G06V 40/16 (2022.01)

G06V 10/774 (2022.01)

G06K 9/62 (2022.01)

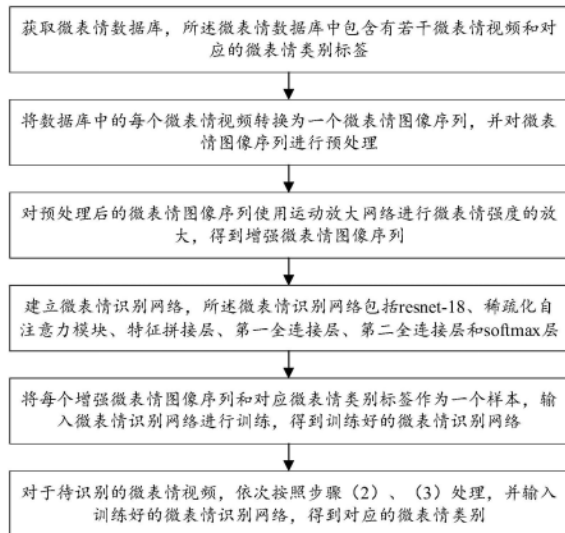
权利要求书4页 说明书9页 附图2页

(54) 发明名称

基于稀疏化自注意力机制的微表情识别方法及系统

(57) 摘要

本发明公开了一种基于稀疏化自注意力机制的微表情识别方法及系统,方法包括:(1)获取微表情数据库(2)将微表情视频转换为微表情图像序列并进行预处理;(3)对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列;(4)建立微表情识别网络,所述微表情识别网络包括resnet-18、稀疏化自注意力模块、特征拼接层、第一全连接层、第二全连接层和softmax层(5)训练微表情识别网络;(6)对于待识别的微表情视频,依次按照步骤(2)、(3)处理,并输入训练好的微表情识别网络,得到对应的微表情类别。本发明识别准确率更高。



1. 一种基于稀疏化自注意力机制的微表情识别方法,其特征在于该方法包括:

(1) 获取微表情数据库,所述微表情数据库中包含有若干微表情视频和对应的微表情类别标签;

(2) 将数据库中的每个微表情视频转换为一个微表情图像序列,并对微表情图像序列进行预处理;

(3) 对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列;

(4) 建立微表情识别网络,所述微表情识别网络包括resnet-18、稀疏化自注意力模块、特征拼接层、第一全连接层、第二全连接层和softmax层,其中,所述resnet-18用于从增强微表情图像序列中提取空间特征,所述稀疏化自注意力模块用于从空间特征中获取运动变化特征,所述特征拼接层用于将空间特征和运动变化特征拼接为综合特征后输入至第一全连接层,第一全连接层和第二全连接层用于依次将综合特征降维,所述softmax层用于根据降维后的综合特征进行识别分类,得到微表情类别;

(5) 将每个增强微表情图像序列和对应微表情类别标签作为一个样本,输入微表情识别网络进行训练,得到训练好的微表情识别网络;

(6) 对于待识别的微表情视频,依次按照步骤(2)、(3)处理,并输入训练好的微表情识别网络,得到对应的微表情类别。

2. 根据权利要求1所述的基于稀疏化自注意力机制的微表情识别方法,其特征在于:步骤(2)中所述预处理包括人脸配准和面部区域剪切。

3. 根据权利要求1所述的基于稀疏化自注意力机制的微表情识别方法,其特征在于:步骤(3)具体包括:

提取预处理后的微表情图像序列的起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ ,并根据起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ 使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列 $I_{mag}$ :

$$I_{mag} = \{I_{mag,i}\}$$

$$I_{mag,i} = f(I_{onset} + (1+\alpha_i) | I_{apex} - I_{onset} |)$$

$I_{mag,i}$ 为预处理后微的表情图像序列中第i帧图像放大之后的图像, $f(\cdot)$ 为运动放大网络, $\alpha_i$ 为第i帧图像的放大倍数,放大倍数按照微表情图像序列的排列顺序从前到后依次递增。

4. 根据权利要求1所述的基于稀疏化自注意力机制的微表情识别方法,其特征在于:所述稀疏化自注意力模块具体包括:

基本矩阵生成单元,用于根据空间特征 $F^T$ 采用下式计算自注意力矩阵的查询矩阵Q、键矩阵K、值矩阵V:

$$Q = F^T \times W^Q$$

$$K = F^T \times W^K$$

$$V = F^T \times W^V$$

式中, $W^Q$ 、 $W^K$ 、 $W^V$ 分别为对应的权重矩阵;

超平面生成单元,用于在空间中随机地生成 $n\_rounds$ 个超平面 $\{h^{(1)}, h^{(2)}, \dots, h^{(n\_rounds)}\}$ , $n\_rounds$ 为大于2的正整数; $h^{(*)}$ 为第\*个超平面;

特征哈希单元,用于基于超平面将空间分为 $2^{(n\_rounds)}$ 个哈希桶,采用下式分别计算Q、K

中每一个特征向量所落入的哈希桶；

$$h(q_u) = a^{(1)}a^{(2)} \dots a^{(n_{rounds})}, u = 1, 2, \dots, length$$

$$a^{(j)} = \begin{cases} 1, q_u^T h^{(j)} > 0 \\ 0, q_u^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n_{rounds}$$

$$h(k_v) = b^{(1)}b^{(2)} \dots b^{(n_{rounds})}, v = 1, 2, \dots, length$$

$$b^{(j)} = \begin{cases} 1, k_v^T h^{(j)} > 0 \\ 0, k_v^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n_{rounds}$$

式中,  $q_u$ 、 $k_v$  表示 Q、K 中第 u、v 个特征向量, length 表示增强微表情图像序列的帧数,  $h(q_u)$ 、 $h(k_v)$  表示  $q_u$ 、 $k_v$  所落入的哈希桶的二进制标号,  $a^{(j)}$ 、 $b^{(j)}$  表示  $h(q_u)$ 、 $h(k_v)$  的第 j 个比特值；

