基于 KCCA 算法的 EEG 和语音特征融合情绪识别

刘晴[1]*

摘要

情绪识别是智能人机交互的重要环节。传统方法主要采用单模态情感特征进行识别，情感识别率低。针对该问题，本文提出了一种核典型相关分析算法（KCCA）的多特征（multi-features）融合情感识别方法。在特征层面上，分别选取外在直观表达信号——语音信号和生理信号——脑电波进行特征提取，然后利用两种特征互补性，采用核典型相关分析算法（KCCA）将它们进行融合，降低特征向量的维数。最后选择 SVM 模型对情感识别的训练集进行建模，并通过具体情感数据集进行仿真实验．实验结果表明，核相关分析算法有效的提高了情感识别的正确率。

关键词

情感识别　核典型相关分析算法　特征融合　脑电特征

[1] 西北工业大学　　***通讯作者**: elbox@qq.com

1. 引言

随着智能技术的发展，传统的人机交互方法已经不能满足人们的需求，在人机交互方向，人们不只满足于机器人对人的语言进行识别，机器人是否具备情绪识别能力是推进智能人机交互的关键。为了实现机器人与人类的自主情感交互，我们需要机器人有更高的情绪识别能力。

情绪识别在智能人机交互、机器人领域是十分值得探究的前沿热点方向。目前，情绪识别研究多数基于人的面部表情、语音和生理信号来进行。本文基于语音和生理信号的进行情绪识别。语音是一种情绪的外在表现，能够直观地反应说话人的情绪，但是，由于语音可能会被人为地伪装，有时难以检测到说话人的真实情绪。并且，单从语音信号来进行情绪分类识别，难以相近语音特征对应情绪（比如生气和惊讶）的识别。EEG 是人的生理反应，可以更直接地反映人的情感状态。然而 EEG 信号复杂度高且十分微弱，并且原始 EEG 信号噪声来源较多，比如机器的工频干扰、电磁干扰，当人进行呼吸、眨眼、说话、晃动时均会对 EEG 信号产生干扰。如果只用 EEG 信号进行情绪分类，则会忽视人的外在情绪表现，而且受设备限制，在有些实际应用场景中难以有效执行。基于这两方面考虑，本文同时应用语音和 EEG 信号，利用两种信号的互补特性，以提高情绪识别准确率。

在特征融合方面，简单的将多种情感特征组合到一起，不但特征维数高，运算代价大，识别慢，并且含有很多无效信息，影响识别率。为此有些学者提出采用主成分分析等算法对特征进行融合和选择，使识别效率得到提升。但 PCA 是一种线性特征融合算法，无法描述特征之间的非线性关系。本文提出了采用核典型相关分析算法 (KCCA) 进行多特征融合。核典型相关分析（KCCA）是相关分析（CCA）的非线性扩展，CCA 是一种非线性特征的提取方法，用核方法来增强典型相关分析方法，不仅能剔除冗余信息，增强鉴别信息，还能解决特征的线性不可分问题，挖掘非线性因素，具有更好的特征提取与表征能力。

2. 情感识别特征的提取

2.1 语音信号特征提取

语音情感识别时，需要选取适当的声学参数来对情感进行识别分类。常用的声学参数主要有三大类：韵律学特征、音质特征和基于谱的相关特征。它们分别从不同侧面对语音的情感信息进行表达，均对语音中的情感有一定的区分能力，相对于使用一种特征获取的语音信息比较单一，采用多种特征进行语音情感识别会获得更多的语音情感信息，可提高识别性能。本文中选择了声音韵律、声音质量和基于谱的相关特征共 3 种特征参数，提取了如下的情感特征：能量，语速，基频，共振峰，MFCC。

1. 能量特征提取

语音的能量反应在语音的响亮度上的，在日常生活中我们都有所体会，当人处于愤怒或者惊讶这种激动的情绪状态时，声音响度通常会比较大，而人悲伤或难过时，声音就会变得相对低沉。所以能

