

doi: 10.3969/j.issn.1001-0505.2015.01.002

基于 LDA + kernel-KNNFLC 的语音情感识别方法

张昕然^{1 2} 查 诚¹ 徐新洲¹ 宋 鹏³ 赵 力^{1 2}

(¹ 东南大学水声信号处理教育部重点实验室, 南京 210096)

(² 东南大学信息科学与工程学院, 南京 210096)

(³ 东南大学儿童发展与学习科学教育部重点实验室, 南京 210096)

摘要: 结合 K 近邻、核学习方法、特征线重心法和 LDA 算法, 提出了用于情感识别的 LDA + kernel-KNNFLC 方法. 首先针对先验样本特征造成的计算量庞大问题, 采用重心准则学习样本距离, 改进了核学习的 K 近邻方法; 然后加入 LDA 对情感特征向量进行优化, 在避免维度冗余的情况下, 更好地保证了情感信息识别的稳定性. 最后, 通过对特征空间再学习, 结合 LDA 的 kernel-KNNFLC 方法优化了情感特征向量的类间区分度, 适合于语音情感识别. 对包含 120 维全局统计特征的语音情感数据库进行仿真实验, 对降维方案、情感分类器和维度参数进行了多组对比分析. 结果表明, LDA + kernel-KNNFLC 方法在同等条件下性能提升效果最显著.

关键词: 语音情感识别; K 近邻; 核学习; 特征重心线; 线性判别分析

中图分类号: TP391.42 **文献标志码:** A **文章编号:** 1001-0505(2015)01-0005-07

Speech emotion recognition based on LDA + kernel-KNNFLC

Zhang Xinran^{1 2} Zha Cheng¹ Xu Xinzhou¹ Song Peng³ Zhao Li^{1 2}

(¹ Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing 210096, China)

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

(³ Key Laboratory of Child Development and Learning Science of Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: Based on KNN (K-nearest neighbor), kernel learning, FLC (feature line centroid) and LDA (linear discriminant analysis) algorithm, the LDA + kernel-KNNFLC method is put forward for emotion recognition according to the characteristics of the speech emotion features. First, in view of the large amount of calculation caused by the prior sample characteristics, the KNN of kernel learning method is improved by learning sample distances with the FLC. Secondly, by adding LDA to emotional feature vectors, the stability of emotional information recognition is ensured and dimensional redundancy is avoided. Finally, by the relearning of feature spaces, LDA + kernel-KNNFLC can optimize the degree of differentiation between emotional feature vectors, which is suitable for speech emotion recognition (SER). An emotional database is used for simulation tests, which contains 120 dimensional global statistical characteristics. Multiple comparison analysis is conducted through the dimension reduction scheme, emotion classifiers and dimension parameters. The results show that the improvement effect for SER by using LDA + kernel-KNNFLC is remarkable under the same conditions.

Key words: speech emotion recognition; K-nearest neighbor; kernel learning method; feature line centroid; linear discriminant analysis

由于语音情感识别 (speech emotion recognition, SER) 融合模式识别、机器学习以及语音信号等领域, 因而得到了广泛的研究. 相比情感特征提取和语音数据库分析领域的大量研究, 对语音情感

收稿日期: 2014-09-17. 作者简介: 张昕然 (1987—), 男, 博士生; 赵力 (联系人), 男, 博士, 教授, 博士生导师. zhaoli@seu.edu.cn.

基金项目: 国家自然科学基金资助项目 (61273266, 61231002, 61375028)、教育部博士点专项基金资助项目 (20110092130004).

引用本文: 张昕然, 查诚, 徐新洲, 等. 基于 LDA + kernel-KNNFLC 的语音情感识别方法[J]. 东南大学学报: 自然科学版, 2015, 45(1): 5-11.
[doi: 10.3969/j.issn.1001-0505.2015.01.002]

识别性能方面的关注度偏少。

近年来在情感计算领域,包括重回归分析和主元素分析等多变量解析方法神经网络技术,都取得了一定的进展^[1],但由于情感特征学习收敛性问题,还存在一定的局限性。另外,采用主元分析法、最大似然贝叶斯分类器和 K 最近邻分类器、人工神经网络、隐马尔科夫模型^[2-3]等技术的情感计算方法,也取得了一定的成果。

