

文章编号:1004-9037(2014)02-0222-05

基于核主分量分析和典型相关分析的语音情感识别

卞金洪¹ 王吉林¹ 余威风² 赵 力²

(1. 盐城工学院信息工程学院,盐城,224051; 2. 东南大学信息科学与工程学院,南京,210096)

摘要:核主分量分析(Kernel principal component analysis ,KPCA)是一种利用核方法将主分量分析(Principal component analysis ,PCA)推广后的学习方法,KPCA方法能够使得输入空间线性不可分的样本在特征空间有更好的可分性。典型相关分析(Canonical correlation analysis ,CCA)是分析两组随机变量之间的相关性的一种统计方法。本文提出将KPCA方法用于语音情感识别中,并采用KPCA和CCA结合的方法用于情感识别。与传统的PCA方法进行了对比,研究结果表明基于KPCA及KPCA+CCA的情感识别有较好的效果。

关键词:语音信号;情感识别;情感特征;核主分量分析;典型相关分析

中图分类号:TP391.42 **文献标识码:**A

Speech Emotion Recognition Based on KPCA and CCA

Bian Jinhong¹, Wang Jilin¹, Yu Weifeng², Zhao Li²

(1. College of Information Engineering, Yancheng Institute of Technology, Yancheng, 224051, China;

2. School of Information Science and Engineering, Southeast University, Nanjing, 210096, China)

Abstract: Kernel principal component analysis (KPCA) is a learning method that uses the kernel method on principal component analysis (PCA), and the KPCA method can make the inseparable samples separable in the feature space. Canonical correlation analysis (CCA) is a statistical analysis of the correlation between the two sets of random variables. In this paper, the KPCA method is proposed for speech emotion recognition, and the KPCA combined with CCA method is also proposed for emotion recognition. Compared with the traditional PCA method, the results show that the emotion recognition based on KPCA and KPCA+CCA have better performance.

Key words: speech signal; emotion recognition; emotional features; kernel principal component analysis; canonical correlation analysis

引 言

通过语音传递信息是人类最重要的信息交换的方式。语音信号不仅有传递语义的作用,而且承载着说话人个性特征以及态度、情感等诸多信息。包含在语音信号中的情感是一种很重要的信息资源,随着计算机多媒体技术的发展,拟人的人机交互越来越受到人们的重视,因此,语音情感识别引起了广泛的关注^[1-6],通过计算机处理分析判断说话人的情感也成为重要的研究课题。然而由于情感信息的社会性、文化性,以及语音信号自身

的复杂性,语音情感识别尚有许多问题有待解决,如情感建模、特征的分析 and 选择、识别方法的改进等^[1-3]。

虽然关于语音情感的研究已经越来越被重视,但相关的研究成果还是比较少,且主要集中在语音情感特征分析和情感语音合成等方面。在情感特征的识别方面,利用重回归分析、主元素分析等多变量解析方法取得了87.1%以上的平均识别率^[2-3],利用神经网络技术进行情感识别的研究也在进行,但由于情感特征学习收敛性存在问题,并没有得到好的结果^[4]。近年来,又出现了许多语音情感识别方面的成果和方法,如主元分析法、最大

似然 Bayes 分类器和 K 最近邻分类器、人工神经网络^[4-5]、隐马尔科夫模型^[6-7]、支持向量机^[8-9]等。

核主分量分析(Kernel principal component analysis, KPCA)是一种利用核方法将主分量分析(PCA)推广后的学习方法,它能够使得输入空间线性不可分的样本在特征空间有更好的可分性。典型相关分析(Canonical correlation analysis, CCA)等价于最优化的 Fisher 鉴别分析,它是分析两组随机变量之间相关性的一种较好的统计方法。本文针对生气、欢快、中性和悲伤 4 种情感的语音信号,提取 111 个全局统计特征参数,提出了基于 KPCA 和 KPCA+CCA 的语音情感的识别方法。通过对实验结果进行比较分析,结果表明基于 KPCA 及 KPCA+CCA 的语音情感识别具有较好的识别效果。