量特征对某些情感有较好的区分能力，在本文中选取能量的最大值、最小值和均值做为特征。

2. 语速特征提取

语速是通过语音信号中包含的音节数和该语句的语音发音时间进行定义的。其值为语音信号发音持续时间与发音中音节数的比值 (音节/s)，如下所示。

$$v_{vs} = \frac{s_{vd}}{n_{ps}} \tag{1}$$

语音发音持续时间是包括音节间的停顿的，因不同的停顿频率、时长能反应出不同的情感状态。本文中，为了避免统计语音信号中音节的复杂性，用相对的方法来表示语速。对情感语料库中的数据进行处理时，将平静时的语音 (即中性情感状态下的语音) 的速度记做 1，则其他情感状态下的相对语速为该语音信号的发音时长与对应的中性情感状态下的发音时长的比值，如下所示。

$$v_{rvs} = \frac{s_{osvd}}{s_{vdnes}} \tag{2}$$

3. 基频特征提取

通常每一个语音帧中会含有 1 到 7 个基音周期，基音周期的倒数为基音频率，即是发浊音时声带振动的频率。通常用基音频率或着基音周期来描述语音的韵律变化，但是个体声道差异问题导致了对基音周期的精确检测存在较大的困难。因为基于其对语音信号的重要性，所以其对基音周期的研究一直是备受关注的课题。常用的基音周期检测算法主要有：倒谱法、小波法、自相关函数法、平均幅度差函数法等，本文中使用的是自相关函数的方法。

假如语音信号为 $s(n)$，通过语音信号预处理，用汉明窗将信号分成 N 帧，加窗之后语音信号为 $s_w(n)$, 语音信号的短时自相关函数 $R_n(k)$ 的计算公式如3所示。

$$R_n(k) = \sum_{m=0}^{N-1-K} x_n(m)x_n(m+k) \tag{3}$$

基音的周期信号为浊音，而浊音的短时自相关函数具有周期性，即浊音信号周期。清音接近于随机噪声，其短时间自相关函数不具备周期性。根据此特性，可以对一个语音信号是清音还是浊音以及浊音的基音周期进行判断。

从3式可以看出：自相关函数的值为两个等长序列乘积的累加和，并且自相关函数的值随 k 的增高而降低。通过自相关函数的计算波形周期，若窗口长度不够，那么其所包含的周期就不够多，这样会导致计算困难。因此，使用修正的短时自相关函数来解决这类问题。

4.MFCC 特征提取

人的耳朵对不同频率语音信号感知差异很大的，在低频部分对声音的感知与声音的频率成线性关系，在高频的部分却为对数关系，由此定义了一种新的频率：Mel 频率，这种频率与平常频率 (以 Hz 为单位的频率) 之间的关系如公式4所示。

$$Mel(f) = 1125\ln(1 + F/700) \tag{4}$$

MFCC(Mel 频率倒谱系数）是根据人耳听觉领域特性而设计的，计算步骤如下：

(1) 通过加窗处理后，现将语音信号变为短时 $x_n(m)$，将时域信号换成频域信号，计算它的短时功率谱 $|X_n(m)|$。

(2) 在 Mel 频率域内配置 K 个通道的三角形滤波器组 $H_k(i)$，然后计算功率谱经过每一个滤波器的输出：

$$m(k) = \sum_i H_k(i)|X_n(i)|(k = 1, 2, ..., K) \tag{5}$$

如5公式中，第 i 个滤波器可以表示为 $H_k(i)$

(3) 滤波器输出对数能量为，如6 所示：

$$m_k = \log m(k) \tag{6}$$

(4) 经过离散余弦变换得到 MFCC 的特征参数为，如7

$$C_{MFCC}(l) = \sqrt{\frac{2}{N}} \sum_{k-1}^{k} m_i \cos(k - 0.5)\frac{l\pi}{k} \tag{7}$$

本文中通过计算出了 MFCC 的特征参数为 $C_1, C_2, ..., C_N$, 然后取它的一阶方差和二阶方差为语音的情感特征。

5. 共振峰特征提取

共振峰是一个反映声道特性的重要参数，还与语音质量相关的声学参数。语音是由空气气流经过声道产生的，这时会在一组频率上产生共振，这称为共振峰频率。当人们处于不同情感状态时，其神经的紧张程度是不同，导致声道产生形变，从而改变共振峰频率。因此，共振峰是同说话人情感状态之间是有密切关系的。