语音情感识别中的 K 近邻分类法(K-nearest neighbor, KNN)因其易于实现而被作为常用的分类器。但是其识别性能在很大程度上依赖于语音情感特征数据库的表征能力。因此,文献[4-5]提出最近特征线(nearest neighbor feature line centroid, NNFLC)的识别方法来扩展样本的表征能力。先提取出样本的低维特征,然后用同一个类中的 2 个样本特征点得到一条直线,用此直线来描述 2 个样本点间的特征变化,从而扩展样本的表征能力,取得了较好的识别结果。文献[6-7]提出的核学习方法,可将原空间的待分类样本映射到一个高维的特征空间(核空间)。该方法利用核学习思想和结合 K 近邻分类方法,通过非线性映射来突出不同类别的样本特征差异,使得原来线性不可分的样本点在核空间变得更加线性可分(或近似线性可分),从而提高了分类效果。

NNFLC 和核学习分类方法虽然在一定程度上提升了语音情感特征的识别性能,然而这些方法存在如下问题:①这 2 种分类方法用于语音情感识别,需要先抽取情感数据库的特征作为样本,如果数据量较大则运算效率会大幅降低,难以保证训练样本的有效性。② NNFLC 是针对线性变换得到的方法,但在实际语音情感数据中,单用线性变换难以全面地表征人类复杂的情感信息;核学习方法虽然利用高维空间映射保留了情感信息的非线性因子,但同时也加大了数据的冗余量,可能会导致计算量过大甚至“维度灾难”。

针对上述问题,本文提出了将核 K 近邻与线性判别分析(linear discriminant analysis, LDA)方法相结合,并采用重心准则学习样本距离,从而得到 LDA + 核 K 近邻重心(LDA + kernel-KNNFLC)方法。这种方法保留了非线性空间部分的语音情感信息,并同时避免了维度过高造成的计算量问题。其中, LDA 保证了更好的情感信息识别稳定性;而核学习通过再生核希尔伯特空间(RKHS)思想^[8],尽可能地保留了数据信息;同时,重心准则降低了距离计算的复杂度并解决了特征线失效问题。

<http://journal.seu.edu.cn>

1 基于核 K 近邻特征线重心分类方法

1.1 基于核学习 K 近邻语音情感识别方法

KNN 算法是最近邻算法的推广,其优点是分类速度快、简单易行且属于非参数分类算法,现已广泛应用于模式识别和数据挖掘等各个领域。然而其性能较大依赖于训练样本特征的表征能力。传统 K 近邻法简单地说就是取未知样本 X 的 K 个近邻,按照近邻数目所属类别把 x 归类^[9]。即:从 N 个已知样本中找出的 X 的 K 个近邻,若 K_1, K_2, \dots, K_c 分别是 K 个近邻中属于 $\omega_1, \omega_2, \dots, \omega_c$ 类的样本数,则可定义判别函数为

$$g_j(X) = K_i \quad i = 1, 2, \dots, c \quad (1)$$

若

$$g_j(X) = \max_i(K_i) \quad (2)$$

则得到决策 $X \in \omega_j$ 。

但语音情感数据保留高自然度和复杂语义属性,其样本边界往往存在线性不可分或分布为非高斯分布等情况, K 近邻法表现出的分类效果较差^[10]。为了提高分类的效果,克服 K 近邻法这一缺点,将统计学习理论中的核学习方法(kernel learning)与 K 近邻法相结合^[11],得到基于语音情感识别的 kernel-KNN 法。

kernel-KNN 分类算法的基本思想是:首先利用一个隐式的非线性映射 $\Phi(x)$ 将样本 x 从输入空间 R 映射到一个高维的特征空间 F 中,如

$$\Phi: R^n \rightarrow F, x \rightarrow \Phi(x) \quad (3)$$

映射的目的是突出不同类别样本之间的特征差异,使得样本在特征空间中变得线性可分(或近似线性可分),然后再在这个高维的特征空间中进行传统的 K 近邻分类。

核学习方法的一个重要特点就是利用核函数取代特征空间的内积运算,因此在计算时无需知道非线性映射函数 $\Phi(x)$ 的形式。样本 x, x' 映射至特征空间 F 中的点 $\Phi(x), \Phi(x')$ 的内积为

