

doi:10.14132/j.cnki.1673-5439.2019.05.007

基于遗传算法的多模态情感特征融合方法

卢官明¹,程晓¹,李霞¹,闫静杰¹,李海波^{1,2}

(1. 南京邮电大学 通信与信息工程学院, 江苏 南京 210003)
(2. 瑞典皇家理工学院, 瑞典 斯德哥尔摩 SE-10044)

摘要: 为了提高情感识别的准确率, 针对单模态情感识别率低以及常规特征融合方法存在的缺点, 提出了一种基于遗传算法的多模态情感特征融合方法, 利用遗传算法对多个模态的情感特征进行选择、交叉以及重组。在 eNTRAFACE'05 表情-语音双模态情感数据库上进行了情感识别测试, 对基于面部表情或语音的单模态情感识别, 以及基于特征层或决策层融合的各种双模态情感识别的性能进行了比较。实验结果表明, 双模态情感识别的性能高于单模态情感识别, 而且文中提出的基于遗传算法的多模态情感特征融合方法比其他几种常规的特征融合方法的效果好, 验证了文中提出的方法的可行性和有效性。

关键词: 多模态情感识别; 特征融合; 特征选择; 遗传算法

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-5439(2019)05-0041-07

Multi-modal emotion feature fusion method based on genetic algorithm

LU Guanming¹, CHENG Xiao¹, LI Xia¹, YAN Jingjie¹, LI Haibo^{1,2}

(1. College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)
(2. KTH Royal Institute of Technology, Stockholm SE-10044, Sweden)

Abstract: To improve the accuracy of the emotion recognition, aiming at the low accuracy of the single-modal emotion recognition and the shortcomings of conventional feature fusion methods, this paper proposes a multi-modal emotion feature fusion method based on a genetic algorithm. The features from multiple emotion modalities are selected, crossed and recombined by using the genetic algorithm. The emotion recognition tests are carried out in eNTRAFACE'05 audio-visual emotion database. The performance of the single-modal emotion recognition based on the facial expression or speech and various bimodal emotion recognitions based on the feature level or the decision level fusion are compared. Experimental results show that the performance of the bimodal emotion recognition is better than that of the single-modal emotion recognition, and the proposed method based on genetic algorithm is superior to other conventional feature fusion methods, thus demonstrating the feasibility and the effectiveness of the proposed method.

Keywords: multi-modal emotion recognition; feature fusion; feature selection; genetic algorithm

收稿日期: 2019-07-25 本刊网址: <http://nyzr.njupt.edu.cn>

基金项目: 国家自然科学基金(61501249)、江苏省重点研发计划(BE2016775)和江苏省研究生科研与实践创新计划(KYLX15_0827, KYLX16_0660, KYCX17_0787)资助项目

作者简介: 卢官明, 男, 博士, 教授, 博士生导师, lugm@njupt.edu.cn

引用本文: 卢官明, 程晓, 李霞, 等. 基于遗传算法的多模态情感特征融合方法[J]. 南京邮电大学学报(自然科学版), 2019, 39(5): 41-47.

情感在人类日常生活的互动中起着重要作用。人类面部表情的变化,说话的声调、节奏和语速的变化以及肢体动作等这些外部表现都传达出不同的情感状态。随着人工智能技术的发展,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。情感计算已迅速发展成为一个跨学科的研究领域^[1]。作为情感计算研究领域的一个重要分支,情感识别在机器人^[2]、健康监测^[3]等人机交互领域具有广阔的应用前景,已受到研究人员的广泛关注。

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,所以,单模态情感识别存在一定的局限性,无法满足人们的期望。而在实际生活中,人们往往是综合语音、面部表情、肢体动作等多种信息来判断一个人的情感状态,这样的判断结果才更加准确。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率^[4-8]。

本文提出一种基于视听数据的多模态情感识别方法。首先提取面部表情以及语音等各个模态的情感特征;然后,利用遗传算法对多个模态的情感特征进行特征融合,包括基于遗传算法的特征选择、交叉以及重组;最后,使用基于支持向量机(SVM)的分类器进行情感分类。

1 相关工作

“多模态”(Multi-Modal)这一概念最早由 Bigün 等于 1997 年提出,他们基于贝叶斯统计理论使用人体多种生理特征或行为特征进行身份和行为的分析、识别。2007 年, Gunes 等^[9]提出了基于面部表情和身体姿态的双模态情感识别方法。2012 年, Wang 等^[10]提出了一种基于核跨模态因子分析(Kernel Cross-Modal Factor Analysis)的双模态情感识别方法,在用 eNTERFACE'05 和 RML 语音-表情