提取共振峰特征时常用的提取方法有：线性预测法、离散傅立叶分析法、带通滤波器组法、倒谱法等。考虑到线性预测描述声道模型的准确性，本文决定要使用线性预测法对共振峰参数特征进行提取，并选取了第一、第二、第三共振峰的最大值、最小值、均值和方差作为特征。

2.2 脑电信号特征提取

用于情绪识别的 EEG 特征值基本分为统计特征、时域特征、频域特征和熵特征等。

1. EEG 信号的均值为：

$$\mu_X = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{8}$$

其中，N 为 EEG 信号的总采样点数，X_i 为第 i 个点的 EEG 信号电压值。

2. 标准偏差

$$\sigma_X = \left(\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\mu_X)^2\right)^{1/2} \tag{9}$$

3. 一阶差分标准偏差

$$\delta_X = \frac{1}{N-1}\sum_{i=1}^{N-1}|X_{i+1}-X_i| \tag{10}$$

4. 二阶差分标准偏差

$$\gamma_X = \frac{1}{N-2}\sum_{i=1}^{N-2}|X_{i+2}-X_i| \tag{11}$$

5. 偏度

偏度是度量数据分布的偏斜方向和程度的参数，计算公式为：

$$S_k = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\mu_X)^3/\sigma_X^3 \tag{12}$$

6. 峰度

峰度是对总体中所有数据的分布形态的陡峭程度的度量，其表示为：

$$K_u = \frac{1}{n-1}\sum_{i=1}^{N}\frac{(X_i-\mu_X)^4}{\sigma_X^3} - 3 \tag{13}$$

7. θ 、 α 、 β 波功率

脑电信号可被分为不同的频带，分别是 θ 波（4-8Hz）、 α 波（8-13Hz）、 β 波（13-30Hz），人们在不同状态时，各类波有显著差异。因此将其分别提取出来表示情绪的不同特征。

8. 各频带波功率比

由于人处于不同状态时，各个频率的波段出现概率不同，使用其比值能够反映状态的差异。

9. 近似熵

近似熵是用来评判数据复杂性的参数。近似熵的抗干扰和抗噪能力比较强，无论是随机信号或是确定性信号，都可以使用近似熵进行分析，尤其是对分析生理信号十分有效，因为生理信号中既含有确定性成分，又含有随机成分。

以下是近似熵公式推导过程：

有 N 个点 u（1），u （2），…,u（N）, 以及两个固定参数 m 和 r，m 是窗长，r 是一个有效阈值，有极限值 ApEn(m，r) 和这 N 个点的统计估计值 ApEn(m，r，N)。

（1）将这 N 个点按顺序组成一组 m 维向量：从 X（1）到 X（N-m+1）, 其中，X(i)=[u(i)，u(i+1)，…，u(i+m-1)],i=1 N-m+1。这些向量为从第 i 个点开始的顺序排列的 m 个 u 值。

（2）X（i）与 X（j）中对应元素差值最大的值我们将其定义为两个向量之间的距离 d[X(i),X(j)]，即

$$d[X(i),X(j)] = max[|u(i+k)-u(j+k)|], k = 0\,m-1 \tag{14}$$

（3）给定阈值 r，对每个 i N-m+1 的值，统计 d[X(i)，X(j)] 小于 r 的数目及此数目与距离总数 N-m 的比值，记为 $C_i^m(r)$，即：

$$C_i^m(r) = \frac{1}{N-m}\{[d[X(i),X(j)]<r]\ \ \}i = 1\ N-m+1 \tag{15}$$

（4）对 $C_i^m(r)$ 取对数，并求平均值 $O^m(r)$，即：

$$O^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\ln C_i^m(r) \tag{16}$$

（5）维数加 1，为 m+1，重复（1）- （4），得到 $C_i^{m+1}(r)$ 和 $O^{m+1}(r)$。（6）则此序列的近似熵为：

$$ApEn(m,r) = \lim_{N\to\infty}[O^m(r)-O^{m+1}(r)] \tag{17}$$

若 N 为有限值，则数据的近似熵为：

$$ApEn(m,r,N) = O^m(r)-O^{m+1}(r) \tag{18}$$

10. 分形维数

几何模型的分形维数值能够反映该模型几何复杂度。

现有有限维的时间序列样本 X(1),X(2),…,X(N)，新的时间序列值表示为：

$$X_k^m : X(m), X(m+k), X(m+2k), ..., X(m+\left[\frac{N-m}{k}\right]k) \tag{19}$$