$$k(x, x') = \Phi(x) \cdot \Phi(x') \geq \Phi^T(x) \Phi(x') \quad (4)$$

式中 $k(x, x')$ 为核函数。

本文利用 KNN 思想提高语音情感识别效率的一个关键环节是选取核函数的类型。常用的核函数有多项式核函数: $k(x, x') = (x^T x' + 1)^d$, 其中 d 为多项式阶数;高斯径向基核函数(RBF)为 $k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$, 其中 σ 为核函数中心 σ 为函数的宽度参数,控制了函数的径向作用范围; $\|\cdot\|$ 为欧式距离;sigmoid 核 $K(x, x_i) =$

$\tanh[u(x, x_i) + r] \mu > 0 \quad r < 0$.

这些函数中应用最广的是RBF核,无论是小样本还是大样本,高维还是低维等,RBF核函数均适用.针对语音情感数据特征,RBF函数相比其他的函数有以下优点:

① RBF核函数可以将一个样本映射到一个更高维的空间,而且线性核函数是RBF的一个特例,而语音情感样本集就包含多种语义信息和特征性质,因此相比线性核函数,RBF更适用于语音情感识别.

② 与多项式核函数相比,RBF需要确定的参数要少,核函数参数的多少直接影响函数的复杂程度.另外,当多项式的阶数比较高时,语音情感样本核矩阵的元素值将趋于无穷大或无穷小,而RBF则具有相对固定的参数,会减少计算情感数据样本数值的困难.

③ 对于需要保留自然度和多种语义的样本参数,RBF比sigmoid具有更好的性能.其映射能力和非线性逼近能力均较强,更加适合大数据量的语音情感识别样本集.

1.2 基于核学习K近邻特征线重心法

先给出K近邻特征线重心法(KNNFLC).将每个样本特征看作样本特征空间中的一点,设有C个情感类别,第c类有 N_c 个语音情感特征样本.用第c情感类中的2个不同特征样本特征点 x_i^c 和 x_j^c 确定一条特征线 $\overline{x_i^c x_j^c}$ ($i \neq j$),用此直线来描述这2个样本点之间的基频、能量、共振峰等特征参数的变化.如图1所示,待识别情感样本特征点为 x_p ,首先计算出 p_{ij}^c 是特征线 $\overline{x_i^c x_j^c}$ 上的任意一点,待识别样本点 x_p 和 p_{ij}^c 之间的特征线距离体现出样本点与第c类情感类型的远近程度, p_{ij}^c 位置的选择直接关系到分类结果的好坏.

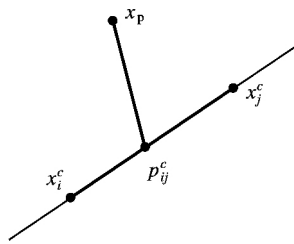


图1 特征线示意图

如果采用待识别样本点到特征线距离为最短距离即垂线距离,也就是 p_{ij}^c 的位置为垂足,那么有可能会出现特征线失效的情况.如图2所示,待识别样本特征 x_p 到情感样本点的距离和到特征线的

距离不一定是正比关系.这时 x_p 到 $\overline{x_i^c x_j^c}$ 的垂线距离比 x_p 到 $\overline{x_k^c x_l^c}$ 的垂线距离要近.但是 x_p 与 x_k^c 和 x_l^c 这2个样本点的距离较近,此时用特征线 $\overline{x_i^c x_j^c}$ 描述 x_p 的情感类型不如用特征线 $\overline{x_k^c x_l^c}$ 准确.

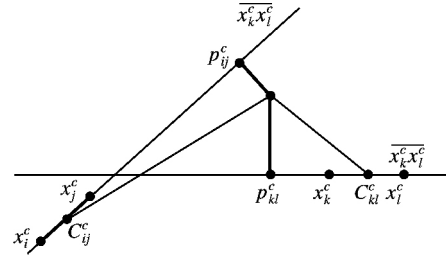


图2 特征线的失效和特征线重心方法示意图

计算情感样本点 x_p 到特征线的垂线距离较复杂,计算量很大,所以不宜采用垂线距离作为特征线距离.为了解决特征线失效,综合考虑各种因素,选择待识别情感样本点到特征线重心的距离作为特征线距离.从图2可以清晰看出,待识别情感样本 x_p 到 x_k^c 和 x_l^c 的中点 C_{kl}^c 的距离最短,而且 x_p 也距 x_k^c 和 x_l^c 最近.改进后的方案可以解决特征线可能失效的问题.2个样本点 x_k^c 和 x_l^c 的重心 C_{kl}^c 即为这2个样本点构成线段的中点,即 $C_{kl}^c = 1/2(x_k^c + x_l^c)$.