特征稀疏单元, 用于只保留 Q、K 中落入同一哈希桶的特征向量的索引号, 存入稀疏后的特征索引集合  $\Omega_{u,v}$  中:

$$\Omega_{u,v} = \{(u, v) \mid h(q_u) = h(k_v)\}, u, v = 1, 2, \dots, length$$

稀疏注意力矩阵计算单元, 用于根据特征索引集合  $\Omega_{i,j}$  计算稀疏注意力矩阵 Attention(Q, K, V):

$$Attention(Q, K, V) = \sum_{(u,v) \in \Omega_{u,v}} softmax\left(\frac{q_u k_v^T}{\sqrt{d_l}}\right) V$$

式中,  $d_l$  为 Q、K、V 的维度, Attention(Q, K, V) 即为运动变化特征。

5. 根据权利要求 1 所述的基于稀疏化自注意力机制的微表情识别方法, 其特征在于: 步骤 (5) 中训练时采用的损失函数为:

$$L = - \sum_{c=1}^C y_c \log(p_c)$$

式中, C 为类别数,  $c = 1, 2, \dots, C$ ,  $y_c$  为微表情的类别标签,  $p_c$  为属于类别 c 的概率。

6. 一种基于稀疏化自注意力机制的微表情识别系统, 其特征在于包括:

微表情数据库, 包含有若干微表情视频和对应的微表情类别标签;

预处理模块, 用于将数据库中的每个微表情视频转换为一个微表情图像序列, 并对微表情图像序列进行预处理;

强度放大模块, 用于对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大, 得到增强微表情图像序列;

微表情识别网络建立模块, 用于建立微表情识别网络, 所述微表情识别网络包括 resnet-18、稀疏化自注意力模块、特征拼接层、第一全连接层、第二全连接层和 softmax 层, 其中, 所述 resnet-18 用于从增强微表情图像序列中提取空间特征, 所述稀疏化自注意力模块用于从空间特征中获取运动变化特征, 所述特征拼接层用于将空间特征和运动变化特征拼接为综合特征后输入至第一全连接层, 第一全连接层和第二全连接层用于依次将综合特征降维, 所述 softmax 层用于根据降维后的综合特征进行识别分类, 得到微表情类别;

网络训练模块, 用于将每个增强微表情图像序列和对应微表情类别标签作为一个样

本,输入微表情识别网络进行训练,得到训练好的微表情识别网络;

微表情识别模块,用于对于待识别的微表情视频,依次预处理和强度放大,并输入训练好的微表情识别网络,得到对应的微表情类别。

7.根据权利要求6所述的基于稀疏化自注意力机制的微表情识别系统,其特征在于:所述预处理模块具体用于将数据库中的每个微表情视频转换为一个微表情图像序列,并进行人脸配准和面部区域剪切。

8.根据权利要求6所述的基于稀疏化自注意力机制的微表情识别系统,其特征在于:所述强度放大模块具体用于提取预处理后的微表情图像序列的起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ ,并根据起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ 使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列 $I_{mag}$ :

$$I_{mag} = \{I_{mag,i}\}$$

$$I_{mag,i} = f(I_{onset} + (1+\alpha_i) |I_{apex} - I_{onset}|)$$

$I_{mag,i}$ 为预处理后微的表情图像序列中第 $i$ 帧图像放大之后的图像, $f(\cdot)$ 为运动放大网络, $\alpha_i$ 为第 $i$ 帧图像的放大倍数,放大倍数按照微表情图像序列的排列顺序从前到后依次递增。

9.根据权利要求6所述的基于稀疏化自注意力机制的微表情识别系统,其特征在于:所述稀疏化自注意力模块具体包括:

基本矩阵生成单元,用于根据空间特征 $F^T$ 采用下式计算自注意力矩阵的查询矩阵 $Q$ 、键矩阵 $K$ 、值矩阵 $V$ :

$$Q = F^T \times W^Q$$

$$K = F^T \times W^K$$

$$V = F^T \times W^V$$

式中, $W^Q$ 、 $W^K$ 、 $W^V$ 分别为对应的权重矩阵;

超平面生成单元,用于在空间中随机地生成 $n\_rounds$ 个超平面 $\{h^{(1)}, h^{(2)}, \dots, h^{(n\_rounds)}\}$ , $n\_rounds$ 为大于2的正整数; $h^{(*)}$ 为第 $*$ 个超平面;

特征哈希单元,用于基于超平面将空间分为 $2^{(n\_rounds)}$ 个哈希桶,采用下式分别计算 $Q$ 、 $K$ 中每一个特征向量所落入的哈希桶;

$$h(q_u) = a^{(1)}a^{(2)} \dots a^{(n\_rounds)}, u = 1, 2, \dots, length$$

$$a^{(j)} = \begin{cases} 1, & q_u^T h^{(j)} > 0 \\ 0, & q_u^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

$$h(k_v) = b^{(1)}b^{(2)} \dots b^{(n\_rounds)}, v = 1, 2, \dots, length$$

$$b^{(j)} = \begin{cases} 1, & k_v^T h^{(j)} > 0 \\ 0, & k_v^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

式中, $q_u$ 、 $k_v$ 表示 $Q$ 、 $K$ 中第 $u$ 、 $v$ 个特征向量, $length$ 表示增强微表情图像序列的帧数, $h(q_u)$ 、 $h(k_v)$ 表示 $q_u$ 、 $k_v$ 所落入的哈希桶的二进制标号, $a^{(j)}$ 、 $b^{(j)}$ 表示 $h(q_u)$ 、 $h(k_v)$ 的第 $j$ 个比特值;

特征稀疏单元,用于只保留 $Q$ 、 $K$ 中落入同一哈希桶的特征向量的索引号,存入稀疏后的特征索引集合 $\Omega_{u,v}$ 中:

$\Omega_{u,v} = \{(u,v) \mid h(q_u) = h(k_v)\}, u, v = 1, 2, \dots, \text{length}$

稀疏注意力矩阵计算单元,用于根据特征索引集合  $\Omega_{i,j}$  计算稀疏注意力矩阵 Attention(Q,K,V):

$$\text{Attention}(Q, K, V) = \sum_{(u,v) \in \Omega_{u,v}} \text{softmax}\left(\frac{q_u k_v^T}{\sqrt{d_l}}\right) V$$