其中，m=1,2,…,k。

那么则有：

$$L_m(k) = \frac{\left\{\left(\sum_{i=1}^{\frac{N-m}{k}}|X(m+ik)-X(m+(i-1)k)|\right)\frac{N-1}{[\frac{N-m}{k}]k}\right\}}{k} \tag{20}$$

这里用 $\langle L(k)\rangle$ 表示 k 个点的 $L\_m(k)$ 的平均值。

分形维数可从由 k 和 $\langle L(k)\rangle$ 所画的对数图中得到。

3. 基于 KCCA 的特征融合

核相关分析算法将样本非线性映射到高维特征空间，然后进行相关分析，得到两组变量间的非线性关联。设 $X=(x_1,x_2,...,x_n)$ 和 $Y=(y_1,y_2,...,y_n)$ 分别表示语音情感特征向量和 EEG 特征向量，核相关分析算法通过两个非线性映射 ϕ 和 ψ 作用于两组特征向量

$$\begin{cases}\phi: x\to\phi(x)\in F_x\\ X\to\phi(x)=[\phi(x_1),\phi(x_2),...,\phi(x_n)]\end{cases}, \quad (21)$$

$$\begin{cases}\psi: y\to\psi(y)\in F_y\\ Y\to\psi(y)=[\psi(y_1),\psi(y_2),...,\psi(y_n)]\end{cases}, \quad (22)$$

语音特征和 EEG 特征的核函数分别是 K_x 和 K_y，则有

$$\begin{cases}K_x=\phi^T(x)\phi(x)\\ K_y=\psi^T(x)\psi(x)\end{cases}, \quad (23)$$

核矩阵中心化是对训练样本进行了零均值化:

$$\overline{K}=K-\frac{1}{N}1_{N\times N}K-\frac{1}{N}K1_{N\times N}+\frac{1}{N}1_{N\times N}K1_{N\times N} \quad (24)$$

核典型相关分析算法的目标是寻找投影方向 α_φ 和 β_ω，使得如下准则函数式最大

$$J(\alpha_\varphi,\beta_\omega)=\frac{\alpha_\varphi^T\phi(x)\psi(y)^T\beta_\omega}{\sqrt{\alpha_\varphi^T\phi(x)\phi(x)^T\alpha_\varphi\cdot\beta_\omega^T\psi(y)\psi(y)^T\beta_\omega}} \quad (25)$$

向量 α_φ 位于样本 $\phi(x_1),\phi(x_2),...,\phi(x_n)$ 张成的空间，根据核再生理论，则存在 N 维向量 ξ 使 $\alpha_\varphi=\phi(x)\xi$，同理，存在 N 维向量 η 使得 $\beta_\omega=\psi(y)\eta$，带入下式中得到：

$$J(\xi,\eta)=\frac{\xi^TK_xK_y\eta}{\sqrt{\xi^TK_x^2\xi\cdot\eta^TK_y^2\eta}} \quad (26)$$

防止产生没有意义的典型相关向量，需要引入正则项对式进行约束：

$$J(\xi,\eta)=\frac{\xi^TK_xK_y\eta}{\sqrt{\xi^T((1-\tau)K_x^2+\tau K_x)\xi\cdot\eta^T((1-\tau)+\tau K_y^2)\eta}} \quad (27)$$

式中，$0\le\tau\le1$。因此，核典型相关分析算法就转化为关于 ξ,η 的约束优化问题，目标函数为：$\max\xi^TK_xK_y\eta$

约束条件：

$$\begin{cases}\xi^T((1-\tau)K_x^2+\tau K_x)\xi=1\\ \eta^T((1-\tau)K_y^2+\tau K_y)\eta=1\end{cases}, \quad (28)$$

