

基于局部保持投影和主元分析的语音情感识别^①

韩志艳, 王 健

(渤海大学 工学院, 锦州 121000)

摘 要: 为了提高情感识别的准确性, 该文以语音信号为研究对象, 提出了一种新型的语音情感识别方法. 将局部保持投影算法(LPP)的思想融入到主元分析(PCA)的目标函数中, 使得在原始变量空间投影到低维空间的过程中, 不仅实现了整体方差的最大化, 而且保持了局部近邻结构不变, 有利于全局和局部特征的全面提取, 克服了传统 PCA 方法只关注全局结构特征而忽略局部特征的缺陷. 对比实验结果验证了该方法的可行性和有效性, 实现了对喜悦、愤怒、悲伤、恐惧和中性 5 种人类基本情感的识别, 研究成果将为情感识别提供新的研究方法, 促进人机交互系统进一步深入发展.

关键词: 语音信号; 情感识别; 局部保持投影; 主元分析; 特征提取; 神经网络

Speech Emotion Recognition Based on Locality Preserving Projections and PCA

HAN Zhi-Yan, WANG Jian

(College of Engineering, Bohai University, Jinzhou 121000, China)

Abstract: In order to improve the accuracy of emotion recognition, this paper proposes a novel speech emotion recognition algorithm, and takes speech signal as the research subject. The idea of locality preserving projection (LPP) is integrated into the objective function of PCA. It not only realizes the maximum of the total variance, but also keeps the local neighbor structure. That is beneficial to comprehensive extraction of global and local features, and overcomes the defects that the traditional PCA can only keep the structure in global and can not maintain the structure in local. Experiment results verify the feasibility and effectiveness of the proposed method, and accomplish recognition for five kinds of human emotion (joy, anger, sadness, fear, neutral). Research results provide new methods into emotion recognition, promote the further development of human-computer interaction system.

Key words: speech signal; emotion recognition; locality preserving projections (LPP); principal component analysis (PCA); feature extraction; neural network

随着信息技术的高速发展和人类对计算机的依赖性的不断增强, 人机的交互能力越来越受到研究者的重视. 如何实现计算机的拟人化, 使其能够感知周围的环境和气氛以及对象的态度、情感等内容, 自适应地为对话对象提供最舒适的对话环境, 尽量消除操作者和机器之间的障碍, 已经成为下一代计算机发展的目标^[1-5]. 斯坦福大学的 Reeves 和 Nass 的研究发现表明, 在人机交互中需要解决的问题实际上与人和人交流中的重要因素是一致的, 最关键的都是“情感智能”

的能力, 计算机要能够更加主动地适应操作者的需要, 首先必须能够识别操作者的情感.

语音作为人类交流的最重要媒介之一, 携带着丰富的情感信息. 因此我们可以考虑从语音信号中提取情感相关的特征来进行情感识别^[6-7]. 1990 年麻省理工大学多媒体实验室构造了一个“情感编辑器”对外界各种情感信号进行采样来识别各种情感, 并让机器对各种情感做出反应^[8]. 中科院的颜永红^[9]采用非均匀子带滤波器来挖掘对语音情感有益的信息, 加大了各类

① 基金项目: 国家自然科学基金(61503038, 61403042)

收稿时间: 2016-02-15; 收到修改稿时间: 2016-03-14 [doi:10.15888/j.cnki.csa.005389]

情感之间的鉴别性,提高了情感识别的性能.北京航空航天大学毛峡^[10]通过用相关密度和分形维数作为情感特征参数来进行语音情感识别,获得了较好的性能.东南大学的邹采荣^[11]提出了一种基于改进模糊矢量量化的语音情感识别方法,有效地改善了现有模糊矢量量化方法的情感识别率.东南大学的赵力等^[12]通过采用全局与时序结构特征并用的方法进行了语音情感识别. Attabi等^[13]将锚模型的思想应用到了语音情感识别中,改进了识别系统的性能. Zheng等^[14]通过对传统的最小二乘回归算法进行改进,提出了不完稀疏最小二乘回归算法,能同时对标记和未标记语音数据进行情感识别. Mao等^[15]通过使用卷积神经网络来选择对情感有显著影响的特征.但上述方法还存在较多局限性,尤其是情感识别设备可能会应用在各种环境中,所以就要求系统具有较强的鲁棒性.