式中,  $d_l$  为 Q、K、V 的维度, Attention(Q,K,V) 即为运动变化特征。

10. 根据权利要求6所述的基于稀疏化自注意力机制的微表情识别系统,其特征在於: 所述网络训练模块中训练时采用的损失函数为:

$$L = - \sum_{c=1}^C y_c \log(p_c)$$

式中, C 为类别数,  $c = 1, 2, \dots, C$ ,  $y_c$  为微表情的类别标签,  $p_c$  为属于类别 c 的概率。

## 基于稀疏化自注意力机制的微表情识别方法及系统

### 技术领域

[0001] 本发明涉及表情识别技术,尤其涉及一种基于稀疏化自注意力机制的微表情识别方法及系统。

### 背景技术

[0002] 微表情是人类试图隐藏或抑制内心真实情绪时,不经意间露出的面部表情,不受人主观意识的控制。微表情是检测人类隐藏情绪时重要的非语言信号,通常能够有效揭示一个人的真实心理状态,被认为是识别谎言的关键线索,对更好地理解人类情感,有着重要作用。因此,微表情的有效运用,对社会生产生活有着重要作用。在刑侦方面,经过一定微表情识别能力训练的审讯员,可以更好地识别犯罪嫌疑人的谎言;在社会治安方面,通过观察微表情可以判断潜伏在日常生活中的危险分子,防恐防暴;在临床医疗方面,通过微表情,医生更好地了解患者的真实想法,例如隐藏病情等等,从而更加有效地与患者交流,更加准确地分析病情,改进治疗方案。但是,人工识别微表情培训成本较高,难以大规模推广。因此近年来,利用计算机视觉技术和人工智能方法,进行微表情识别的需求日益增加。

[0003] 由于微表情涉及运动强度低的特性,传统的识别任务在特征提取之前会利用放大手段放大微表情的强度,如欧拉视频放大,拉格朗日放大等。这些方法在放大过程中容易引入噪声和斑点,对图像其他区域的恢复效果并不理想,近年来,深度学习的发展进入到图像,语音,自然语言处理等各个方面,基于学习的运动放大方法也逐渐增加,相比于传统的手工方法,深度学习的方法对图像的重建效果更好。

[0004] 无论是传统的放大方式还是基于深度学习的放大方式,都面临这样一个问题:人与人之间的面部构造差异很大,表情与表情的呈现所涉及的肌肉运动也不尽相同,如果对于所有样本都应用统一的放大强度,很可能造成有的微表情放大不充分,有的放大程度过大导致脸部变形,这样获得的放大图像很可能影响后续的识别结果,识别准确率有待提高。

### 发明内容

[0005] 发明目的:本发明针对现有技术存在的问题,提供一种识别准确率更高的基于稀疏化自注意力机制的微表情识别方法及系统。

[0006] 技术方案:本发明所述的基于稀疏化自注意力机制的微表情识别方法包括:

[0007] (1) 获取微表情数据库,所述微表情数据库中包含有若干微表情视频和对应的微表情类别标签;

[0008] (2) 将数据库中的每个微表情视频转换为一个微表情图像序列,并对微表情图像序列进行预处理;

[0009] (3) 对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列;

[0010] (4) 建立微表情识别网络,所述微表情识别网络包括resnet-18、稀疏化自注意力

模块、特征拼接层、第一全连接层、第二全连接层和softmax层,其中,所述resnet-18用于从增强微表情图像序列中提取空间特征,所述稀疏化自注意力模块用于从空间特征中获取运动变化特征,所述特征拼接层用于将空间特征和运动变化特征拼接为综合特征后输入至第一全连接层,第一全连接层和第二全连接层用于依次将综合特征降维,所述softmax层用于根据降维后的综合特征进行识别分类,得到微表情类别;

[0011] (5) 将每个增强微表情图像序列和对应微表情类别标签作为一个样本,输入微表情识别网络进行训练,得到训练好的微表情识别网络;

[0012] (6) 对于待识别的微表情视频,依次按照步骤(2)、(3)处理,并输入训练好的微表情识别网络,得到对应的微表情类别。

[0013] 进一步的,步骤(2)中所述预处理包括人脸配准和面部区域剪切。

[0014] 进一步的,步骤(3)具体包括:

[0015] 提取预处理后的微表情图像序列的起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ ,并根据起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ 使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列 $I_{mag}$ :

$$[0016] \quad I_{mag} = \{I_{mag,i}\}$$

$$[0017] \quad I_{mag,i} = f(I_{onset} + (1 + \alpha_i) | I_{apex} - I_{onset} |)$$

[0018]  $I_{mag,i}$ 为预处理后微的表情图像序列中第*i*帧图像放大之后的图像, $f(\cdot)$ 为运动放大网络, $\alpha_i$ 为第*i*帧图像的放大倍数,放大倍数按照微表情图像序列的排列顺序从前到后依次递增。

[0019] 进一步的,所述稀疏化自注意力模块具体包括:

[0020] 基本矩阵生成单元,用于根据空间特征 $F^T$ 采用下式计算自注意力矩阵的查询矩阵Q、键矩阵K、值矩阵V:

$$[0021] \quad Q = F^T \times W^Q$$

$$[0022] \quad K = F^T \times W^K$$

$$[0023] \quad V = F^T \times W^V$$

[0024] 式中, $W^Q$ 、 $W^K$ 、 $W^V$ 分别为对应的权重矩阵;

[0025] 超平面生成单元,用于在空间中随机地生成 $n\_rounds$ 个超平面 $\{h^{(1)}, h^{(2)}, \dots, h^{(n\_rounds)}\}$ , $n\_rounds$ 为大于2的正整数; $h^{(*)}$ 为第\*个超平面;