双模态情感数据库进行实验时,识别率分别达到 72.47% 和 82.22%。2013 年, 闫静杰等^[11]研究了基于视频的表情和姿态双模态情感识别,提出了一种基于双边稀疏偏最小二乘(BSPLS)的情感识别方法,首先从视频中分别提取表情和姿态两种模态的空时特征,然后使用基于 BSPLS 的方法对特征降维后的两组单模态的特征向量进行特征选择并融合,生成新的情感特征向量,最后分别采用 K-近邻(KNN)分类器和基于支持向量机(SVM)的分类器进行情感分类识别,在 FABO 数据库上取得了 65.22% 和 76.40% 的平均识别率。2018 年, 闫静杰等^[12]针对语音和面部表情的双模态情感识别,提出了一种基于稀疏典型相关分析的特征融合方法,在 eNTRAFACE'05 数据库上取得了 60.09% 的平均识别率。Tzirakis 等^[6]提出了一种基于深度神经网络的双模态情感识别算法,其中语音和面部表情模态分别使用 CNN 和 50 层的 ResNet 网络提取情感特征,然后对双模态特征进行直接级联,最后使用两层 LSTM 模型进行时序建模,然后以端到端的方式训练系统。上述研究都表明,与基于语音或面部表情的单模态情感识别方法相比,多模态情感识别方法能够取得更高的识别率。

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。目前常用的决策层融合方法包括线性加权法^[13]、Dempster-Shafer 证据理论(D-S 证据理论)^[14-15]和人工神经网络(Artificial Neural Network, ANN)法等^[16]。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。常用的特征层融合方法包括典型相关分析(Canonical Correlation Analysis, CCA)^[17]、判别典型相关分析(Discriminative Canonical Correlation Analysis, DCCA)^[18]、核典型相关分析(Kernel Canonical Correlation Analysis, KCCA)^[19]、基于核矩阵的融合(Kernel Matrix Fusion, KMF)^[10]、多核学习(Multiple Kernel Learning, MKL)^[20]、遗传

算法^[21]等。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,是进行多特征融合的关键,目前仍然是面临挑战的开放课题。为了对不同模态的情感特征进行合理的融合,本文提出一种基于遗传算法的多模态情感特征融合方法,以实现不同情感特征在多模态情感识别系统中的信息互补。

2 多模态情感特征提取

2.1 面部表情特征提取

由于 Gabor 小波具有多尺度和多方向的特点,

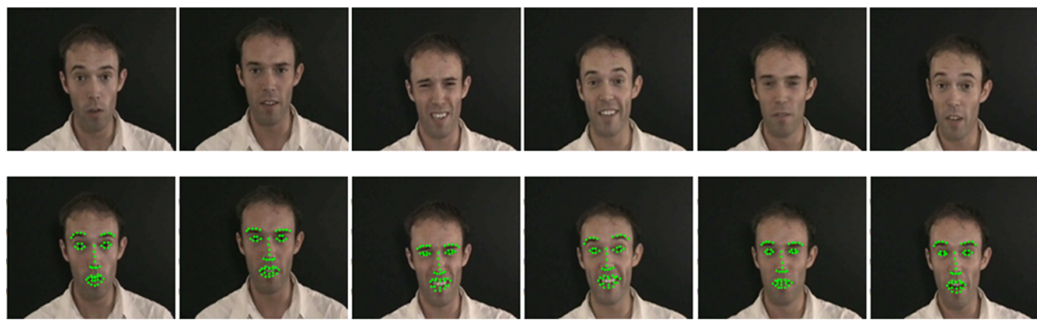


图1 eNTERFACE'05 数据库中某一对象的6种情感样本图像及标定的关键点

2.2 语音情感特征提取

目前,在语音情感识别领域,常用的语音情感特征包括韵律特征、声谱特征、音质特征^[22]。本文使用 OpenSMILE 工具包^[23]提取“emobase2010”特征集内的语音特征,包括基频(Pitch)、帧能量(Frame Energy)、过零率(Zero-Crossing Rate)、梅尔频率倒谱系数(Mel Frequency Cepstral Coefficients, MFCC)、频率微扰和振幅微扰(Jitter and Shimmer)等低阶描述子(Low-Level Descriptors, LLD),然后通过常用的统计量,例如均值、方差、标准差、偏度(Skewness)、峰度(Kurtosis)、回归系数等将低阶描述子转化为一定维数的特征向量,构成一个1582维的语音情感特征向量。