PCA是一种将多个相关变量转化为少数几个相互独立变量的有效分析方法,它不依赖于过程机理,只需通过过程数据的信息进行统计建模,即可实现以较少维数的数据描述对象的主要特征^[16].然而,PCA方法只能发现数据的全局欧几里德结构特征,而忽略了数据的局部结构特征.最近,大量研究表明^[17-19],流形学习能够有效地发现隐含在高维数据集中的低维特征,对局部结构特征具有较好的提取能力.多种流形学习算法也被提出,如:局部线性嵌入(LLE)、保距特征映射(Isomap)、拉普拉斯特征映射(LE)等.其中,LE的线性化映射—局部保持投影法(LPP)在继承LE算法能够保持局部流形特征的基础上,实现了线性计算,得到广泛的应用.因此该文结合PCA和LPP算法的优点,提出了一种新的数据降维和特征提取方法,以求更为准确的反应语音中所蕴含的真实情感.

1 PCA和LPP算法分析

1.1 PCA 算法分析

PCA方法认为方差是数据的最主要信息,其目的是为原始的高维数据找到一个方差最大化的低维表示,因此PCA的目标函数可以写为:

$$J(W)_{PCA} = \max_W \sum_{i=1}^n (y_i - \bar{y})^2 \quad (1)$$

其中 W 是投影向量, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. PCA提取的特征空间 Y 和原始变量空间 X 有着相同的最大方差方向,因此

也具有相似的“外部形状”.但这种方法由于没有考虑到数据点之间的局部近邻关系,使得PCA算法不能保持住原始数据中的本征几何结构,缺点也是很明显的.例如:假设有一串数据点 $x_1, x_2, \dots, x_n$,且 $[x_{l-1}, x_l]$ 在原始变量空间 X 是相邻的,其中 $l=2, 3, \dots, n$.在降维后的空间里,对应点 $y_1, y_2, \dots, y_n$ 的顺序可能已经被打乱,因为PCA在投影时没有考虑到数据中包含的这种近邻结构,从而导致了内部结构的破坏和信息的丢失.

1.2 LPP 算法分析

与PCA只关注数据的方差信息不同,LPP算法的基本思想是通过保存样本空间的局部结构,来提取高维数据中嵌入的低维几何流形.LPP的目标函数可以写为:

$$J(W)_{LPP} = \min_W \sum_{i,j=1}^n (y_i - y_j)^2 S_{ij} \quad (2)$$

其中 S 是 $n \times n$ 阶的关系矩阵,矩阵里元素 S_{ij} 代表了点 x_i 和 x_j 之间的近邻关系,它的值和 x_i 与 x_j 之间的距离成反比比例.

假设 x_i 与 x_j 是原始变量空间中彼此靠近的两个点,那么这两点之间的距离关系就表达了样本空间里的一种局部结构.由于 x_i 与 x_j 彼此靠的很近,此时 S_{ij} 被赋予了一个较大的数值,因此,只要最小化LPP的目标函数就可以保证它们的投影 y_i 和 y_j 在低维空间里依然是相近.但由于没有对原始变量空间 X 中那些相互远离的点进行约束,LPP很可能把这些点投影到一个小的区域里,从而导致了方差信息的损失和对整个数据集外部形状的破坏.另外,在模式分类问题中,彼此远离的点一般属于不同的类别,而LPP的目标函数里没有考虑这些远离点所包含的信息,因此很可能把它们投影到一起,使得在低维空间里无法区分它们的类别.从而限制了LPP应用于模式分类问题的效果.