[0026] 特征哈希单元,用于基于超平面将空间分为 $2(n\_rounds)$ 个哈希桶,采用下式分别计算Q、K中每一个特征向量所落入的哈希桶;

$$[0027] \quad h(q_u) = a^{(1)}a^{(2)} \dots a^{(n\_rounds)}, u = 1, 2, \dots, length$$

$$[0028] \quad a^{(j)} = \begin{cases} 1, & q_u^T h^{(j)} > 0 \\ 0, & q_u^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

$$[0029] \quad h(k_v) = b^{(1)}b^{(2)} \dots b^{(n\_rounds)}, v = 1, 2, \dots, length$$

$$[0030] \quad b^{(j)} = \begin{cases} 1, & k_v^T h^{(j)} > 0 \\ 0, & k_v^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

[0031] 式中, $q_u$ 、 $k_v$ 表示Q、K中第*u*、*v*个特征向量,length表示增强微表情图像序列的帧数, $h(q_u)$ 、 $h(k_v)$ 表示 $q_u$ 、 $k_v$ 所落入的哈希桶的二进制标号, $a^{(j)}$ 、 $b^{(j)}$ 表示 $h(q_u)$ 、 $h(k_v)$ 的第*j*个特值;

[0032] 特征稀疏单元,用于只保留Q、K中落入同一哈希桶的特征向量的索引号,存入稀疏后的特征索引集合  $\Omega_{u,v}$  中:

[0033]  $\Omega_{u,v} = \{(u,v) | h(q_u) = h(k_v)\}, u, v = 1, 2, \dots, \text{length}$

[0034] 稀疏注意力矩阵计算单元,用于根据特征索引集合  $\Omega_{i,j}$  计算稀疏注意力矩阵 Attention(Q,K,V):

$$[0035] \text{Attention}(Q, K, V) = \sum_{(u,v) \in \Omega_{u,v}} \text{softmax}\left(\frac{q_u k_v^T}{\sqrt{d_l}}\right) V$$

[0036] 式中,  $d_l$  为Q、K、V的维度, Attention(Q,K,V) 即为运动变化特征。

[0037] 进一步的,步骤(5)中训练时采用的损失函数为:

$$[0038] L = - \sum_{c=1}^C y_c \log(p_c)$$

[0039] 式中,C为类别数,  $c = 1, 2, \dots, C$ ,  $y_c$  为微表情的类别标签,  $p_c$  为属于类别c的概率。

[0040] 本发明所述的基于稀疏化自注意力机制的微表情识别系统包括:

[0041] 微表情数据库,包含有若干微表情视频和对应的微表情类别标签;

[0042] 预处理模块,用于将数据库中的每个微表情视频转换为一个微表情图像序列,并对微表情图像序列进行预处理;

[0043] 强度放大模块,用于对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列;

[0044] 微表情识别网络建立模块,用于建立微表情识别网络,所述微表情识别网络包括 resnet-18、稀疏化自注意力模块、特征拼接层、第一全连接层、第二全连接层和softmax层,其中,所述resnet-18用于从增强微表情图像序列中提取空间特征,所述稀疏化自注意力模块用于从空间特征中获取运动变化特征,所述特征拼接层用于将空间特征和运动变化特征拼接为综合特征后输入至第一全连接层,第一全连接层和第二全连接层用于依次将综合特征降维,所述softmax层用于根据降维后的综合特征进行识别分类,得到微表情类别;

[0045] 网络训练模块,用于将每个增强微表情图像序列和对应微表情类别标签作为一个样本,输入微表情识别网络进行训练,得到训练好的微表情识别网络;

[0046] 微表情识别模块,用于对于待识别的微表情视频,依次预处理和强度放大,并输入训练好的微表情识别网络,得到对应的微表情类别。

[0047] 进一步的,所述预处理模块具体用于将数据库中的每个微表情视频转换为一个微表情图像序列,并进行人脸配准和面部区域剪切。

[0048] 进一步的,所述强度放大模块具体用于提取预处理后的微表情图像序列的起始帧  $I_{\text{onset}}$  和峰值帧  $I_{\text{apex}}$ ,并根据起始帧  $I_{\text{onset}}$  和峰值帧  $I_{\text{apex}}$  使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列  $I_{\text{mag}}$ :

$$[0049] I_{\text{mag}} = \{I_{\text{mag},i}\}$$

$$[0050] I_{\text{mag},i} = f(I_{\text{onset}} + (1 + \alpha_i) | I_{\text{apex}} - I_{\text{onset}} |)$$

[0051]  $I_{\text{mag},i}$  为预处理后微的表情图像序列中第i帧图像放大之后的图像,  $f(\cdot)$  为运动放大网络,  $e_i$  为第i帧图像的放大倍数,放大倍数按照微表情图像序列的排列顺序从前到后依次递增。



[0052] 进一步的,所述稀疏化自注意力模块具体包括:

[0053] 基本矩阵生成单元,用于根据空间特征 $F^T$ 采用下式计算自注意力矩阵的查询矩阵Q、键矩阵K、值矩阵V:

$$[0054] \quad Q = F^T \times W^Q$$

$$[0055] \quad K = F^T \times W^K$$

$$[0056] \quad V = F^T \times W^V$$

[0057] 式中, $W^Q$ 、 $W^K$ 、 $W^V$ 分别为对应的权重矩阵;

[0058] 超平面生成单元,用于在空间中随机地生成 $n\_rounds$ 个超平面 $\{h^{(1)}, h^{(2)}, \dots, h^{(n\_rounds)}\}$ , $n\_rounds$ 为大于2的正整数; $h^{(*)}$ 为第 $*$ 个超平面;

[0059] 特征哈希单元,用于基于超平面将空间分为 $2^{(n\_rounds)}$ 个哈希桶,采用下式分别计算Q、K中每一个特征向量所落入的哈希桶:

$$[0060] \quad h(q_u) = a^{(1)} a^{(2)} \dots a^{(n\_rounds)}, u = 1, 2, \dots, length$$

$$[0061] \quad a^{(j)} = \begin{cases} 1, & q_u^T h^{(j)} > 0 \\ 0, & q_u^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

$$[0062] \quad h(k_v) = b^{(1)} b^{(2)} \dots b^{(n\_rounds)}, v = 1, 2, \dots, length$$

$$[0063] \quad b^{(j)} = \begin{cases} 1, & k_v^T h^{(j)} > 0 \\ 0, & k_v^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