1 语音情感特征的提取

1.1 语音情感数据库的选取

情感语音数据库的选取是进行语音识别研究的前提,根据数据获取途径,目前国际上语音情感研究人员所用的数据按照获取途径大致可分为 4 类^[9-10]:表演数据、激励数据、启发数据和摘引数据。本文选取的语音库为免费的柏林情感语音库,其采样频率为 16 kHz,16 bit 量化。该语音库共有 500 句情感语音信号,分别由 10 名专业演员(5 男,5 女)在不同情感状态下(高兴、愤怒、平静、悲伤、害怕、厌烦、憎恨)朗读 10 句不同文本的德语组成。本实验选取其中的部分情感(生气、欢快、中性和悲伤)进行实验。

1.2 语音情感特征提取

选取语音的特征对语音识别有着很大的影响,在过去的几十年里,针对语音信号中的何种特征能有效地体现情感,研究者从心理学、声学、语言学和语义等角度出发,作了大量的研究。目前主要分为韵律特征,情感识别所采用的特征大多是韵律特征,如基音、强度、持续时间;音质特征,如共振峰、MFCC、LPCC 等,韵律特征和音质特征并不是相互孤立的^[11]。根据情感的三维模型:“激励维(Arousal)—效价维(Valence)—控制维(Power)”,Tato 等人指出^[12],情感特征的提取不仅要考虑韵律特征参数,而且要考虑语音的音质特征参数。本文综合采用语音的韵律和音质特征,用于识别和建

模的特征向量一般有两种构造方法,全局统计特征和动态特征。由于动态特征对音位信息的依赖性太强^[10],不利于建立与文本无关的情感识别系统,因此在本文中使用了 111 个全局统计特征,如下所示:

特征 1~30:短时能量、短时幅度、短时过零率以及它们一阶差分的均值、最大值、最小值、中值、方差。

特征 31~46:基音及其一阶、二阶差分的均值、最大值、最小值、中值、方差,基音范围。

特征 47~61:第一共振峰及其一阶差分,第二共振峰的均值、最大值、最小值、中值、方差。

特征 62~111:MFCC 倒谱系数,差量倒谱系数的均值和方差。

2 基于核子空间分析的语音情感识别

2.1 核主分量分析

核方法由统计理论中得来,提供了一种从线性学习获得相应非线性学习算法的途径。该算法的思想为利用一非线性映射,把原空间语音信号的特征数据 R^d 映射到一隐特征空间 F 中: $\Phi: R^d \rightarrow F$, $x \rightarrow \Phi(x)$ 在核的特征空间进行通常的线性分析,在计算上,不需要知道具体的非线性变换函数 Φ ,核方法的特点就是利用核函数取代特征空间的内积运算。

$$k(x, y) = \Phi(x)^T \Phi(y) \quad (1)$$

常用的核函数有多项式核函数、径向基核函数、Sigmoid 核函数和复合核函数等,多项式核函数的形式为

$$k(x, y) = (x^T y + 1)^p \text{ 或 } k(x, y) = (x^T y)^p \quad (2)$$

本文选择了 2 阶的多项式核函数,一是因为多项式核函数比较简单,容易实现;二是与其他核函数相比识别效果相差不大。根据主分量分析的原理,求解在隐特征空间 F 的主元,等于求解如下特征值与特征向量的问题

$$S_\Phi w_\Phi = \lambda w_\Phi \quad (3)$$

式中 S_Φ 为非线性映射后的协方差矩阵。最后,原空间的数据 x 在 w_Φ 上的投影为

$$(x_{\text{KPCA}})_k = w_{\Phi k}^T \Phi(x) = \sum_{i=1}^N a_k^i k(x_i, x) \quad (4)$$

