

文章编号:1672-058X(2013)01-0062-07

基于 KPCA 和模糊核 Fisher 判别的语音情感情识别^{*}

邢玉娟, 李恒杰, 胡建军, 王万军

(甘肃联合大学 电子与信息工程学院, 甘肃 兰州 730000)

摘要:针对语音情感情识别率不高和实时性差的问题,提出一种基于 KPCA 核主成分空间的模糊 KFD 算法,应用于语音情感情识别。首先采用 KPCA 对语音情感特征向量降维去噪,根据转换矩阵得到核主成分空间,然后在该特征空间利用模糊 C 均值聚类计算语音特征向量的隶属度,进而对 LDA 算法中的类间离散度和类内离散度重新定义,生成模糊 KFD 分类器进行语音情感情识别。仿真实验结果表明,提出的方法相比于传统 SVM 和核 Fisher 判别算法具有较高的识别率和良好的抗噪性能,是一种行之有效的语音情感情识别新方法。

关键词:语音情感情识别; 模糊核 Fisher 判别; 核主成分分析; 模糊 C 均值聚类

中图分类号:TP391

文献标志码: A

语音情感情识别^[1]就是根据说话人的语音特征识别出说话人的情感状态(如愤怒、喜悦、惊讶、悲伤、害怕、厌恶等),包含语音情感特征参数的提取以及情感状态的判定,在信号处理和人工智能领域有着重要的研究意义。

在语音情感情识别中,为了获得理想的识别率,主要提取基音频率、短时能量、共振峰、MFCC 参数以及它们的派生参数作为语音的情感特征。特征向量的维数很高,并且高维特征向量往往也存在着一些对情感分类贡献小,甚至没有贡献的特征^[2]。采用这样的高维特征向量进行情感情识别,势必会产生识别率低和实时性差的问题。近几年来,基于核函数的机器学习方法如支持向量机法、核 Fisher 判别技术等以其出色的分类性能成为模式识别领域的研究热点。核 Fisher 判别(Kernel Fisher Discriminant, KFD)技术在分类时采用了所有的训练样本,而不是少量的称之为“支持向量”的样本,因此 KFD 的识别率在某种程度上优于支持向量机^[3]。然而 KFD 的求解难度随着样本数目的增加急剧提高,同时在 KFD 的计算中并没有考虑到样本对所属类别的相对重要性或对所属类别贡献的大小问题,并且语音特征向量中存在噪声点和野值向量,导致 KFD 训练速度慢且识别率不高。因此此处提出一种基于核主成分空间的(Kernel Principle Component Space, KPCS)模糊核 Fisher 判别(Fuzzy Kernel Fisher Discriminant, FKFD)语音情感情识别方法。采用核主成分分析(Kernel Principle Component Analysis, KPCA)和模糊 C 均值聚类(Fuzzy C-means Clustering, FCM)对特征向量进行降维去噪、聚类的同时,对不同的样本赋予不同的惩罚权系数,得到样本的隶属度,使得不同样本对 Fisher 目标函数的贡献不同,进而提高 FKFD 判别的准确率。

收稿日期:2012-06-05;修回日期:2012-07-04.

* 基金项目:甘肃省教育厅基金项目(1113-01).

作者简介:邢玉娟(1981-),女,甘肃天水人,讲师,硕士研究生,从事生物特征识别的研究.