[0064] 式中, $q_u$ 、 $k_v$ 表示Q、K中第 $u$ 、 $v$ 个特征向量, $length$ 表示增强微表情图像序列的帧数, $h(q_u)$ 、 $h(k_v)$ 表示 $q_u$ 、 $k_v$ 所落入的哈希桶的二进制标号, $a^{(j)}$ 、 $b^{(j)}$ 表示 $h(q_u)$ 、 $h(k_v)$ 的第 $j$ 个比特值;

[0065] 特征稀疏单元,用于只保留Q、K中落入同一哈希桶的特征向量的索引号,存入稀疏后的特征索引集合 $\Omega_{u,v}$ 中:

$$[0066] \quad \Omega_{u,v} = \{(u, v) \mid h(q_u) = h(k_v)\}, u, v = 1, 2, \dots, length$$

[0067] 稀疏注意力矩阵计算单元,用于根据特征索引集合 $\Omega_{i,j}$ 计算稀疏注意力矩阵Attention(Q,K,V):

$$[0068] \quad \text{Attention}(Q, K, V) = \sum_{(u,v) \in \Omega_{u,v}} \text{softmax}\left(\frac{q_u k_v^T}{\sqrt{d_l}}\right) V$$

[0069] 式中, $d_l$ 为Q、K、V的维度,Attention(Q,K,V)即为运动变化特征。

[0070] 进一步的,所述网络训练模块中训练时采用的损失函数为:

$$[0071] \quad L = - \sum_{c=1}^C y_c \log(p_c)$$

[0072] 式中, $C$ 为类别数, $c=1, 2, \dots, C$ , $y_c$ 为微表情的类别标签, $p_c$ 为属于类别 $c$ 的概率。

[0073] 有益效果:本发明与现有技术相比,其显著优点是:本发明识别准确率更高,且减少人为设定的超参数,更方便。

## 附图说明

[0074] 图1是本发明提供的基于稀疏化自注意力机制的微表情识别方法的一个实施例的

流程示意图；

[0075] 图2是微表情识别示意图；

[0076] 图3是注意力矩阵稀疏化过程的示意图；

[0077] 图4是一个局部敏感哈希算法划分超平面和分配哈希桶的示意图。

### 具体实施方式

[0078] 本实施例提供了一种基于稀疏化自注意力机制的微表情识别方法,如图1所示,包括如下步骤:

[0079] (1) 获取微表情数据库,所述微表情数据库中包含有若干微表情视频和对应的微表情类别标签。

[0080] 其中,数据库可以采用CASME II,SAMM,SMIC-HS这三个常用数据库。

[0081] (2) 将数据库中的每个微表情视频转换为一个微表情图像序列,并对微表情图像序列进行预处理。

[0082] 其中,所述预处理包括人脸配准和面部区域剪切。

[0083] (3) 对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列。

[0084] 该步骤提取预处理后的微表情图像序列的起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ ,并根据起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ 使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列 $I_{mag}$ :

[0085]  $I_{mag} = \{I_{mag,i}\}$

[0086]  $I_{mag,i} = f(I_{onset} + (1 + \alpha_i) |I_{apex} - I_{onset}|)$

[0087]  $I_{mag,i}$ 为预处理后微的表情图像序列中第 $i$ 帧图像放大之后的图像, $f(\cdot)$ 为运动放大网络, $\alpha_i$ 为第 $i$ 帧图像的放大倍数,放大倍数按照微表情图像序列的排列顺序从前到后依次递增,这样得到的运动变化会更加明显,比如 $\alpha_1 = 1, \alpha_2 = 1.375, \alpha_3 = 1.75, \dots, \alpha_{32} = 13$ ,可以得到32张图像的增强微表情图像序列。

[0088] (4) 建立微表情识别网络,如图2所示,所述微表情识别网络包括resnet-18、稀疏化自注意力模块、特征拼接层、第一全连接层、第二全连接层和softmax层,其中,所述resnet-18用于从增强微表情图像序列中提取空间特征,所述稀疏化自注意力模块用于从空间特征中获取运动变化特征,所述特征拼接层用于将空间特征和运动变化特征拼接为综合特征后输入至第一全连接层,第一全连接层和第二全连接层用于依次将综合特征降维,所述softmax层用于根据降维后的综合特征进行识别分类,得到微表情类别。

[0089] 原始的自注意力机制中主要由三个特征矩阵起作用:Q用来匹配其他的信息,K作为被匹配的键值,V用来计算最终获取的信息,注意力矩阵的获得可以表示如下:

[0090] 
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

[0091] 其中 $\sqrt{d_k}$ 为Q,K,V三个特征向量的维度。

[0092] 但是本发明中由于个体差异所带来的放大程度无法统一的问题,对所有序列使用一致的放大程度不可避免会出现面部图像严重变形的情况,为了解决这一问题,本发明摒

弃传统的注意力机制算法,在注意力矩阵计算时引入稀疏化限制,采用稀疏化自注意力模块提取运动变化特征,将严重的变形视为噪声信息,如图3所示。稀疏化自注意力模块从空间特征 $F^T$ 中获取运动变化特征,具体包括:

[0093] 基本矩阵生成单元,用于根据空间特征 $F^T$ 采用下式计算自注意力矩阵的查询矩阵Q、键矩阵K、值矩阵V:

$$[0094] \quad Q = F^T \times W^Q$$

$$[0095] \quad K = F^T \times W^K$$

$$[0096] \quad V = F^T \times W^V$$

[0097] 式中, $W^Q$ 、 $W^K$ 、 $W^V$ 分别为对应的权重矩阵;

[0098] 超平面生成单元,用于在空间中随机地生成 $n\_rounds$ 个超平面 $\{h^{(1)}, h^{(2)}, \dots, h^{(n\_rounds)}\}$ , $n\_rounds$ 为大于2的正整数; $h^{(*)}$ 为第\*个超平面;例如,当 $n\_rounds = 3$ ,用python中自带的随机数生成函数,如`np.random.randn(3,1024)`表示生成三个大小为 $1 \times 1024$ 的向量;