3 基于遗传算法的特征融合

遗传算法(Genetic Algorithm, GA)是一种基于自然选择、基因遗传以及优胜劣汰的生物种群进化思想进行问题求解的启发式优化算法。尽管后来提出了多种改进的遗传算法,但其基本原理还是没有改变。对遗传编码、适应度函数及遗传算子的设计和选择是实施遗传算法的重要环节。遗传编码的目的是将遗传算法中的染色体与实际问题建立联系,每一个染色体实质上都是问题的一个可能解。适应

Gabor 滤波器是具有方向选择和频率选择特性的带通滤波器,在空间和频率域具有最佳的联合分辨率,所以,本文采用 Gabor 滤波器来提取人脸面部表情特征。

在提取人脸面部表情特征时,为了减少后期处理的数据量,首先使用 IntraFace 软件对输入的人脸图像标定49个关键特征(Landmark)点,如图1所示;然后选择5个尺度、8个方向($0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ, 157.5^\circ$)的40个 Gabor 滤波器来提取每个关键特征点的表情特征,得到一个1960维的面部表情特征向量。

度即染色体个体对于生存环境的适应程度,适应度的值越高,表示该染色体个体的生存概率越大,有更高的可能性被保留至下一代染色体种群。初始种群,也称父代,父代经过一个操作循环得到第一代子代种群,这个操作循环由选择、交叉、变异3个算子组成。选择算子是根据适应度函数的值,保留染色体种群中效果较好的个体,进行下一代种群的繁衍操作,这就相当于多模态情感的特征选择,即选择了区分度高的特征用于分类识别;交叉操作模拟了基因重组的过程,在保留上一代种群优良特性的同时,获得新的个体;变异算子模拟种群进化过程中的基因突变,其在保留染色体个体大部分基因的同时对个别的编码进行局部调整,目的是为了保持种群的多样性,以提高遗传算法的局部搜索能力。

本文提出的基于遗传算法的特征融合方法的具体步骤如下。

(1) 对于多模态情感数据库(假设数据库包含 L 类情感的样本,每类情感的样本数为 n ,总样本数为 $N = nL$)中的每个样本,提取 T 种不同模态的情感特征,如语音特征、表情特征、姿态特征等,其中第 t 种模态的情感特征用 d_t 维特征向量表示, $t = 1, 2, \dots, T$,则每个样本的多模态情感特征向量维数为 $M = (d_1 + d_2 + \dots + d_T)$;对于数据库中的 N 个样本,

构造一个 $N * M$ 大小的多模态情感特征矩阵为

$$\mathbf{A} = \begin{bmatrix} a_{1,1,1} & \cdots & a_{1,1,k} & \cdots & a_{1,1,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1,j,1} & \cdots & a_{1,j,k} & \cdots & a_{1,j,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{1,n,1} & \cdots & a_{1,n,k} & \cdots & a_{1,n,M} \\ \vdots & & \vdots & & \vdots \\ a_{i,1,1} & \cdots & a_{i,1,k} & \cdots & a_{i,1,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,j,1} & \cdots & a_{i,j,k} & \cdots & a_{i,j,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,n,1} & \cdots & a_{i,n,k} & \cdots & a_{i,n,M} \\ \vdots & & \vdots & & \vdots \\ a_{L,1,1} & \cdots & a_{L,1,k} & \cdots & a_{L,1,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{L,j,1} & \cdots & a_{L,j,k} & \cdots & a_{L,j,M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{L,n,1} & \cdots & a_{L,n,k} & \cdots & a_{L,n,M} \end{bmatrix} \quad (1)$$

其中,矩阵元素 $a_{i,j,k}$ 是属于第 i 类情感的第 j 个样本特征向量的第 k 维特征值, $i=1,2,\dots,L, j=1,2,\dots,n, k=1,2,\dots,M$; 设置迭代次数 F 。

(2) 对于第 k 维特征,将属于第 i 类情感的 n

$$\text{个样本的特征值进行组合,构成数组 } \mathbf{A}_{i,k} = \begin{pmatrix} a_{i,1,k} \\ a_{i,2,k} \\ \vdots \\ a_{i,n,k} \end{pmatrix},$$