1 基于 KPCA 核主成分空间的模糊核 Fisher 判别在语音情感识别中的应用

此处提出的语音情感识别方法的系统框图如图 1 所示。

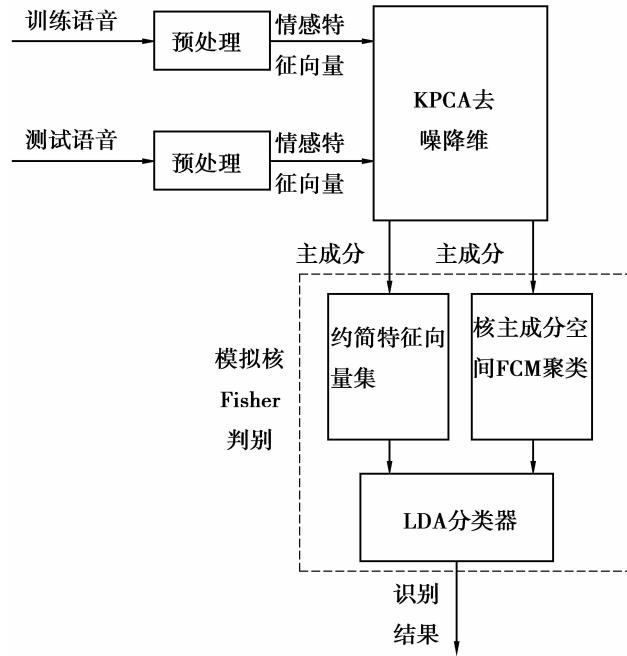


图 1 语音情感识别系统框图

1.1 KPCA 降维去噪

KPCA 的核心思想是通过核函数技巧, 将原始数据映射到高维特征空间, 然后在高维特征空间中利用主成分分析法(Principle Component Analysis, PCA)求出最佳投影方向, 从而获得非线性特征^[4]。

假设情感语音样本为 $\{x_1, x_2, \dots, x_N\}$, $x_i \in \mathbf{R}^m$, 通过非线性映射 $\Phi: x \in \mathbf{R}^m \rightarrow z \in \mathbf{R}^h$ 将样本映射到高维空间 H 。在该空间样本的协方差矩阵可以表示为:

$$C = \frac{1}{n} \sum_{i=1}^n \bar{\Phi}(x_i) \bar{\Phi}(x_i)^T \quad (1)$$

其中 $\bar{\Phi}(x_i) = \Phi(x_i) - \frac{1}{n} \sum_{i=1}^n \Phi(x_i)$ 。与 PCA 的原理类似,为了得到向量 x 的主分量,需要求解协方差矩阵 C

的非负特征值 λ 和特征向量 V 。由等式 $\lambda v = Cv$ 可得：

$$Cv = \left(\frac{1}{n} \sum_{i=1}^n \bar{\Phi}(x_i) \bar{\Phi}(x_i)^T \right) v = \frac{1}{n} \sum_{i=1}^n \langle \bar{\Phi}(x_i)^T, v \rangle \bar{\Phi}(x_i) \quad (2)$$

其中 $\langle x, v \rangle$ 代表 x 与 v 的点积。公式(2)意味着 $\lambda \neq 0$ 所对应的 v 都一定在 $\overline{\Phi}(x_1), \dots, \overline{\Phi}(x_n)$ 中，因此，存在系数 $\alpha_i (i = 1, \dots, n)$ 可以满足：

$$v = \sum_{i=1}^n \alpha_i \bar{\Phi}(x_i) \quad (3)$$

假设 K 为 $n \times n$ 的核矩阵: $[K]_{ij} = K_{ij} = \langle \bar{\Phi}(x_i), \bar{\Phi}(x_j) \rangle$, 将式(3)代入式(2)可得:

$$n\lambda\alpha = K\alpha \quad (4)$$

将式(4)得到的特征值由大到小排序, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$, 其对应的特征向量为 $\mu_1, \mu_2, \dots, \mu_n$ 。语音特征向量经过 KPCA 变换之后, 选择前 q 个特征分量构成核主成分空间: $P^{(s)} = \{\mu_1^{(s)}, \mu_2^{(s)}, \dots, \mu_q^{(s)}\}, s = 1, \dots, N$ 。

1.2 模糊 C 均值聚类

在核主成分空间 $P^{(s)}$ 中, 对 KPCA 降维去噪后得到的约简特征向量集, 采用模糊 C 均值聚类计算向量的隶属度。假设约简数据集 $X = \{x_i\}_{i=1}^N$, 其中 $x_i \in \mathbf{R}^q$, FCM 聚类主要是最小化目标函数:

$$J_f(D, U, C) = \sum_{j=1}^c \sum_{i=1}^n [u_j(x_i)]^b x_i - m_j^2 \quad (5)$$

其中 n 是训练样本的数目, c 表示指定的聚类数目。 $X = \{x_1, x_2, \dots, x_n\}$ 和 $M = \{m_1, m_2, \dots, m_c\}$ 分别表示训练样本集和聚类中心。模糊矩阵 $U = (u_j(x_i))_{c \times n}$ 由每一个训练样本 x_i 对于聚类 m_j 的模糊隶属度组成。同时, $b > 1$ 用于控制分类结果模糊量的指数权重。 $d_{ij}^2 = |x_i - m_j|^2$ 是样本 x_i 到聚类中心 m_j 的欧几里得距离。根据定义, 每个样本 x_i 满足条件:

$$\sum_{j=1}^c \mu_j(x_i) = 1, 0 \leq \mu_j(x_i) \leq 1 \quad (6)$$

在此条件下, 最小化目标函数。这是一种受限优化问题, 通过拉格朗日乘子法将其转换为一个非受限优化问题。

$$m_j = \frac{\sum_{i=1}^n [u_j(x_i)]^b x_i}{\sum_{i=1}^n [u_j(x_i)]^b}, j = 1, 2, \dots, c \quad (7)$$

$$u_j(x_i) = \frac{(1/x_i - m_j^2)^{\frac{1}{b-1}}}{\sum_{k=1}^c (1/x_i - m_k^2)^{\frac{1}{b-1}}}, i = 1, 2, \dots, n; j = 1, 2, \dots, c \quad (8)$$

FCM 算法的具体步骤如下:

Step1: 确定聚类的数目 $c, b = 2$, 给定非负的收敛误差 ε , 同时设定 $t = 0$ 用于记录聚类中心。

Step2: 随机初始化聚类中心 $m_j^t, j = 1, 2, \dots, c$ 和隶属度 $u_j(x_i)^t$ 。

Step3: 更新 m_j^t 和 $u_j(x_i)^t$

Step31: 计算新的 m_j^{t+1} 使用式 (7)。

Step32: 计算新的 $u_j(x_i)^{t+1}$ 使用式 (8)。

Step4: 如果 $\|U^t - U^{t-1}\| < \varepsilon$ 停止聚类, 否则转到 step3。

1.3 KFKD 最终判别

由于核主成分空间已经是高维线性空间, 因此在该空间 KFD 的核函数映射计算可以省略, 直接采用线性判别(Linear Discriminant Analysis, LDA)算法进行最终分类识别。LDA 核心问题是样本隶属度的确定以及如何在目标函数中引入隶属度。在目标函数中起决定作用的是类间离散度矩阵 S_b 和类内离散度矩阵 S_w 。因此, 提出在用 FCM 得到的模糊隶属度矩阵 $U = (u_{ij})_{c \times N}$ 和聚类中心矩阵 $M = [m_1, m_2, \dots, m_c]$ 对 S_b 和 S_w 进行重新定义^[5]。

类内离散度矩阵和类间离散度矩阵可重新定义为:

$$FS_w = \sum_{i=1}^c \left(\sum_{x_j \in C_i} u_{ij}^p (x_j - m_i) (x_j - m_i)^T \right) \quad (9)$$

$$FS_b = \frac{1}{N} \sum_{i=1}^c d_i \left(\frac{1 - \sum_{x_j \in C_i} u_{ij}^p}{\sum_{j=1}^n u_{ij}^p} \right) (m_i - \bar{m}) (m_i - \bar{m})^T \quad (10)$$

其中 p 是控制模糊量的指数权重, m_i 是第 i 类样本的聚类中心, \bar{m} 是样本的整体均值向量。将重新定义的 FS_b 和 FS_w 代入 Fisher 目标函数式可得:

$$J(w) = \frac{w^T FS_b w}{w^T FS_w w} \quad (11)$$

重新计算输入样本的最优投影方向 $w = S_w(m_1 - m_2)$, 最后使用分类判别函数式(12)即可得出识别结果。

$$f(x) = \text{sgn}(w^T x + b) \quad (12)$$

综上所述, FKFD 判别算法描述如下:

Step 1: 在 KPCA 的核主成分空间进行模糊 C - 均值聚类, 得到模糊隶属度矩阵 $U = (u_{ij})_{C \times N}$ 和类中心矩阵 $M = [m_1, m_2, \dots, m_C]$ 。