与 PCA 类似,基于 KPCA 的情感识别同样是将语音情感在特征向量进行投影,根据投影矢量间的距离进行分类。

2.2 典型相关分析(CCA)及 KPCA+CCA 方法

Fisher 线性鉴别分析(Fisher linear discriminant analysis, FLDA)求解最佳鉴别矢量,使得降维后样本特征类间散度最大且类内散度最小,设有 C 个类,第 i 个类有 n_i 个样本,第 i 个类的第 m 个样本记为 x_m^i ,类间散度阵 S_b ,类内散度阵 S_w 和总散度阵 S_t 定义为

$$S_b = \sum_{i=1}^C n_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{5}$$

$$S_w = \sum_{i=1}^C \sum_{m=1}^{n_i} (x_m^i - \mu)(x_m^i - \mu)^T \tag{6}$$

$$S_t = \sum_{i=1}^C \sum_{m=1}^{n_i} (x_m^i - \mu)(x_m^i - \mu)^T = S_b + S_w \tag{7}$$

定义 Fisher 准则

$$J(a) = \frac{a^T S_b a}{a^T S_w a} \tag{8}$$

求解最佳鉴别矢量 a 的问题等价于求解如下广义特征方程的特征值问题

$$S_b a_i = \lambda_i S_w a_i \tag{9}$$

将所有的鉴别矢量 a_k 构成变换矩阵 $W=[a_1, a_2, \dots, a_{C-1}]$,对原样本变换可得到低维的特征矢量 $y, y=W^T x$ 。但当 S_w 为奇异阵时,无法求解式(9),可以考虑用典型相关分析转换问题。

典型相关分析是分析两组随机变量的统计分析方法,其目的是寻找两组随机变量使得各自的线性组合,现实组合使得两个随机变量的相关性最大。设有两组均值为零的随机变量 $x=(c_1, c_2, \dots, c_k)^T$ 和 $y=(d_1, d_2, \dots, d_l)^T$,CCA 寻求矢量 a_1 和 b_1 使得 $u_1=x^T a_1$ 和 $v_1=y^T b_1$ 的相关系数 $r(u_1, v_1)$ 最大,并满足 $\text{var}(u_1)=\text{var}(v_1)=1$;寻找所有的矢量 a_i 和 b_i 使得 $r(u_i, v_i)$ 最大,并满足条件 $\text{var}(u_i)=\text{var}(v_i)=1$ 和 $\text{cov}(u_i, u_j)=\text{cov}(v_i, v_j)=0, i \neq j$,利用矢量对和相关系数可分析一组变量中的某些随机变量和另一组变量的某些随机关系。求解矢量 a 和 b 相当于如下

$$b = \frac{1}{r(u, v)} \Sigma_y^{-1} \Sigma_{yx} a \tag{10}$$

$$\Sigma_x^{-1} \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy} a = r^2(u, v) a = \lambda^{\text{CCA}} a \tag{11}$$

$$Y = \begin{bmatrix} I_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & I_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_C} & 0_{n_C} & \cdots & I_{n_C} \end{bmatrix} \tag{12}$$

式(12)为构造的编码矩阵 Y 与样本矩阵 X ,利用该矩阵估计协方差矩阵 S_x, S_y, S_{xy}

$$S_x = \frac{1}{n-1} X^T X, S_y = \frac{1}{n-1} Y^T Y, S_{xy} = \frac{1}{n-1} X^T Y \tag{13}$$

从而得到最佳鉴别矢量构成的变换阵 $A=(\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{C-1})$ 。KPCA+CCA 的语音情感识别步骤如下:

- (1) 对 n 个情感特征的训练样本进行 KPCA,得到核主分量特征;
- (2) 利用核主分量特征和式(12)构成矩阵 X_{KPCA} 和 Y ;
- (3) 利用式(13)估计协方差矩阵 S_x, S_y, S_{xy} ;
- (4) 由式(11)计算最佳鉴别矢量 $\tilde{a}_k$,得到变换阵 A ,得到 $x_{\text{CCA}}=x_{\text{KPCA}}A$;
- (5) 对待识别的样本 y 得到 y_{KPCA} ,并抽取最佳鉴别特征得到 $y_{\text{CCA}}=y_{\text{KPCA}}A$,利用最近邻作为归类,用 2 范数作为距离测度。