计算这个数组的均值 $Mean_{\mathbf{A}_{i,k}} = \frac{1}{n} \sum_{j=1}^n a_{i,j,k}$ 和方差

$$Var_{\mathbf{A}_{i,k}} = \frac{1}{n} \sum_{j=1}^n (a_{i,j,k} - Mean_{\mathbf{A}_{i,k}})^2, \text{ 定义第 } p \text{ 类情感}$$

与第 q 类情感的类内类间距离为

$$R_{p,q,k} = \frac{|Mean_{\mathbf{A}_{p,k}} - Mean_{\mathbf{A}_{q,k}}|}{\sqrt{Var_{\mathbf{A}_{p,k}} + Var_{\mathbf{A}_{q,k}}}} \quad (2)$$

$$(p, q = 1, 2, \dots, L; k = 1, 2, \dots, M)$$

定义适应度函数为

$$R_k = \frac{2}{L(L-1)} \sum_{\substack{p,q=1,2,\dots,L \\ p < q}} R_{p,q,k} \quad (3)$$

那么 M 维特征的适应度函数值可以分别表示为 $R_1, R_2, \dots, R_M$ 。

(3) 定义多模态情感特征矩阵 $\mathbf{A}$ 中第 s 列特征的插入率为

$$\rho_s = \frac{R_s}{\sum_{k=1}^M R_k} \quad (4)$$

给定一个 $[0,1]$ 区间的均匀随机数 α , 若 $\rho_s \geq \alpha$, 则多模态情感特征矩阵 $\mathbf{A}$ 中第 s 列特征被选择保留, 否则不被选择。假定经选择后的多模态情感特征矩阵 $\mathbf{B}$ 是一个 $N * m$ 大小的矩阵, 其中 $m < M$ 。

(4) 使用单点交叉算子对特征矩阵 $\mathbf{B}$ 进行交叉操作, 具体过程如下。

将矩阵 $\mathbf{B}$ 的奇、偶列分离出来, 分别构成矩阵 $\mathbf{B}_{\text{even}}$ 和矩阵 $\mathbf{B}_{\text{odd}}$, 大小分别为 $N * m_1$ 和 $N * m_2$, 其中 m_1 是矩阵 $\mathbf{B}$ 偶数列的列数, m_2 是矩阵 $\mathbf{B}$ 奇数列的列数, m_1 和 m_2 的取值为

$$m_1 = \begin{cases} \frac{m}{2}, & m \text{ 为偶数} \\ \frac{m-1}{2}, & m \text{ 为奇数} \end{cases} \quad (5)$$

$$m_2 = \begin{cases} \frac{m}{2}, & m \text{ 为偶数} \\ \frac{m+1}{2}, & m \text{ 为奇数} \end{cases} \quad (6)$$

产生 m_1 个 $[0,1]$ 区间的随机数 $x_1, x_2, \dots, x_{m_1}$, 将这些随机数依次与事先设定的交叉率 p_1 进行比较, 小于 p_1 则返回逻辑值“1”, 否则返回逻辑值“0”, 这样就得到了一个长度为 m_1 的逻辑向量, 用 $\mathbf{D} = (d_1, d_2, \dots, d_{m_1})$ 表示; 重新产生 m_1 个 $[0,1]$ 区间的随机数 $y_1, y_2, \dots, y_{m_1}$, 通过式 (7) 计算矩阵 $\mathbf{B}_{\text{even}}$ 和矩阵 $\mathbf{B}_{\text{odd}}$ 第 r 列的交叉点位置 $S_r (r=1, 2, \dots, m_1)$ 。

$$S_r = ((N-1) * y_r * d_r + N-1) \% N + 1 \quad (7)$$

其中, $\%$ 表示取余运算, 对向量进行取余运算是指对向量的每个元素均进行取余运算; 依次将矩阵 $\mathbf{B}_{\text{even}}$ 第 r 列的第 S_r 个元素之后的数据与矩阵 $\mathbf{B}_{\text{odd}}$ 第 r 列的第 S_r 个元素之后的数据进行互换, 得到 $\mathbf{B}'_{\text{even}}$ 和 $\mathbf{B}'_{\text{odd}}$, 将矩阵 $\mathbf{B}'_{\text{odd}}$ 的每一列依次插入矩阵 $\mathbf{B}'_{\text{even}}$ 的每两列之间, 即构成交叉后的特征矩阵 $\mathbf{H}$ (当 m 为奇数时, 特征矩阵 $\mathbf{H}$ 的最后一列与特征矩阵 $\mathbf{B}$ 的最后一列相同), 计算交叉后的特征的适应度函数值, m 维特征的适应度函数值可以分别表示为 $R'_1, R'_2, \dots, R'_m$ 。