[0099] 特征哈希单元,用于基于超平面将空间分为 $2^{(n\_rounds)}$ 个哈希桶,如图4所示,采用下式分别计算Q、K中每一个特征向量所落入的哈希桶;

$$[0100] \quad h(q_u) = a^{(1)} a^{(2)} \dots a^{(n\_rounds)}, u = 1, 2, \dots, length$$

$$[0101] \quad a^{(j)} = \begin{cases} 1, & q_u^T h^{(j)} > 0 \\ 0, & q_u^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

$$[0102] \quad h(k_v) = b^{(1)} b^{(2)} \dots b^{(n\_rounds)}, v = 1, 2, \dots, length$$

$$[0103] \quad b^{(j)} = \begin{cases} 1, & k_v^T h^{(j)} > 0 \\ 0, & k_v^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

[0104] 式中, $q_u$ 、 $k_v$ 表示Q、K中第 $u$ 、 $v$ 个特征向量,大小为 $1 \times 1024$ , $length$ 表示增强微表情图像序列的帧数, $h(q_u)$ 、 $h(k_v)$ 表示 $q_u$ 、 $k_v$ 所落入的哈希桶的二进制标号, $a^{(j)}$ 、 $b^{(j)}$ 表示 $h(q_u)$ 、 $h(k_v)$ 的第 $j$ 个比特值;例如,当 $n\_rounds = 3$ ,共有8个哈希桶,如果 $q_u^T h_1 > 0$ , $q_u^T h_2 < 0$ , $q_u^T h_3 > 0$ ,则 $h(q_u)$ 可以记为101;

[0105] 特征稀疏单元,用于只保留Q、K中落入同一哈希桶的特征向量的索引号,存入稀疏后的特征索引集合 $\Omega_{u,v}$ 中:

$$[0106] \quad \Omega_{u,v} = \{(u,v) \mid h(q_u) = h(k_v)\}, u, v = 1, 2, \dots, length$$

[0107] 稀疏后的特征索引集合是为了形成稀疏化的注意力矩阵 $QK_r$ ,只计算落入同一个哈希桶中的计算结果,根据局部敏感哈希算法的特性,在这一过程中,由于放大产生的噪声与其他携带运动信息的特征向量之间的相似度很低,在空间上可以表示为两个特征向量间的距离会比较远,因此,噪声的特征向量很可能被哈希到离散的哈希桶中,如果该哈希桶中没有与之对应的向量,在最终的注意力矩阵上会表现为零值。所以,最终保留落入同一哈希桶的特征向量的索引号;

[0108] 稀疏注意力矩阵计算单元,用于根据特征索引集合 $\Omega_{i,j}$ 计算稀疏注意力矩阵 $Attention(Q,K,V)$ :

$$[0109] \quad Attention(Q,K,V) = \sum_{(u,v) \in \Omega_{u,v}} softmax\left(\frac{q_u k_v^T}{\sqrt{d_l}}\right) V$$

[0110] 式中, $d_l$ 为Q、K、V的维度, $Attention(Q,K,V)$ 即为运动变化特征。

[0111] 得到运动变化特征后,特征拼接层将空间特征和运动变化特征拼接为综合特征,记直接在空间特征后加上运动变化特征,之后输入至第一全连接层,第一全连接层和第二全连接层先后对综合特征降维,之后softmax层根据降维后的综合特征进行识别分类,得到微表情类别。

[0112] (5) 将每个增强微表情图像序列和对应微表情类别标签作为一个样本,输入微表情识别网络进行训练,得到训练好的微表情识别网络。

[0113] 训练时采用的损失函数为:

$$[0114] \quad L = - \sum_{c=1}^C y_c \log(p_c)$$

[0115] 式中,C为类别数, $c=1,2,\dots,C$ , $y_c$ 为微表情的类别标签, $p_c$ 为属于类别c的概率。

[0116] (6) 对于待识别的微表情视频,依次按照步骤(2)、(3)处理,并输入训练好的微表情识别网络,得到对应的微表情类别。

[0117] 本实施例还提供了一种基于稀疏化自注意力机制的微表情识别系统,包括:

[0118] 微表情数据库,包含有若干微表情视频和对应的微表情类别标签;

[0119] 预处理模块,用于将数据库中的每个微表情视频转换为一个微表情图像序列,并对微表情图像序列进行预处理;

[0120] 强度放大模块,用于对预处理后的微表情图像序列使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列;

[0121] 微表情识别网络建立模块,用于建立微表情识别网络,所述微表情识别网络包括resnet-18、稀疏化自注意力模块、特征拼接层、第一全连接层、第二全连接层和softmax层,其中,所述resnet-18用于从增强微表情图像序列中提取空间特征,所述稀疏化自注意力模块用于从空间特征中获取运动变化特征,所述特征拼接层用于将空间特征和运动变化特征拼接为综合特征后输入至第一全连接层,第一全连接层和第二全连接层用于依次将综合特征降维,所述softmax层用于根据降维后的综合特征进行识别分类,得到微表情类别;

[0122] 网络训练模块,用于将每个增强微表情图像序列和对应微表情类别标签作为一个样本,输入微表情识别网络进行训练,得到训练好的微表情识别网络;

[0123] 微表情识别模块,用于对于待识别的微表情视频,依次预处理和强度放大,并输入训练好的微表情识别网络,得到对应的微表情类别。

[0124] 进一步的,所述预处理模块具体用于将数据库中的每个微表情视频转换为一个微表情图像序列,并进行人脸配准和面部区域剪切。

[0125] 进一步的,所述强度放大模块具体用于提取预处理后的微表情图像序列的起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ ,并根据起始帧 $I_{onset}$ 和峰值帧 $I_{apex}$ 使用运动放大网络进行微表情强度的放大,得到增强微表情图像序列 $I_{mag}$ :

$$[0126] \quad I_{mag} = \{I_{mag,i}\}$$

$$[0127] \quad I_{mag,i} = f(I_{onset} + (1+\alpha_i) | I_{apex} - I_{onset} |)$$

[0128]  $I_{mag,i}$ 为预处理后微的表情图像序列中第i帧图像放大之后的图像, $f(\cdot)$ 为运动放大网络, $\alpha_i$ 为第i帧图像的放大倍数,放大倍数按照微表情图像序列的排列顺序从前到后依次递增。