Step 2: 根据模糊隶属度矩阵 U 和类中心矩阵 M 重新计算 LDA 中的类内离散度矩阵 S_w 和类间离散度矩阵 S_b 。

Step 3: 将 FS_b 和 FS_w 代入目标函数计算样本的最优投影方向。

Step 4: 根据最优投影方向采用判别函数得出识别结果。

在语音情感识别中, 很明显识别模型是一个多元分类器, 在此参照支持向量机构造多类分类器的方法, 设计了一个二叉树 FKFD 多元分类器, 其结构如图 2 所示。

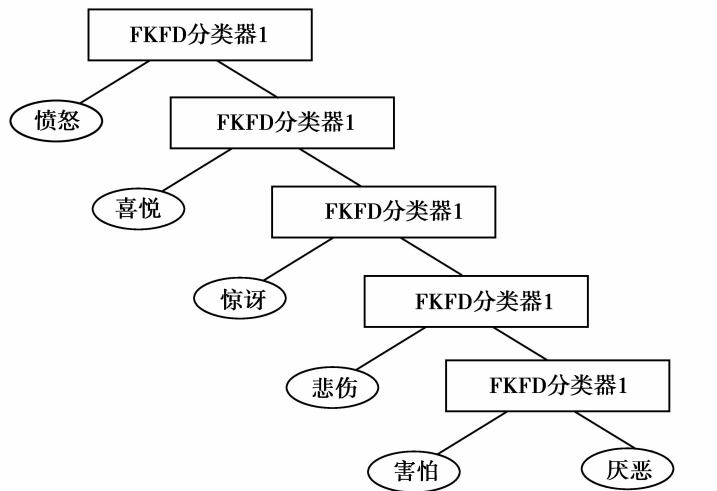


图 2 二叉树 FKFD 多元分类器

2 实验结果及分析

2.1 实验环境及语音情感数据库

仿真实验平台为 PC2.6G/1G, Windows XP2003 操作系统/Matlab7.0, 结合语音工具箱 Voicebox 以及 SVM Toolbox 1.0 验证语音情感识别的性能。核函数采用分类性能较好的 RBF 核函数($\sigma = 1.3$)。采用自己录制的语音数据, 录音软件采用 Cool Edit pro 2.0, 录音时采用单声道、11.025 kHz 采样频率、16 位采样精度, 录制语音保存为 PCM 编码的 wav 格式。录制 26 位说话人的 6 种情感状态语音: 愤怒、喜悦、惊讶、悲伤、害怕、厌恶。其中男性 13 名, 女性 13 名, 每人每种情感录制 10 条语句, 其中 5 句用于训练, 剩余 5 句用于测试。利用 Cool Edit 中的降噪器工具, 清除各种背景杂音。

2.2 情感特征提取

对语音库中每条语句进行端点检测、预加重, 以帧长 30 ms、帧移 15 ms 为语音信号加汉明窗, 窗长 $N = 200$ 。提取基音频率、第一共振峰、短时能量、MFCC(Mel-frequency Cepstral Coefficients)4 类基本情感声学特征及其派生特征参数作为语音情感特征向量。在仿真实验中, 逐帧计算语音的基音频率、第一共振峰和短

时能量的最大值、最小值、均值、范围、标准差、斜度、峰度、曲线抖动均值以及曲线抖动范围,各 9 维共 27 维特征向量。逐帧提取前 13 维 MFCC 参数。通过语音情感参数的提取,每位说话人的每条语句共可提取 40 维的特征向量。全部特征参数数据归一化到 [0,1]。