(5) 将初始的多模态情感特征矩阵 $\mathbf{A}$ 和交叉后的特征矩阵 $\mathbf{H}$ 进行重组, 具体过程如下。

设定替换率 p_2 , 假定用特征矩阵 $\mathbf{H}$ 中的 r 列数据替换特征矩阵 $\mathbf{A}$ 中的 r 列数据, 则 r 的数值为 p_2 与 m 相乘后再取整, 构造数组 $\mathbf{Q}_1 = (R_1, R_2, \dots, R_M)$ 和 $\mathbf{Q}_2 = (R'_1, R'_2, \dots, R'_m)$, 将 $\mathbf{Q}_1$ 和 $\mathbf{Q}_2$ 的元素按照

从大到小的顺序分别重新排序,排序后的位置信息分别用向量 $\psi = (\varphi_1, \varphi_2, \dots, \varphi_M)$ 和向量 $\theta = (\vartheta_1, \vartheta_2, \dots, \vartheta_m)$ 来表示,即将数组 Q_1 中第 φ_k ($k = 1, 2, \dots, M$) 个元素排到第 k 位,将数组 Q_2 中第 ϑ_w ($w = 1, 2, \dots, m$) 个元素排到第 w 位,例如, $a = (3, 5, 9, 7)$, 从大到小排序为 $b = (9, 7, 5, 3)$, 位置信息为 $c = (3, 4, 2, 1)$, 即:将 a 中第 3 个元素排到第 1 位,第 4 个元素排到第 2 位,第 2 个元素排到第 3 位,第 1 个元素排到第 4 位,得到排序后的 b ;用特征矩阵 H 中第 $\vartheta_1, \vartheta_2, \dots, \vartheta_r$ 列的数据依次替换特征矩阵 A 中第 $\varphi_1, \varphi_2, \dots, \varphi_r$ 列的数据,得到重组后的特征矩阵 A' ,大小为 $N * M$;将该特征作为步骤(2)的输入,再进行上述过程,依次迭代,当迭代次数达到事先设定的 F 值时,停止循环,得到最终的融合特征。

4 实验结果与分析

为了验证所提出的基于遗传算法的特征融合方法的有效性,本文选用 eNTRAFACE'05 表情-语音双模态情感数据库^[24] 对不同的融合方法进行对比实验。实验时,使用基于支持向量机(SVM)的分类器,并采用 5-折(5-fold)交叉验证的方法,获得最终的平均识别率。

4.1 eNTRAFACE'05 双模态情感数据库

eNTRAFACE'05 双模态情感数据库包含的样本是从来自 14 个不同国家的 43 个成年人的视频中采集得到,包含了人的 6 种基本情感:愤怒、厌恶、高兴、悲伤、恐惧以及惊讶。数据库中的样本以视频的形式记录,主要通过对表演者(采集对象)讲述与 6 种基本情感相对应的小故事来诱导他们表达相应的情感,对于每一种情感类型,每一个表演者用英语说 5 句台词,每人说一句台词的视频为一个样本,其中语音的采样频率为 48 000 Hz,视频的帧率为 25 fps,帧图像的分辨率为 720 × 576 像素。

从记录每个表演者不同情感的样本中截取等长的 10 个视频片段,每个视频片段的时长为 0.12 s。对于每一个视频片段,采用视频截图软件截取其中一帧图像,作为一个面部表情样本,而视频片段中包含的语音信号作为一个语音样本。本文选取其中 42 人的表情样本和语音样本作为实验对象,其中每人每种情感的 4 个表情样本和 4 个语音样本作为训练样本,每人每种情感的 1 个表情样本和 1 个语音样本作为测试样本。

4.2 单模态情感识别实验

对于基于面部表情的单模态情感识别,首先对

面部表情样本提取一个 1 960 维的特征向量,然后使用主成分分析(PCA)方法进行特征降维,最后使用基于支持向量机(SVM)的分类器进行情感分类识别,得到 74.6% 的平均识别率。对于基于语音的单模态情感识别,首先使用 OpenSMILE 工具包,对语音样本提取 1 582 维的 'emobase2010' 特征向量,然后使用主成分分析(PCA)方法进行特征降维,最后使用基于支持向量机(SVM)的分类器进行情感分类识别,得到 44.5% 的平均识别率。