2 该文算法描述

该文将PCA和LPP算法的优点相结合,提出了一种新的数据降维和特征提取方法.其主要思想是寻找一个投影矩阵 $W \in R^{M \times d}$,使得原始变量空间 $X = [x_1, x_2, \dots, x_n] \in R^M$ 通过投影映射 $y_i = W^T x_i$, $i=1, 2, \dots, n$ 得到的低维空间 $Y = [y_1, y_2, \dots, y_n] \in R^d$ ($d < M$)不仅和原始变量空间 X 具有相似的局部近邻结构,而且保留了原始变量空间 X 的大部分方差信息,实现了全局和局部特征的全面提取.

2.1 全局目标函数描述

假设原始向量为 $x_1, x_2, \dots, x_n$, $J(W)_{global}$ 的目标是寻找投影矩阵 W , 即寻找 d 个投影向量 $w_1, w_2, \dots, w_d$, 使得通过投影映射 $y_i = W^T x_i$, $i=1, 2, \dots, n$ 得到的低维空间 Y 能保留原始数据空间 X 中的大部分方差信息, 即:

$$\begin{aligned} J(W)_{global} &= \max_W \sum_{i=1}^n (y_i - \bar{y})^2 \\ &= \max_W \sum_{i=1}^n W^T (x_i - \bar{x})(x_i - \bar{x})^T W \\ &= \max_W W^T C W \end{aligned} \quad (3)$$

其中, $W = [w_1, w_2, \dots, w_d]$, $\bar{x} = (\sum_{i=1}^n x_i)/n$, $\bar{y} = (\sum_{i=1}^n y_i)/n$, $C = (x_i - \bar{x})(x_i - \bar{x})^T$.

可以得出, $J(W)_{global}$ 和 PCA 的目标函数都是通过线性空间变换将高维数据空间投影到低维数据空间的, 它们是一致的. 由于低维数据空间保留了原始数据空间的大部分方差信息, 因此实现了全局结构特征的提取. 但是, 全局目标函数中没有考虑样本点的局部几何关系, 在低维空间里, 样本点之间的局部几何关系有可能被打乱, 导致重要的信息在低维空间中的丢失.

2.2 局部目标函数描述

假设原始变量为 $x_1, x_2, \dots, x_n$, $J(W)_{local}$ 的目标是寻找投影矩阵 W , 即寻找 d 个投影向量 $w_1, w_2, \dots, w_d$, 使得通过投影映射 $y_i = W^T x_i$, $i=1, 2, \dots, n$ 得到的低维空间 Y 和原始变量空间 X 具有相似的局部近邻结构. 即:

$$\begin{aligned} J(W)_{local} &= \min_W \frac{1}{2} \sum_{i,j=1}^n (y_i - y_j)^2 S_{ij} \\ &= \min_W \frac{1}{2} \sum_{i,j=1}^n (W^T x_i - W^T x_j)^2 S_{ij} \\ &= \min_W \left(\sum_{i,j=1}^n W^T x_i S_{ij} x_j^T W - \sum_{i,j=1}^n W^T x_i S_{ij} x_j^T W \right) \\ &= \min_W \left(\sum_{i=1}^n W^T x_i D_{ii} x_i^T W - W^T X S X^T W \right) \\ &= \min_W (W^T X D X^T W - W^T X S X^T W) \\ &= \min_W W^T X (D - S) X^T W \\ &= \min_W W^T X L X^T W \end{aligned} \quad (4)$$

其中, S 是权重矩阵, S_{ij} 表示 x_i 和 x_j 之间的近邻关系, 一般取:

$$S_{ij} = \begin{cases} e^{-\|x_i - x_j\|^2/t}, & \text{若 } x_i \text{ 和 } x_j \text{ 是 } k \text{ 近} \\ 0, & \text{其它} \end{cases} \quad (5)$$

s.t. $0 < i < n, 0 < j < n$

其中, D 为对角阵, $D_{ii} = \sum_j S_{ij}$; $L = D - S$ 是 Laplacian 矩阵; $W = [w_1, w_2, \dots, w_d]$ 是投影矩阵.