图 3 所示为语音库中一名男性语音“你可真伟大呀!”的喜悦和悲伤两种情感语音的短时能量和基音周期,其中基音周期采用短时自相关函数法检测。

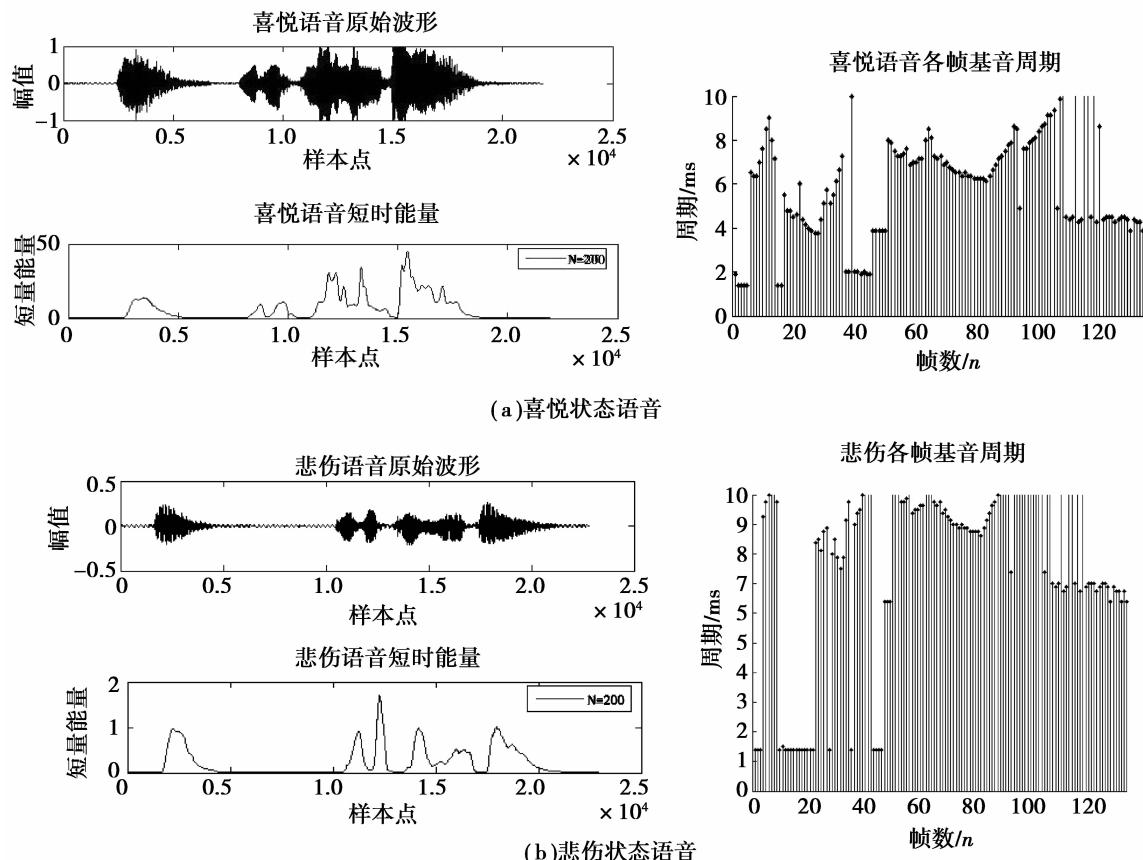


图 3 喜悦和悲伤两种情感语音短时能量和基音周期

2.3 实验结果

实验 1 识别性能实验。在实验中,此处提出的方法和传统的支持向量机法、核 Fisher 判别法进行比较分析,实验结果如表 1 所示。

表 1 情感识别率

识别方法	愤怒	喜悦	惊讶	悲伤	害怕	厌恶	平均识别率
SVM	82.11	80.28	81.62	78.21	77.01	74.81	79.01
KFD	84.67	84.39	83.91	80.33	79.24	75.59	81.35
KPCA + FKFD	89.29	86.14	88.61	82.42	80.26	79.47	84.36

由表 1 可知,在 3 种方法中,此处提出的方法在愤怒和惊讶状态识别率都有显著的提高。在愤怒状态,KPCA + FKFD 方法比 SVM 提高了 7.81%,比 KFD 方法提高了 4.62%;在惊讶状态,KPCA + FKFD 方法比 SVM 提高了 6.99%,比 KFD 方法提高了 4.7%。而 3 种方法的悲伤、害怕和厌恶的识别率都普遍不高,主要

是由于这3种情感在发音时,许多生理特征相似,较易混淆。

实验2 抗噪实验。为了测试提出方法的抗噪性能,采用 Matlab7.0 对测试语音进行加噪处理,产生带白噪声的语音测试库^[6]。选择的信噪比分别为:40 dB、35 dB、30 dB、25 dB、20 dB、15 dB、10 dB、5 dB。实验结果如表2所示。为了更好地描述不同信噪比下识别算法的情感平均正确识别率的变化趋势,绘制了情感平均识别率曲线图,如图4所示。

表2 不同信噪比下的情感平均正确识别率 %

识别方法	40 dB	35 dB	30 dB	25 dB	20 dB	15 dB	10 dB	5 dB
SVM	80.33	78.25	74.91	70.62	62.01	53.44	41.92	30.01
KFD	81.23	78.34	75.29	72.11	63.74	58.27	44.71	35.61
KPCA + FKFD	85.21	82.07	80.34	76.55	70.23	64.81	55.44	48.17