4.3 双模态情感识别实验

在 4.2 节实验中,我们得到了面部表情、语音样本的特征向量,以及基于面部表情、语音的单模态情感识别结果,这里将进行面部表情和语音双模态情感融合实验,分别使用基于核典型相关分析(KCCA)、核矩阵融合(KMF)、遗传算法(GA)的特征层融合方法以及基于求和(Sum)规则的决策层方法进行融合。

(1) 基于核典型相关分析(KCCA)的特征层融合方法:分别选用高斯核函数、多项式核函数以及线性核函数,对表情、语音两种不同模态的特征进行基于 KCCA 的特征层融合,得到相应的融合特征。当选用高斯核函数,且将融合特征向量的维数降到 85 维时,得到 80.6% 的平均识别率。

(2) 基于核矩阵融合(KMF)的特征层融合方法:选用高斯核函数,将不同模态的特征进行核矩阵加权融合,当表情的权重取 0.75、语音的权重取 0.25 时,得到 79.4% 的平均识别率。

(3) 基于遗传算法(GA)的特征层融合方法:当交叉率取 0.75,插入率取 0.9,迭代次数取 9 时,得到 87.5% 的平均识别率。

(4) 基于求和(Sum)规则的决策层融合方法:根据 4.2 节得到的基于面部表情或语音的单模态情感识别结果,将所得的后验概率采用求和规则进行决策层融合,得到 78.6% 的平均识别率。

4.4 结果分析

基于面部表情或语音的单模态情感识别以及基于特征层或决策层融合的双模态情感识别方法的识别率比较如表 1 所示。

表 1 不同方案下的情感识别率

方案	平均识别率/%
基于面部表情的单模态	74.6
基于语音的单模态	44.5
基于 KCCA 特征层融合的双模态	80.6
基于 KMF 特征层融合的双模态	79.4
基于 GA 特征层融合的双模态	87.5
基于求和规则决策层融合的双模态	78.6

由表 1 的对比实验结果可知,基于面部表情和语音双模态的情感识别率高于基于表情或语音的单模态情感识别率,而本文提出的基于遗传算法(GA)的特征融合方法获得的平均识别率最高,达到 87.5%,从而验证了本文提出的基于遗传算法(GA)的多模态情感特征融合方法的有效性。

5 结束语

由于人类情感的表达是通过多种形式同时进行的,因此单一模态的情感识别具有一定的局限性。本文利用不同模态信息之间的互补性,提出了一种基于遗传算法的多模态情感特征融合方法,利用遗传算法对多个模态的情感特征进行选择、交叉以及重组,在充分利用语音、面部表情等多个模态的互补信息的同时也消除了单模态的不稳定性,从而在一定程度上提高了情感识别的准确率。在 eN-TRAFACE'05 表情-语音双模态情感数据库上进行了情感识别测试,对基于面部表情或语音的单模态情感识别,以及基于特征层或决策层融合的各种双模态情感识别的性能进行了比较。实验结果表明,双模态情感识别的性能高于单模态情感识别,而且本文提出的基于遗传算法的多模态情感特征融合方法比其他几种常规的特征融合方法的效果好,验证了本文提出的方法的可行性和有效性。

参考文献:

- [1] CALVO R A, D' MELLO S. Affect detection: an interdisciplinary review of models, methods, and their applications [J]. IEEE Transactions on Affective Computing, 2010, 1 (1): 18–37.
- [2] MANOHAR V, CRANDALL J W. Programming robots to express emotions: interaction paradigms, communication modalities, and context [J]. IEEE Transactions on Human-Machine Systems, 2014, 44 (3): 362–373.
- [3] HOSSAIN M S, MUHAMMAD G. Cloud-assisted speech and face recognition framework for health monitoring [J]. Mobile Networks and Applications, 2015, 20 (3): 391–399.
- [4] PORIA S, CAMBRIA E, BAJPAI R, et al. A review of affective computing: from unimodal analysis to multimodal fusion [J]. Information Fusion, 2017, 37: 98–125.
- [5] ZHENG W L, LIU W, LU Y F, et al. EmotionMeter: a multimodal framework for recognizing human emotion [J]. IEEE Transactions on Cybernetics, 2019, 49 (3): 1110–1122.
- [6] TZIRAKIS P, TRIGEORGIS G, NICOLAOU M A, et al. End-to-end multimodal emotion recognition using deep neu-