可以看出, $J(W)_{local}$ 和 LPP 的目标函数都是通过使在原始变量空间 X 中相邻的点映射到低维空间 Y 中也相邻, 实现原始变量空间 X 和低维空间 Y 具有相似的局部流形结构, 完成对样本点局部特征的提取.

2.3 该文算法的目标函数描述

为了克服 PCA 算法的缺点, 该文将 LPP 算法的思想融入到 PCA 的目标函数中, 使得投影得到的低维空间能够获得更全面的特征信息, 其全局目标函数定义为:

$$J(W)_{global} = \max_W W^T C W \quad (6)$$

局部目标函数定义为:

$$\begin{aligned} J(W)_{local} &= \min_W W^T X L X^T W \\ &= \min_W (W^T X D X^T W - W^T X S X^T W) \end{aligned} \quad (7)$$

当我们施加的约束为 $W^T X D X^T W = I$ 时, 得到:

$$\begin{aligned} J(W)_{local} &= I - \min_W W^T X S X^T W \\ &= \max_W W^T X S X^T W \end{aligned} \quad (8)$$

该文算法的目标函数定义为:

$$\begin{aligned} J(W)_{LPPCA} &= \max_W (W^T C W + \lambda W^T X S X^T W) \\ &= \max_W (W^T (C + \lambda X S X^T) W) = \max_W W^T \tilde{S} W \end{aligned} \quad (9)$$

其中, $\tilde{S} = C + \lambda X S X^T$, λ 为调节参数, 用以调节全局特征和局部特征的提取比例, λ 越小越侧重于全局特征的提取, λ 越大越侧重于局部特征的提取. 该文算法的最终目标函数为:

$$\begin{aligned} \arg \max_W W^T \tilde{S} W \\ \text{s.t. } W^T X D X^T W = I \end{aligned} \quad (10)$$

使公式(10)达到最大的投影矩阵, 可以利用拉格朗日方法, 通过求解下式最大的 d 个特征值 $a_1, a_2, \dots, a_d$ 所对应的特征向量 $w_1, w_2, \dots, w_d$ 的方式得到:

$$\tilde{S} W = a X D X^T W \quad (11)$$

即投影矩阵

$$W = [w_1, w_2, \dots, w_d] \quad (12)$$

3 仿真实验及结果分析

该文选用 8 个汉语语句作为情感分析用的语音资料, 并由 9 位(5 男 4 女)善于表演的说话者用喜悦、愤怒、悲伤、恐惧和中性情感对每一个句子各发 5 遍, 共采集到 360 句实验数据, 其中 100 句为训练数据集, 260 句为测试数据集, 采样频率为 16kHz. 建立的 PCA 模型选取的主元数目为 6 个, 选取低维的维度也为 6, 选取近邻参数 $k=10$. 实际应用中调节参数 λ 是通过多

次实验获得的经验参数, 该文取 $\lambda=0.5$.

该文选用的原始特征参数共 42 个, 其特征参数排列顺序为:

- 1 号: 第一共振峰频率的平均值;
- 2 号: 第二共振峰频率的平均值;
- 3 号: 第三共振峰频率的平均值;
- 4 号: 第四共振峰频率的平均值;
- 5 号: 谐波噪声比的均值;
- 6 号: 谐波噪声比的最大值;
- 7 号: 谐波噪声比的最小值;
- 8 号: 谐波噪声比的方差;
- 9 号: 过零率;

10 号-21 号: 线性预测倒谱系数 LPCC (Linear Predictor Cepstral Coefficient);