待识别情感样本特征 x_p 和特征线 $\overline{x_i^c x_j^c}$ 的距离 $l(x_p, \overline{x_i^c x_j^c}) = l(x_p, C_{ij}^c)$.在情感识别阶段,分别计算 x_p 和每一条特征线之间的距离,根据K近邻分类原则将 x_p 归入较小距离对应的特征线 $\overline{x_i^c x_j^c}$ 所属的情感类别c中.

下面进一步将KNNFLC法推广到非线性领域,即可得到基于核的K近邻特征线重心算法(kernel-KNNFLC).先将样本映射到非线性空间,任意2个样本 $\Phi(x_i^c)$ 和 $\Phi(x_j^c)$ 构成的特征线 $\overline{\Phi(x_i^c) \Phi(x_j^c)}$ 的重心为

$$C_{\Phi_{ij}}^c = \frac{1}{2} [\Phi(x_i^c) + \Phi(x_j^c)] \quad (5)$$

则待识别情感样本 $\Phi(x)$ 到特征线 $\overline{\Phi(x_i^c) \Phi(x_j^c)}$ 的距离为

$$d(\Phi(x_p), \overline{\Phi(x_i^c) \Phi(x_j^c)})^2 = d(\Phi(x_p), C_{\Phi_{ij}}^c)^2 =$$

$$\|\Phi(x_p) - C_{\Phi_{ij}}^c\|^2 = \frac{1}{4}k(x_i^c, x_i^c) +$$

$$\frac{1}{4}k(x_j^c, x_j^c) + \frac{1}{2}k(x_i^c, x_j^c) -$$

$$k(x_p, x_i^c) - k(x_p, x_j^c) + k(x_p, x_p) \quad (6)$$

识别情感时,根据式(6)分别求出在高维特征空间中待识别样本点和每个特征线重心的距离,然

后依据 K 近邻分类原则将待识别样本点归入较小距离对应的特征线 $\Phi(x_i^c)$ $\Phi(x_j^c)$ 所属的情感类别 c 中.

1.3 基于 LDA 的 kernel-KNNFLC 语音情感识别方法

LDA 的基本思想是将高维的模式样本投影到最佳判别矢量空间,以达到抽取分类信息和压缩特征空间维数的效果.投影后保证模式样本在新的子空间有最大的类间距离和最小的类内距离.利用 LDA 在该空间中有最佳的分离性,将语音情感数据样本类内离散度矩阵定义为

$$S_b = \sum_{i=1}^C P_i (\mu_i - \mu) (\mu_i - \mu)^T \quad (7)$$

式中 μ_i 为 C_i 类情感的均值; μ 为所有情感样本的均值; P_i 为先验概率.

情感样本类内离散度矩阵定义为

$$S_w = \sum_{i=1}^C P_i S_i \quad (8)$$

$$S_i = E[(x - \mu_i)(x - \mu_i)^T | x \in C_i] \quad (9)$$

情感样本类内离散度越大越好,而情感样本类内离散度越小越好.因此如果 S_w 是非奇异矩阵,最优的投影 W_{opt} 就是使得情感样本类内离散度矩阵和情感样本类内离散度矩阵的行列式比值最大的正交特征向量^[12].因此, Fisher 准则函数定义为

$$J(W_{opt}) = \arg \max_w \frac{|W^T S_b W|}{|W^T S_w W|} \quad (10)$$

据此,可以将该问题转化为求 W_{opt} 满足如下等式的最优解问题:

$$S_b W_i = \lambda_i S_w W_i \quad (11)$$

W_{opt} 也是对应于矩阵 $S_w^{-1} S_b$ 较大的特征值 λ_i 的特征向量.该矩阵最多只有 $C-1$ 个非零特征值, C 为情感类别数目.将所有的最优投影 W_{opt} 构成变换矩阵 $W = [W_1, W_2, \dots, W_{C-1}]$. 对原情感样本 x 变换可得到低维的特征矢量 y