利用拉格朗日乘数法求解上述带约束的极值问题，则相应的拉格朗日方程为：

$$L(\xi,\eta)=\xi^TK_xK_y\eta-\frac{\lambda_1}{2}(\xi^T((1-\tau)K_x^2+\tau K_x)-\frac{\lambda_2}{2}(\eta^T((1-\tau)K_y^2+\tau K_y)\eta-1). \quad (29)$$

式中，λ_1 和 λ_2 为拉格朗日乘数。分别求 $L(\xi,\eta)$ 关于 ξ,η 的偏导数并令其为零，即

$$\begin{cases}\frac{\partial L}{\xi}=K_xK_y\eta-\lambda_1((1-\tau)K_x^2+\tau K_x)\xi=0\\ \frac{\partial L}{\eta}=K_xK_y\xi-\lambda_2((1-\tau)K_y^2+\tau K_y)\eta=0\end{cases}, \quad (30)$$

两边同时左乘 ξ^T 并由 $\xi^T((1-\tau)K_x^2+\tau K_x)\xi=1$ 得到 $\xi^TK_xK_y\eta=\lambda_1$ 同理得到 $\eta^TK_xK_y\xi=\lambda_2$ 由此可知

$$J(\xi,\eta)=\eta^TK_xK_y\xi=\xi^TK_xK_y\eta=\lambda_1=\lambda_2 \quad (31)$$

又因为 $J(\xi,\eta)\leqslant1$，当取得最大值 1 时，必有 $\lambda_1=\lambda_2=1$。从而，核典型相关分析算法等价于求解如下广义特征方程对应的特征向量问题，即

$$\begin{cases}K_xK_y\eta=((1-\tau)K_x^2+\tau K_x)\xi\\ K_xK_y\xi=((1-\tau)K_y^2+\tau K_y)\eta\end{cases}, \quad (32)$$

求解出 ξ,η 提取 x 和 y 之间的非线性相关特征

$$\begin{cases}\mu=\xi K_x\\ \nu=\eta K_y\end{cases}, \quad (33)$$

式中，μ 和 ν 是变换后的特征分量。将其线性变换，

$$Z=\begin{pmatrix}\mu\\ \nu\end{pmatrix}=\begin{pmatrix}\xi & 0\\ 0 & \eta\end{pmatrix}=\begin{pmatrix}K_x\\ K_y\end{pmatrix} \quad (34)$$

4. SVM 情感分类器

当前语音情感识别的分类器有神经网络，支持向量机以及 K 近邻算法等，其中 K 近邻算法的情感识别率低，而神经网络要求训练样本多，分类结果不稳定，相对其他算法，支持向量机的情感分类效果更佳。并且 SVM 在解决实际问题中应用广泛，其理论完善，具有鲁棒性，计算简单，有众多可利用软件工具。

5. 仿真实验

为了测试核相关分析的情数据库采用感识别性能，在 matlab 上进行实验验证。支持向量机采用 LSSVM 软件，选择 CIAIC 实验室测量的语音和脑电数据进行实验。该数据在全消声暗示中测量，共有 18 位被试，包含 4 种情绪：中立，悲伤，愤怒，高兴。通过观看影片的方式激发被试的情绪，随后朗读对应情绪的语音文本，并记录下对应的脑电数据和语音数据。在最终的数据集中，语音数据被截取为独立语句的形式，脑电数据也按照对应顺序进行截取。本实验选取数据库中的部分数据进行测试，每种情绪随机选取 200 句语音数据和对应的脑电数据作为训练数据，再在每种情绪中另取 39 个数据组成测试集。KCCA 函数采用 RBF 核函数，SVM 选择高斯核。为使 KCCA-SVM 的结果更有说服力，选择与使用单一特征以及直接拼接特征的方法进行对比。各模型的平均识别率如下表所示：

表 1. 情感识别正确率对比

特征	中立	悲伤	愤怒	高兴	识别率
Speech	0.21	0.49	0.67	0.49	0.49
EEG	0.77	0.36	0.36	0.67	0.54
EEG+Speech	0.69	0.54	0.44	0.77	0.61
KCCA	0.82	0.62	0.67	0.87	0.74