- ral networks [J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11 (8): 1301–1309.
- [7] SENG K, ANG L M, OOI C S. A combined rule-based and machine learning audio-visual emotion recognition approach [J]. IEEE Transactions on Affective Computing, 2018, 9 (1): 3–13.
- [8] YAN J W, ZHENG W M, CUI Z, et al. Multi-cue fusion for emotion recognition in the wild [J]. Neurocomputing, 2018, 309: 27–35.
- [9] GUNES H, PICCARDI M. Bi-modal emotion recognition from expressive face and body gestures [J]. Journal of Network and Computer Applications, 2007, 30 (4): 1334–1345.
- [10] WANG Y J, GUAN L, VENETSANOPOULOS A N. Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition [J]. IEEE Transactions on Multimedia, 2012, 14 (3): 597–607.
- [11] 闫静杰, 郑文明, 辛明海, 等. 表情和姿态的双模态情感识别 [J]. 中国图象图形学报, 2013, 18 (9): 1101–1106.
YAN Jingjie, ZHENG Wenming, XIN Minghai, et al. Bi-modal emotion recognition based on body gesture and facial expression [J]. Journal of Image and Graphics, 2013, 18 (9): 1101–1106. (in Chinese)
- [12] 闫静杰, 卢官明, 李海波, 等. 基于人脸表情和语音的双模态情感识别 [J]. 南京邮电大学学报(自然科学版), 2018, 38 (1): 60–65.
YAN Jingjie, LU Guanming, LI Haibo, et al. Bimodal emotion recognition based on facial expression and speech [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2018, 38 (1): 60–65. (in Chinese)
- [13] DOBRIŠEK S, GAJŠEK R, MIHELIC F, et al. Towards efficient multi-modal emotion recognition [J]. International Journal of Advanced Robotic Systems, 2013, 10 (1): 53.
- [14] NEMATI S, NAGHSH-NILCHI A R. An evidential data fusion method for affective music video retrieval [J]. Intelligent Data Analysis, 2017, 21 (2): 427–441.
- [15] 王晓华, 侯登永, 胡敏, 等. 时空 LBP 矩和 Dempster-Shafer 证据融合的双模态情感识别 [J]. 光电工程, 2016, 43 (12): 154–161.
WANG Xiaohua, HOU Dengyong, HU Min, et al. Dual-modality emotion recognition model based on temporal-spatial LBP moment and Dempster-Shafer evidence fusion [J]. Opto-Electronic Engineering, 2016, 43 (12): 154–161. (in Chinese)
- [16] SONG K S, NHO Y H, SEO J H, et al. Decision-level fusion method for emotion recognition using multimodal emotion recognition information [C] // Proceedings of the 15th International Conference on Ubiquitous Robots. 2018: 472–476.

- [17] CORREA N M, ADALI T, LI Y O, et al. Canonical correlation analysis for data fusion and group inferences[J]. IEEE Signal Processing Magazine, 2010, 27(4): 39–50.
- [18] GAO L, QI L, CHEN E Q, et al. Discriminative multiple canonical correlation analysis for information fusion[J]. IEEE Transactions on Image Processing, 2018, 27(4): 1951–1965.
- [19] LI B, QI L, GAO L. Multimodal emotion recognition based on kernel canonical correlation analysis[C]//Proceedings of the IEEE Workshop on Electronics, Computer and Applications. 2014: 934–937.
- [20] GÖNEN M, ALPAYDIN E. Multiple kernel learning algorithms[J]. Journal of Machine Learning Research, 2011, 12: 2211–2268.
- [21] BOUBENNA H, LEE D. Feature selection for facial emotion recognition based on genetic algorithm[C]//Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery. 2016: 511–517.
- [22] EL AYADI M, KAMEL M S, KARRAY F. Survey on speech emotion recognition: features, classification schemes, and databases[J]. Pattern Recognition, 2011, 44(3): 572–587.
- [23] EYBEN F, WÖLLMER M, SCHULLER B. OpenSMILE—the Munich versatile and fast open-source audio feature extractor[C]//Proceedings of the ACM International Conference on Multimedia. 2010: 1459–1462.
- [24] MARTIN O, KOTSIA I, MACQ B, et al. The eNTERFACE'05 audio-visual emotion database[C]//Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW06). 2006: 1–8.