$$y = W^T x \quad (12)$$

在使用核函数将情感样本映射到高维的特征空间时,其优点是能够使得在原空间中不易线性可分的数据在高维的特征空间变得线性可分(或近似线性可分).但是由于情感样本被核函数映射到高维空间,使得特征的维数增加,影响识别计算的速度,甚至造成“维数灾难”.同时,本文在仿真实验中提取了 120 维语音情感特征,这些特征不一定都是最有效的,所以需要进行特征选择.而线性判别分析(LDA)可以有效地进行特征提取,在降低

特征维数的同时提高不同类别情感特征之间的区分度.

为验证结合 LDA 识别方案的区分效果,对语音情感库中的愤怒、烦躁、高兴 3 种情感特征进行 LDA 线性判别分析.样本特征降维至 2 维,经 LDA 变化后各样本特征在新特征空间的分布如图 3 所示.由图可见,经过 LDA 变换后情感特征之间区分度增大,聚类效果明显,进而可以针对 5 类情感数据进行进一步的实验设计.

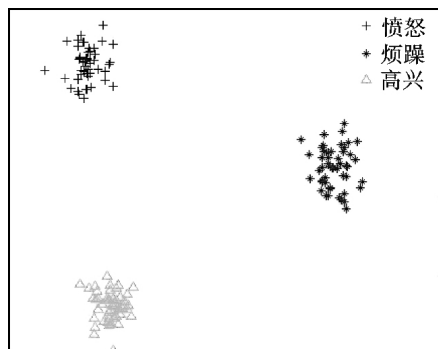


图 3 LDA 变换后样本在特征空间分布图

本文提出将 LDA 与核 KNNFLC 结合用于语音情感识别,在识别时先对 C 类语音情感特征进行 LDA 线性判别分析,将情感特征降维至 $C-1$ 维,再使用核学习的 KNNFLC 方法进行分类识别. LDA + kernel-KNNFLC 训练识别步骤如下:

① 对 C 类训练样本提取出 n 维情感特征,进行 LDA 线性判别分析,使用式(11)计算出变换矩阵 W ,利用式(12)将 n 维情感特征降维至 $C-1$ 维,构成样本库.

② 从待识别样本中提取 n 维情感特征 x ,由式(12)将 n 维特征降维至 $C-1$ 维特征 y .

③ 利用式(6)在高维空间中计算出特征 y 与样本库中每个特征线重心的距离 d .

④ 依据 K 近邻法则将待识别样本归类.

本文提出的 LDA + kernel-KNNFLC 语音情感识别流程如图 4 所示.

2 实验结果

实验在实验室录制的语音情感库上进行^[13].使用了 120 个全局统计特征,构成用于识别的情感特征向量.

1) 特征 1~4. 短时能量最大值、最小值、均值、方差.

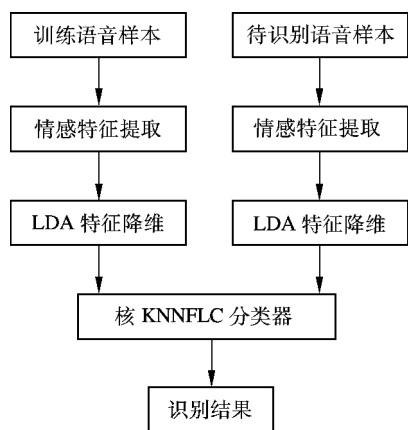


图4 语音情感识别流程图

2) 特征 5 ~ 8. 基音最大值、最小值、均值、方差.

3) 特征 9 ~ 12. 基音一阶差分最大值、最小值、均值、方差.

4) 特征 13 ~ 16. 第一、第二、第三共振峰最大值、最小值、均值、方差.

5) 特征 17 ~ 68. 梅尔倒谱系数 MFCC1 ~ MFCC13 最大值、最小值、均值、方差.

6) 特征 69 ~ 120. 梅尔倒谱系数 MFCC1 ~ MFCC13 一阶差分最大值、最小值、均值、方差.