[0129] 进一步的,所述稀疏化自注意力模块具体包括:

[0130] 基本矩阵生成单元,用于根据空间特征 $F^T$ 采用下式计算自注意力矩阵的查询矩阵Q、键矩阵K、值矩阵V:

$$[0131] \quad Q = F^T \times W^Q$$

$$[0132] \quad K = F^T \times W^K$$

$$[0133] \quad V = F^T \times W^V$$

[0134] 式中, $W^Q$ 、 $W^K$ 、 $W^V$ 分别为对应的权重矩阵;

[0135] 超平面生成单元,用于在空间中随机地生成 $n\_rounds$ 个超平面 $\{h^{(1)}, h^{(2)}, \dots, h^{(n\_rounds)}\}$ , $n\_rounds$ 为大于2的正整数; $h^{(*)}$ 为第 $*$ 个超平面;

[0136] 特征哈希单元,用于基于超平面将空间分为 $2^{(n\_rounds)}$ 个哈希桶,采用下式分别计算Q、K中每一个特征向量所落入的哈希桶:

$$[0137] \quad h(q_u) = a^{(1)} a^{(2)} \dots a^{(n\_rounds)}, u = 1, 2, \dots, length$$

$$[0138] \quad a^{(j)} = \begin{cases} 1, & q_u^T h^{(j)} > 0 \\ 0, & q_u^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

$$[0139] \quad h(k_v) = b^{(1)} b^{(2)} \dots b^{(n\_rounds)}, v = 1, 2, \dots, length$$

$$[0140] \quad b^{(j)} = \begin{cases} 1, & k_v^T h^{(j)} > 0 \\ 0, & k_v^T h^{(j)} < 0 \end{cases}, j = 1, 2, \dots, n\_rounds$$

[0141] 式中, $q_u$ 、 $k_v$ 表示Q、K中第 $u$ 、 $v$ 个特征向量, $length$ 表示增强微表情图像序列的帧数, $h(q_u)$ 、 $h(k_v)$ 表示 $q_u$ 、 $k_v$ 所落入的哈希桶的二进制标号, $a^{(j)}$ 、 $b^{(j)}$ 表示 $h(q_u)$ 、 $h(k_v)$ 的第 $j$ 个比特值;

[0142] 特征稀疏单元,用于只保留Q、K中落入同一哈希桶的特征向量的索引号,存入稀疏后的特征索引集合 $\Omega_{u,v}$ 中:

$$[0143] \quad \Omega_{u,v} = \{(u, v) \mid h(q_u) = h(k_v)\}, u, v = 1, 2, \dots, length$$

[0144] 稀疏注意力矩阵计算单元,用于根据特征索引集合 $\Omega_{i,j}$ 计算稀疏注意力矩阵Attention(Q,K,V):

$$[0145] \quad \text{Attention}(Q, K, V) = \sum_{(u,v) \in \Omega_{u,v}} \text{softmax}\left(\frac{q_u k_v^T}{\sqrt{d_l}}\right) V$$

[0146] 式中, $d_l$ 为Q、K、V的维度,Attention(Q,K,V)即为运动变化特征。

[0147] 进一步的,所述网络训练模块中训练时采用的损失函数为:

$$[0148] \quad L = - \sum_{c=1}^C y_c \log(p_c)$$

[0149] 式中, $C$ 为类别数, $c=1, 2, \dots, C$ , $y_c$ 为微表情的类别标签, $p_c$ 为属于类别 $c$ 的概率。

[0150] 为验证本发明的有效性,在CAME2微表情数据库、SAMM微表情数据库和SMIC数据库的HS子数据库之间做了微表情识别,验证结果及与其他最新方法的比较如表1所示:

[0151] 表1

[0152]

方法	年份	CASME2	SAMM	SMIC-HS
LBP-SIP	2014	45.36	36.76	42.12

MagGA	2018	63.30	N/A	N/A
DSSN	2019	70.78	57.35	63.41
TSCNN-I	2020	74.05	63.53	72.74
LBPA <sub>CC</sub> P <sup>u2</sup>	2021	69.03	N/A	76.59
本方法	2021	81.06	77.61	79.88

[0153] 表1中,N/A表示没有相关记录。

[0154] 对CASME2数据库的表情做如下处理:将样本个数少于10个的类别省略,避免样本严重不平衡问题。完成5种类别的识别任务,分别是happy、regression、disgust、fear和surprise对SAMM数据库的表情做如下处理:将样本个数少于10个的类别省略,避免样本严重不平衡问题,完成5种类别的识别任务,分别是、happiness、angry、disgust、fear和surprise。SMIC数据库本身类别为positive、negative、surprise,完成三分类的识别任务。

[0155] 实验结果表明,基于本发明提出的微表情识别方法取得了较高的微表情识别率。相较于传统的微表情放大方式,本发明可以免去部分超参数人为设置的繁琐,对个体的适应性更强,更方便。

[0156] 以上所揭露的仅为本发明一种较佳实施例而已,不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