3 语音情感识别实验与分析

本文将用基于以下的三种分类器训练方法用于语音情感识别:(1)PCA (2)KPCA (3)KPCA+CCA。针对生气、欢快、中性和悲伤 4 种情感,所有的识别方法用矢量的 2 范数作为距离测度, Fisher 鉴别信息定义如式(14),其中 $\mu_{k,i}$ 和 $\sigma_{k,i}^2$ 分别是第 i 类样本在第 k 个分量上的均值和方差,本文提出一个改进方法是在计算距离时可以用该公式按重要性加入加权系数,用 K-近邻法作为归类的标准。

$$\text{FDI}_k = \sum_{i=1}^C \sum_{j=1}^C \frac{(\mu_{k,i} - \mu_{k,j})^2}{\sigma_{k,i}^2 + \sigma_{k,j}^2} \quad k=1,2,\dots,q \tag{14}$$

由图 1 可以得到以下结论:(1)KPCA 方法比 PCA 的效果要好,这是因为核方法能够使低维空间

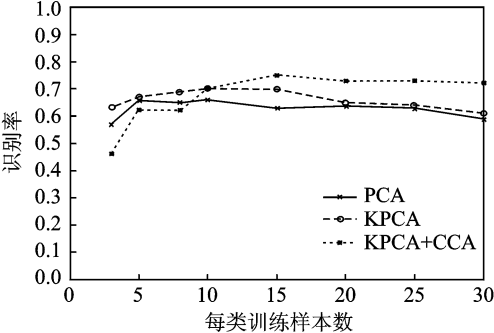


图 1 语音情感识别结果

Fig. 1 Results of speech emotion recognition

[4] Li Wu ,Zhang Yanhui ,Fu Yingzi. Speech emotion recognition in e-learning system based on affective computing[C]// Proc of the 3rd International Conference on Natural Computation. Washington DC : IEEE Computer Society,2007:809-813.

[5] Stibbard R M. Vocal expression of emotions in non-laboratory speech: An investigation of the reading/leeds emotion in speech porject annotation data[D]. UK;University of the Reading, 2001.

[6] Nwe T L, Foo S W , Silva L C D. Speech emotion recognition using hidden Markov models[J]. Speech Communication,2003,41(4):603-623.

[7] 张石清, 赵知劲, 戴育良, 等. 支持向量机应用于语音情感识别的研究[J]. 声学学报,2008,27(1):87-90.

Zhang Shiqing, Zhao Zhijin, Dai Yuliang, et al. A study of support vector machine for speech emotion recognition[J]. Technical Acoustics, 2008, 27(1): 87-90.

[8] Picard R W. Toward computers that recognize and respond to user emotion[J]. IBM Technical Journal, 2000, 38(2): 705-719.

[9] Zhiping Wang, Li Zhao, Cairong Zou. Support vector machines for emotion recognition in Chinese speech[J]. Journal of Southeast University ,2003,19(4): 307-310.

[10] 赵力,将春辉,邹采荣,等. 语音信号中的情感特征分析和识别的研究[J]. 电子学报,2004,32(4):606-609.

Zhao Li, Jiang Chunhui, Zou Cairong, et al. A study on emotional feature analysis and recognition in speech[J]. Acta Electronica Sinica, 2004, 32(4): 606-609.

[11] Ohala J J. Ethological theory and the expression of emotion in the voice[C]// International Conference on Spoken Language Processing. Philadelphia, USA: [s. n.], 1996:1812-1815.

[12] Tato R, Santos R, Kompe R, et al. Emotion space improves emotion recognition[C] // Proc IC SLP 2002. Denver, Colorado:[s. n.],2002:2029-2032.

作者简介:卞金洪(1966-),男,副教授,研究方向:信息与信号处理;王吉林(1969-),男,副教授,研究方向:语音信号处理等;余威风(1990-),男,博士研究生,研究方向:语音信号处理;赵力(1958-),男,教授,研究方向:语音信号处理,E-mail:zhaoli@seu.edu.cn。