用于实验的样本集包括了愤怒、烦躁、高兴、悲伤和平静 5 种情感的语句各 800 条. 其中每种情感选 600 条作为训练样本, 另外 200 条作为待识别样本, 而且训练样本和待识别样本中, 男女声音样本比例基本为 1:1. 提取出语音情感的 120 维特征, 采用全局统计特征原理构成语音情感特征的向量. 仿真实验环境为 Inter Core i5-2300 2.8 GHz 4 GB 内存, Windows 7 操作系统, 编程环境为 Matlab R2011b.

本文分别对 KNN, LDA + KNN 和 LDA + kernel-KNNFLC 算法进行了语音情感识别仿真实验, 并同时分别使用主成分分析 (principal component analysis, PCA) 和朴素贝叶斯分类 (naive Bayesian classification) 对降维部分和情感识别分类部分进行对比实验, 测试本文提出的语音情感识别方案的性能.

2.1 3 种 KNN 分类方案情感识别实验分析

针对 KNN 算法所构建的仿真实验中基于核的非线性分类器均是直接对 120 维的语音情感矢量进行识别的, 而线性分类器均先用 LDA 抽取主分量特征, 所选最优投影 W_{opt} 使 LDA 的重构误差为 4%. 为消除单次选择样本的随机性, 每类随机选取 200 个样本共 1 000 个训练样本, 剩下的 3 000

个作为测试样本, 重复 10 次取平均识别率. 表 1 为 K 近邻、LDA + K 近邻以及 LDA + kernel-KNNFLC 三种方案针对 5 类情感类别的识别率和平均识别率.

表1 3 种语音情感识别方案的识别率

情感类别	识别率/%		
	KNN	LDA + KNN	LDA + kernel-KNNFLC
愤怒	35.5	56.5	69.5
烦躁	79.5	87.0	96.5
高兴	36.5	61.5	68.0
悲伤	65.5	87.5	95.5
平静	57.0	63.5	81.0
平均值	54.8	71.2	82.1

从表 1 中可以看出, KNN 方法对 5 种情感的识别率普遍较低, 平均识别率不到 60%. 2 类情感识别率只有 36% 左右, 效果很不理想. LDA + KNN 方法通过对原始特征进行 LDA 变换, 提高了样本特征之间的区分度, 识别效果相对于 KNN 方法有了较大的提高, 平均识别率超过 70%. LDA + kernel-KNNFLC 方法在 LDA 变换的基础上, 利用核学习思想改进了传统的 KNN 方法, 通过引入特征线重心提高了样本特征的代表能力, 平均识别率达到了 82.1%.

2.2 2 种 kernel-KNNFLC 降维方案情感识别实验分析

采用 PCA 方法替换 LDA 对语音情感数据样本进行降维, 实验模型采用与上述实验相同的结构, 并设置降维后特征空间维数为 5, K 近邻中临近加权参数设 $K=3$ (维度及加权参数的实验分析在 LDA + kernel-KNNFLC 参数对比实验中进行). 对比实验结果如图 5 所示.

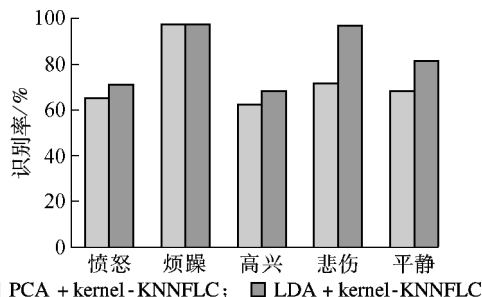


图5 2 种方案对 5 类情感降维识别对比图

由图 5 中对比实验可以观察到 2 种降维方法均采用核学习的 K 近邻特征线重心分类方法, 与 KNN 和 LDA + KNN 方法相比, 正确识别率有了明显的提升. “烦躁”、“悲伤”和“平静”3 种情绪的识别率在 70% 以上, 其中“烦躁”达到了 95% 以上. 对比 2 种降维方案, 在“愤怒”、“高兴”、“悲伤”