22 号-33 号: 梅尔频率倒谱系数 MFCC (Mel-frequency Cepstral Coefficient);

34 号: 语句发音持续时间与相应的平静语句持续时间的比值;

35 号: 基音频率平均值;

36 号: 基音频率最大值;

37 号: 基音频率平均值与相应平静语句的基音频率平均值的差值;

38 号: 基音频率最大值与相应平静语句的基音频率最大值的差值;

39 号: 振幅平均能量;

40 号: 振幅能量的动态范围;

41 号: 振幅平均能量与相应平静语句的振幅平均能量的差值;

42 号: 振幅能量动态范围与相应平静语句的振幅能量动态范围的差值.

在测试集中随机选取每种情感的 20 句数据样本进行分类显示. 图 1 给出了在测试数据集中基于 PCA 的特征提取方法获得的前两维特征向量, 其中横坐标为第一维特征, 纵坐标为第二维特征. 图 2 给出了在测试数据集中基于该文方法获得的前两维特征向量.

从图 1 和图 2 可以看出, 该文方法明显优于 PCA 方法, 其主要原因在于该文方法将全局和局部两个目标函数进行结合, 其全局目标函数能使低维空间可以保留原始数据空间的大部分方差信息, 局部目标函数能使低维空间和原始数据空间具有相似的局部近邻结构.

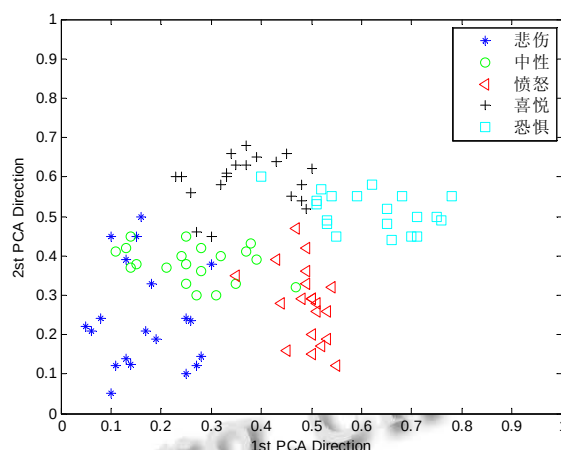


图 1 基于 PCA 的方法

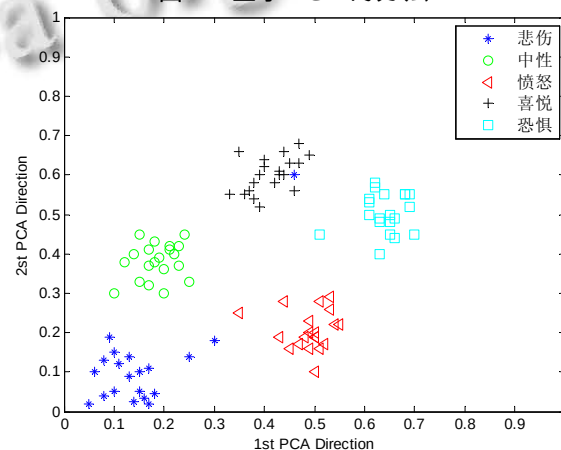


图 2 基于该文的方法

为了比较 PCA 方法和该文方法的性能, 采用遗传小波神经网络作为识别分类器, 具体算法参见文献 [20], 表 1 为对比实验结果.

表 1 对比实验结果

情感类别	PCA 法识别正确率(%)	该文方法识别正确率(%)
喜悦	92.15	98.33
愤怒	90.36	95.83
悲伤	93.28	98.21
恐惧	88.75	96.01
中性	85.39	96.22
平均正确识别率(%)	89.99	96.92

由表 1 可知, 该文方法获得了较高的识别正确率, 达到了 96.92%, 而 PCA 方法只获得了 89.99% 的平均识别正确率. 由于该文方法不仅实现了全局特征的保持, 而且兼顾了局部特征的提取, 因此获得了较好的识别结果.