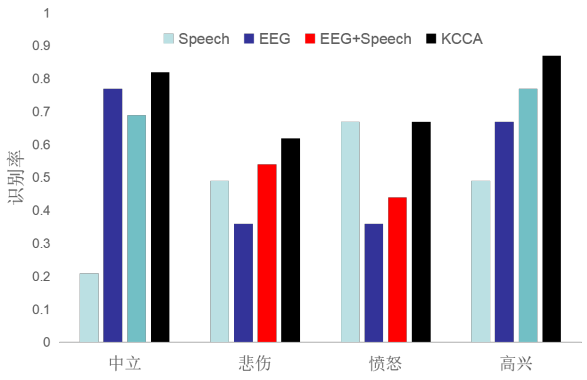

图 1. 情感识别正确率对比

从结果我们可以看出：1. 脑电信号在总体的识别率上略高于语音信号。2. 多模态特征对于情感识别率的提升效果是明显的，即使是简单的特征直接组合的方式，也可以得到较好的结果。3. 利用 KCCA 算法进行特征融合，进一步显著的提升了识别率，可见用该方法，不仅能剔除冗余信息，增强鉴别信息，还能解决特征的线性不可分问题，挖掘非线性因素，具有很好的特征提取与表征能力。4. 不同特征对于不同情绪识别有明显的差异，这一点可以从数据本身的特性来解释，脑电数据反应的是生理上的反应，而在实际录制数据过程中，即使存在激励视频，被测试人可能不会产生十分强烈的愤怒和悲伤情绪，而这两种情绪在语音上是可以明显的表达的，因此产生这样的识别结果是符合实际情况的。

6. 思考与改进

总的来说本次仿真实验其实还有很多的不足之处。首先是在数据上，一方面是对脑电数据还不够了解，对于提取特征时所用的一些参数还有待考证。另一方面，脑电数据与语音并不完全是一一对应的，实验室的数据没有记录文档，导致当时一些录制的数据细节现在无法考证，并且部分数据不完整，使用时需重新从原始数据提取。在数据的数量上面，由于时间关系，暂时仅提取了一个人的脑电特征，即实验中仅使用了一个人的脑电数据，没有进一步考虑个体差异的问题。接下来是在数据预处理上，都是直接使用提取的特征进行算法研究，没有进一步进行数据的优化等，比如使用较好的方式进行降维，语音特征有 93 维，影响可能不大，但是对于脑电数据特征有 1705 维，可能含有较多的冗余信息，如果进行数据的优化可能会有对结果有一定的提升，即使没有提升，能够降低计算代价也是比较好的。在编程过程中，尝试过在将特征送入分类器前进行 PCA 降维，但是对于结果基本没有影响，所以就没有使用。

在算法的实现上，最开始是自己根据理论尝试编程，理论上思路是十分清晰明了的，但是从理论到编程的实现还是有很大差距的。问题基本集中在最后的广义特征方程组那，借助 matlab 的函数，虽然程序能够调通，但是融合后基本没有识别出来，最后还是通过参考一个工具包中的源程序进行编程，最后才能成功运行。虽然最后结果有明显提升，但是从中间一些结果来看，其实还存在一些问题，还有改进的空间，距离将其做成一个通用的工具还是有很大距离的。

参考文献

[1] 刘颖, 贺聪, 张清芳. 基于核相关分析算法的情感识别模型 [J]. 吉林大学学报: 理学版, 2017, 55(6):1539-

1544.

[2] 刘付民, 张治斌, 沈记全. 核典型相关分析算法的多特征融合情感识别 [J]. 计算机工程与应用, 2014, 50(9):193-196.

[3] 张前进, 王华东. 基于核典型相关分析和支持向量机的语音情感识别模型 [J]. 南京理工大学学报, 2017, 41(2):191-197.

[4] 林克正, 王海燕, 李骜, 等. 高效求解方法的核典型相关分析算法 [J]. 计算机科学与探索, 2017, 11(2):286-293.