<http://journal.seu.edu.cn>

和“平静”4个情绪类别中,LDA + kernel-KNNFLC 方案相比其对照组,语音情感识别率分别提高了 4.5%、6.5%、25% 和 13%。这是由于 PCA 特征空间的降维优化是以区分 5 类情感为准则的,因此在同一个特征空间中,就很难兼顾到所有 5 类语音情感的最佳情感特征。特别地,从图 5 中可以看出,PCA 对“悲伤”的情感特征的优化并不理想。而本文提出的 LDA + kernel-KNNFLC 方案,利用特征向量间的离散度加权,使得情感特征类间区分度极大化,增加了对语音情感样本集的认识能力。

2.3 kernel-KNNFLC 与 NB 和 GMM 分类器对比实验分析

本组实验对朴素贝叶斯(naive Bayes, NB)和高斯混合模型(GMM)2种常用的分类器进行实验仿真,识别测试中将 LDA 降维后与 NB、GMM 相结合,分析 3 种不同分类方案的情感识别性能。

表 2 3 种语音情感识别分类器的识别率

情感类别	识别率/%		
	LDA + NB	LDA + GMM	LDA + kernel-KNNFLC
愤怒	56.5	60.0	69.5
烦躁	79.0	74.5	96.5
高兴	54.5	68.5	68.0
悲伤	81.0	71.5	95.5
平静	51.0	65.5	81.0
平均值	64.4	68.0	82.1

从表 2 可以看出,引入 LDA 降维后,LDA + NB 和 LDA + GMM 分类器比 KNN 独立识别语音情感,识别率有了一定的提高,但与 LDA + kernel-KNNFLC 方案相比,还有较大差距。这是由于 NB 算法有一定的局限性,其假定了特征向量之间是独立分布的,而在语音情感特征空间中的几个特征向量显然不是独立分布的,因而得到的 LDA + NB 分类器性能就比较有限。而 GMM 是一种单状态的混合连续分布隐马尔可夫模型,能够融合含有不同情感的语音在发音时的声学特性和时间上的变动^[14]。然而在选用 GMM 模型进行建模时一般需要比较多的训练数据,当训练数据较少时会影响 GMM 模型的学习性能^[15]。因此相比 kernel-KNNFLC 方法,其语音情感识别稳定性较差。

2.4 LDA + kernel-KNNFLC 方案中参数对比实验分析

首先针对 K 近邻的参数中 K 的不同取值进行仿真实验,以确定达到适合本文语音情感数据库的最佳参数。其中,当 K=1 时,表示采用最近邻的原则进行情感样本识别。采用的高斯径向基核函数为 $k(x, y) = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$,其中 $\sigma = 4$ (由交叉

验证的方法确定出参数)。每类选 600 个训练样本,剩下的 200 个作为测试样本,识别率为 5 种语音情感的平均识别率。

由图 6(a) 可看出,结合 LDA 降维算法改进后,基于核的 KNNFLC 分类器对语音情感识别性能比较稳定,并在维度为 4 和 5 时达到最佳效果。这是由于求解广义特征值原理中,求解最小特征数目能达到最优化的目的^[16]。本文实验采用 5 类语音情感来进行,因此维度降至 4 和 5,即可达到最佳的识别效果。从图 6(b) 可以看出,用最近特征线重心方法改进的 RBF 核函数,在近邻参数 K=3 时效果达到最佳。

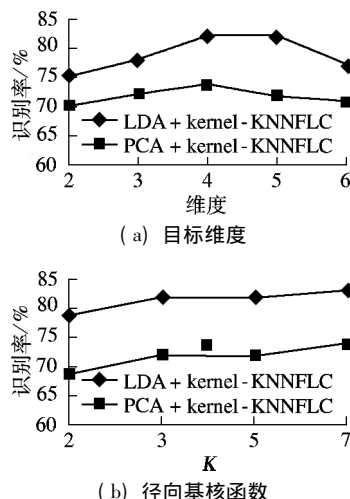


图 6 kernel-KNNFLC 算法参数变化时 2 种方案情感识别率对比图

3 结语

本文在 LDA + KNNFLC 方法核学习的基础上,结合 LDA 降维方法,同时引入了特征线重心对 KNN 算法进行改进,扩展了样本特征的代表能力,进而提高了分类识别能力。LDA 的引入解决了由于核空间维数较高,易造成“维数灾难”影响识别计算速度的问题,提高了不同类别特征之间的区分度。