4 结语

该文提出了一种新的数据降维及特征提取方法,可以通过调整调节参数来改变整体特征和局部特征的保持比例.而且仿真实验结果也证实了该方法的可行性.但是该文只是针对特定文本的语音情感进行识别,因为语音信号的情感特征总是和特定文本信息密切相关,如果进行非特定文本的情感识别,还需提取更为实用的情感特征参数,所以非特定文本的语音情感识别将成为我们下一步的研究方向.

参考文献

- 1 Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 2001, 18(1): 32–80.
- 2 Murray I, Arnott J. Towards the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of Acoustic Society of America*, 1993, 93(2): 1097–1108.
- 3 Kim EH, Hyun KH, Kim SH, Kwak YK. Improved emotion recognition with a novel speaker-independent feature. *IEEE Trans. on Mechatronics*, 2009, 14(3): 317–325.
- 4 Cowie R, Cornelius RR. Describing the emotional states that are expressed in speech. *Speech Communication*, 2003, 40(1-2): 5–32.
- 5 Camurri A, Volpe G, Poli GD, Leman M. Toward communicating expressiveness and affect in multimodal interactive systems for performing arts and cultural applications. *IEEE Multimedia*, 2005, 12(1): 43–53.
- 6 赵力,钱向民,邹采荣,吴镇杨.语音信号中的情感识别研究. *软件学报*, 2001, 12(7): 1050–1055.
- 7 赵力.语音信号处理.第2版.北京:机械工业出版社, 2009.
- 8 余伶俐,蔡自兴,陈明义.语音信号的情感特征分析与识别研究综述. *电路与系统学报*, 2007, 12(4): 76–84.
- 9 颜永红,周瑜,孙艳庆,李军锋.一种用于语音情感识别的语音情感特征提取方法. *Int. Cl:GIOL15/02 Schweiz Pat* 2010102729713. [2010-09-03].
- 10 Mao X, Chen LJ. Speech emotion recognition based on parametric filter and fractal dimension. *IEICE Trans on Information and Systems*, 2010, 93(8): 2324–2326.
- 11 邹采荣,赵力,赵艳,魏昕.一种基于改进模糊矢量量化的语音情感识别方法. *Int. Cl:GIOL15/06 Schweiz Pat* 2008101228062. [2008-07-03].
- 12 赵力,王治平,卢韦,邹采荣,吴镇杨.全局和时序结构特征并用的语音信号情感特征识别方法. *自动化学报*, 2004, 30(3): 423–429.
- 13 Attabi Y, Dumouchel P. Anchor models for emotion recognition from speech. *IEEE Trans. on Affective Computing*, 2013, 4(3): 280–290.
- 14 Zheng WM, Xin MH, Wang XL, Wang B. A novel speech emotion recognition method via incomplete sparse least square regression. *IEEE Signal Processing Letters*, 2014, 21(5): 569–572.
- 15 Mao Q, Dong M, Huang Z, Zhan Y. Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Trans. on Multimedia*, 2014, 16(8): 2203–2213.
- 16 肖应旺,徐保国.改进 PCA 在发酵过程监测与故障诊断中的应用. *控制与决策*, 2005, 20(5): 571–574.
- 17 张沐光,宋执环. LPMVP 算法及其在故障诊断中的应用. *自动化学报*, 2009, 35(6): 766–772.
- 18 Shao JD, Rong G. Nonlinear process monitoring based on maximum variance unfolding projections. *Expert Systems with Applications*, 2009, 36(8): 11332–11340.
- 19 Hu K, Yuan J. Batch process monitoring with tensor factorization. *J of Process Control*, 2009, 19(2): 288–296.
- 20 韩志艳,王健,伦淑娴.基于遗传小波神经网络的语音识别分类器设计. *计算机科学*, 2010, 37(11): 243–246.