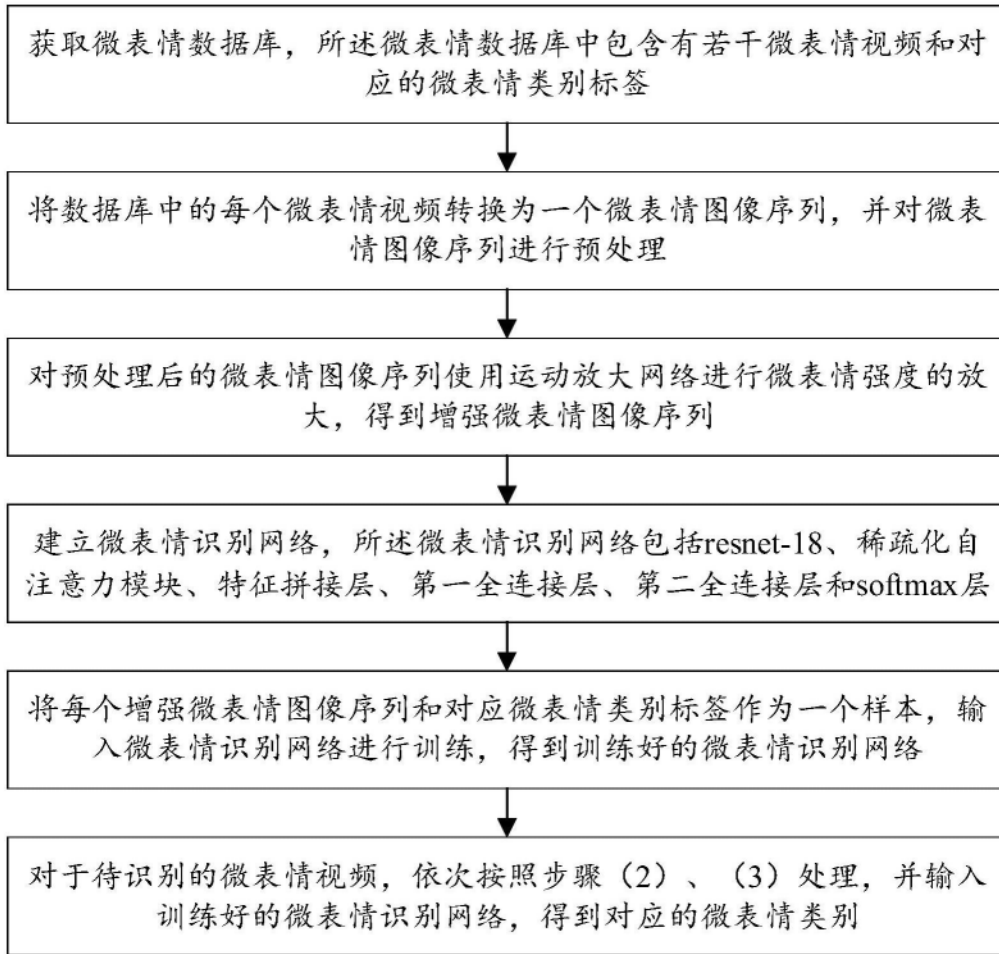


图1

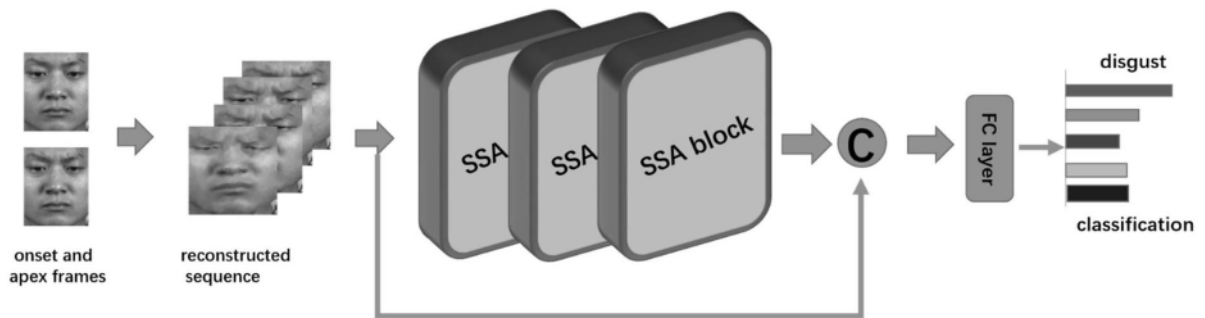


图2

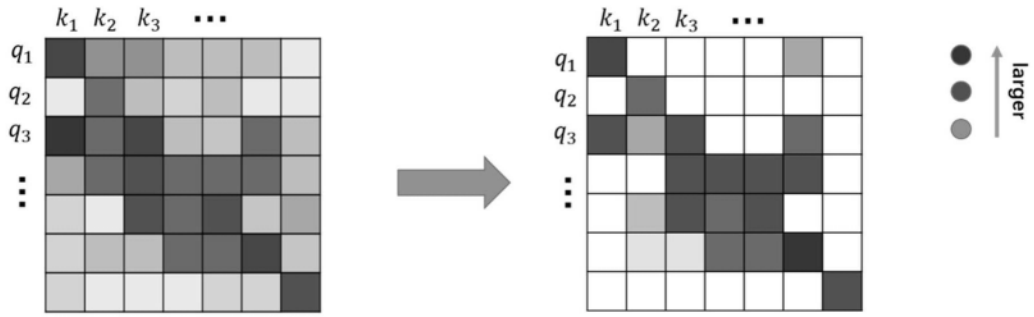


图3

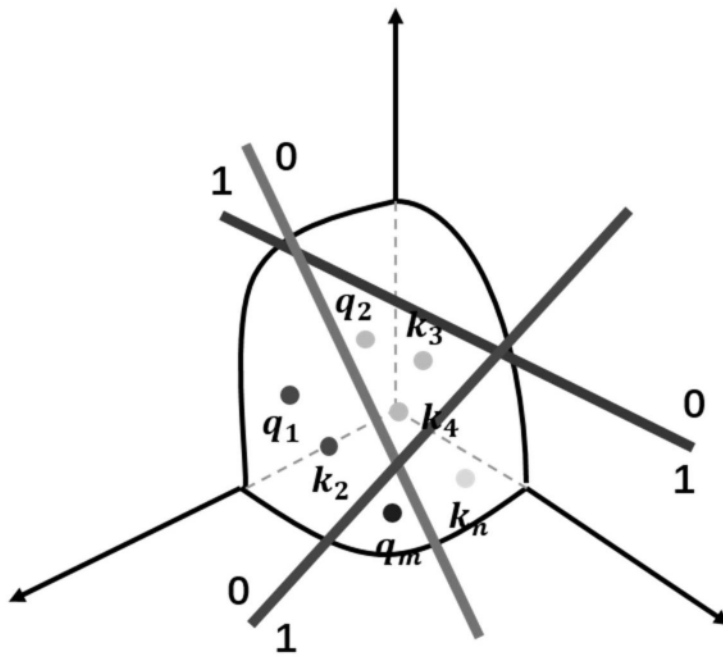


图4