由对比实验可以看出,结合 LDA 的核学习 K 近邻特征线重心的语音情感识别方法,相比其他分类方案具有更好的识别效果;同时,针对 5 种情感样本集,相较于其他降维方法也具有更稳定的工作性能。分别针对语音情感识别中,预先抽取样本造成的训练数据量过大问题,以及非线性映射和数据冗余的平衡问题进行的方案改进,相比原有几种识别方案改进后的方案体现出明显的优势。

另外根据本文进行的多组仿真对比实验,可以

观察出这 5 种情感中,“烦躁”情感识别率在各个识别方案中都是最高的,而“愤怒”和“高兴”较差,原因主要是后 2 类情感在发音时,许多生理特征相似,较易混淆,同时本文提取的情感特征对于“烦躁”情感表征更准确.这也说明了在后续的研究工作中,情感特征的优化还有一段提升的空间.

参考文献 (References)

- [1] Scherer K R. Vocal communication of emotion: a review of research paradigms [J]. *Speech Communication*, 2003, **40**(1/2): 227-256.
- [2] Scherer K R, Mortillaro M, Mehu M. Understanding the mechanisms underlying the production of facial expression of emotion: a componential perspective [J]. *Emotion Review*, 2013, **5**(1): 47-53.
- [3] Lin J C, Wu C H, Wei W L. Error weighted semi-coupled hidden Markov model for audio-visual emotion recognition [J]. *IEEE Transactions on Multimedia*, 2012, **14**(1): 142-156.
- [4] Li S Z, Lu J. Face recognition using the nearest feature line method [J]. *IEEE Transactions on Neural Networks*, 1999, **10**(2): 439-443.
- [5] Li S Z. Content-based audio classification and retrieval using the nearest feature line method [J]. *IEEE Transactions on Speech and Audio Processing*, 2000, **8**(5): 619-625.
- [6] Scholkopf B, Smola A, Muller K. Non-linear component analysis as a kernel eigenvalue problem [J]. *Neural Network*, 1999, **9**(4): 1299-1319.
- [7] Muller K, Mika S, Ratsch G, et al. An introduction to kernel-based learning algorithms [J]. *IEEE Transactions on Neural Networks*, 2001, **12**(2): 181-201.
- [8] Jung A, Schmutzhard S, Hlawatsch F. The RKHS approach to minimum variance estimation revisited: variance bounds, sufficient statistics, and exponential families [J]. *IEEE Transactions on Information Theory*, 2014, **60**(7): 4050-4065.
- [9] Carmona P L, Sánchez J S, Fred A L N. *Mathematical methodologies in pattern recognition and machine learning* [M]. New York: Springer, 2013: 101-107.
- [10] Wu Chung-Hsien, Liang Wei-Bin. Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels [J]. *IEEE Transactions on Affective Computing*, 2011, **2**(1): 10-21.
- [11] Zeng Hong, Cheung Yiu-ming. Feature selection and kernel learning for local learning-based clustering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8): 1532-1547.
- [12] Yan Shuicheng, Xu Dong, Zhang Benyu, et al. Graph embedding and extensions: a general framework for dimensionality reduction [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(1): 40-51.
- [13] 黄程韦, 赵艳, 金赞 等. 实用语音情感的特征分析与识别的研究 [J]. 电子与信息学报, 2011, **33**(1): 112-116.
Huang Chengwei, Zhao Yan, Jin Yun, et al. A study on feature analysis and recognition of practical speech emotion [J]. *Journal of Electronics & Information Technology*, 2011, **33**(1): 112-116. (in Chinese)
- [14] Dileep A D, Sekhar C C. GMM-based intermediate matching kernel for classification of varying length patterns of long duration speech using support vector machines [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(8): 1421-1432.
- [15] Wu Chung-Hsien, Wei Wen-Li, Lin Jen-Chun, et al. Speaking effect removal on emotion recognition from facial expressions based on eigenface conversion [J]. *IEEE Transactions on Multimedia*, 2013, **15**(8): 1732-1744.
- [16] Satapathy S C, Udgata S K, Biswal B N, et al. *Speech emotion recognition using regularized discriminant analysis* [M]. Bhubaneswar, Switzerland: Springer International Publishing, 2014: 363-369.