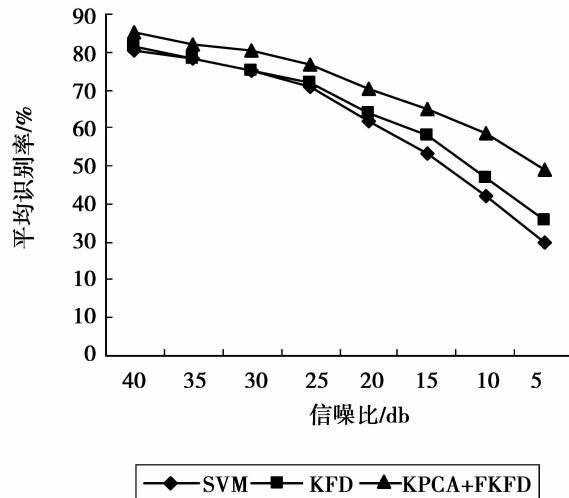


图4 不同信噪比下情感平均识别率曲线图

由表2和图4可知,在不同的信噪比实验中,随着信噪比的降低(噪声的增加),SVM和KFD分类器的情感平均正确识别率下降较快,而此处提出的方法(KPCA+FKFD)下降较慢,而且平均正确识别率均高于其他方法。即使当信噪比降低到5 dB时,此时语音信号中的噪声强度已经非常高,KPCA+FKFD的情感平均正确识别率也达到了48.17%,而其他两种的情感平均正确识别率降至20%左右。因此,在上述3种情感识别方法中,KPCA+FKFD具有最好的抗噪声性能。主要原因是KPCA在对语音数据降维的同时,也去除了语音向量中的冗余信息,这些冗余信息主要是一些噪声信息。

3 总 结

在深入分析研究语音情感识别算法的基础上,提出基于核主成分空间的模糊核 Fisher 判别算法。该算法一方面借助于 KPCA 和 FCM 对语音数据在维度和数量上进行了约简,可以有效地降低后续 LDA 的计算复杂度;另一方面,通过 KPCA 的转换矩阵得到核主成分空间,在该空间进行 FCM,对不同输入样本根据其贡献度引入隶属度,根据隶属度重新定义 LDA,不仅大大地简化了 KFD 的计算步骤,并且有效地提高了 KFD 的高分类性能。仿真实验结果表明,此处提出的方法和传统 SVM、KFD 相比,在纯净语音和噪声语音的情况下都取得了较为理想的情感平均正确识别率,具有良好的抗噪性能,是一种有效可行的语音情感识别算

法。然而,仿真实验也表明,悲伤、害怕和厌恶3种情感状态的识别率不高,因此在后续的研究工作中,将侧重于寻找更有效的参数提取方法,使得这3种情感状态识别率也能具有较高的识别率。

参考文献:

- [1] AYADI E I M. Survey on speech emotion recognition: Features, classification schemes, and databases[J]. Pattern Recognition, 2011, 44(3) :572-587
- [2] BITOUK D, VERMA R, NENKOVA A. Class-level spectral features for emotion recognition[J]. Speech Communication, 2010, 52(7/8) :613-625
- [3] WANG J H, LI Q, YOU J, et al. Fast kernel Fisher discriminant analysis via approximating the kernel principal component analysis[J]. Neurocomputing, 2011, 74(17) :3313-3322
- [4] 李童. 基于改进的2DPCA 人脸识别方法研究[J]. 重庆工商大学学报:自然科学版, 2012, 29(04) :45-49
- [5] ZHENG Y J, YANG J Y, WANG W D, et al. Fuzzy Kernel Fisher Discriminant Algorithm with Application to Face Recognition [A]. The proceedings of Sixth World Congress on Intelligent Control and Automation (WCICA2006)[C]. 中国, 大连:2006
- [6] 张石清, 赵知劲. 噪声背景下的语音情感识别[J]. 西南交通大学学报, 2009, 44(3) :442-447

Speech Emotion Recognition Based on KPCA and Fuzzy Fisher Discrimination

XING Yu-juan, LI Heng-jie, HU Jian-jun, WANG Wan-jun

(School of Electronics and Information Engineering, Gansu Lianhe University, Gansu Lanzhou 730000, China)

Abstract: A kind of novel fuzzy Kernel Fisher Discrimination (KFD) algorithm based on principal component space of Kernel Principle Component Analysis (KPCA) is proposed to improve speech emotion discrimination rate and real-time processing and is used in speech emotion discrimination. KPCA was firstly used to reduce dimensions of and remove noises for speech emotion feature vector, kernel principal component space was obtained according to transformation matrix, then in this feature space, fuzzy C-mean cluster was utilized to compute the membership of each speech feature vector, and furthermore, inter-class divergence factor and in-class divergence factor in LDA algorithm were redefined to form fuzzy KFD classifier to conduct speech emotion discrimination. Simulation experiment results show that the proposed method in this paper has higher recognition rate and better anti-noise performance than traditional SVM and kernel Fisher discrimination algorithm and is a kind of novel effective method for speech emotion discrimination.

Key words: speech emotion discrimination; fuzzy Kernel Fisher Discrimination; kernel principal component analysis;fuzzy C-mean clustering

责任编辑:李翠薇