文章编号:1004-9037(2014)05-0730-05

基于改进多核学习的语音情感识别算法

奚 吉[1] 赵 力[2] 左加阔[2]

(1. 常州工学院计算机信息工程学院,常州,213002; 2. 东南大学信息科学与工程学院,南京,210096)

摘要:提出一种基于改进多核学习的语音情感识别算法。算法以高斯径向基核函数为基准,通过采样不同的样本,采用不同的评价标准并获得不同的参数,来提高分类性能。此外,通过引入多核技术,将得到的高斯核函数构建多核学习的基核,并通过利用松弛因子构建的软间隔多核学习的目标函数改善了学习效率。对比仿真实验结果表明,本文提出的基于多核学习语音情感识别算法有效提高了语音情感识别性能。

关键词:语音情感识别;多核学习;支持向量机

中图分类号:TN912.3 **文献标志码**:A

Speech Emotion Recognition Based on Modified Multiple Kernel Learning Algorithm

Xi Ji[1], *Zhao Li*[2], *Zuo Jiakuo*[2]

(1. School of Computer Information and Engineering, Changzhou Institute of Technology, Changzhou, 213002, China; 2. School of Information Science and Engineering, Southeast University, Nanjing, 210096, China)

Abstract: An improved algorithm of speech emotion recognition is proposed based on modified multiple kernel learning. The algorithm based on Gaussian radial basis function improves classification performance by gathering different samples, utilizing different evaluation criteria and acquiring different parameters . In addition, with the help of multiple kernel technology, the trained Gauss kernel functions are used to construct the basis of multiple kernel learning and the efficiency of learning are greatly improved by utilizing the relaxation factor to construct objective function of soft margin . Experimental results show the presented algorithm can effectively improve the performance of speech emotion recognition.

Key words: speech emotion recognition; multiple kernel learning; support vector machine (SVM)

引 言

语音情感识别方面的研究工作近年来得到了较大的发展,目前在很多领域内得到了广泛的应用[1]。比如在人机交互过程中,机器需要对交流对象的情绪进行分析,并由情绪分析的结果给出相应的反应,从而实现机器的智能化操作。在航天、潜艇、极地等极限环境条件下,语音信号相对于图像、视频信号,具有传输所需带宽较小的特点[2]。而极限环境下工作人员的情绪状态产生波动时可能会对信号的传输产生较大的负面影响,所以极限环境下需要使用语音情感识别措施。此外,在儿童孤独症的治疗研究等领域,语音情感识别也有较大的研究价值。在汉语语音情感识别方面,同样有着一些研究成果,但由于数据资源相对缺乏,没有公认的标准数据库,各方面的研究都处于初步阶段[3]。

核方法在模式识别领域内已得到了广泛的应用,核方法的一个改进方法思想是采用了基于多核学习的技术,为了优化减少核方法的计算量,Lanckriet[4-5]等人提出了核函数是一组预定义基核的线性组合,支持向量机的训练和基核的组合系

基金项目:国家自然科学基金(61375028,61301219)资助项目;江苏省自然科学基金(BK20130241)资助项目。

收稿日期:2014-04-12;**修订日期**:2014-06-18

数的训练同步完成，多核学习（Multiple kernel learning，MKL），核矩阵通过转换的得到，并且要求训练和测试样本都满足半正定的条件。同样的要求下优化的算法必须能够有较低的计算复杂度，同时在优化中需要限定条件：每个线性组合系数都满足非负；组合系数的和为 1。

本文将结合两种方法对基于核的语音情感识别进行优化。首先，基于单核的思想，对高斯径向核函数的参数进行优化，得到较为合理的核函数，用于多核学习中基核的构造；进而采用多核学习的思想，对于语音情感分类器进行优化，得到合适的核矩阵的组合，用在语音情感识别中；最后，用上述改进的算法和语音情感中一些常用方法进行对比，以验证本文算法的有效性。

1　基于训练的高斯核函数

核方法是通过一个非线性的变换 $\Phi: R^d \to H$，将样本映射到高维空间，通常经过非线性映射后变到希尔伯特空间进行线性分析。对于支持向量机（Support vector machine，SVM）来说，映射后的样本集为 $T=\{\Phi(x_i), y_i\}$；对于核方法来说，希尔伯特空间比原始空间维数更高，甚至接近无穷维。根据核函数的理论[6]，在空间的内积运算都转化为核函数为 $k(\boldsymbol{X},\boldsymbol{Y})=\Phi(\boldsymbol{X})^{\mathrm{T}}\Phi(\boldsymbol{Y})$，其中高斯径向核函数应用得最为广泛而且相关的研究也较多[7]

$$k(\boldsymbol{X},\boldsymbol{Y})=\exp\left(-\frac{\|\boldsymbol{X}-\boldsymbol{Y}\|^2}{\sigma^2}\right) \tag{1}$$

式中：σ 决定了非线性映射的几何结构，并能够控制核函数的适应性能，如果 σ 取得过小，适应范围能力将会变广，但是容易出现过拟合的情况，如果取得过大将使得核函数接近一个常量函数，使得其无法应用于非平凡的分类问题，因此选择一个合适的参数值，有利于整个系统的分类，本文优先考虑优化高斯核函数的参数选取。这里首先介绍优化的目标准则，即核目标排列（Kernel target alignment，KTA）准则[8]，估计 KTA 全局和局部的极值特性等。

给定核矩阵 $\boldsymbol{K}: \boldsymbol{X}^2 \to [-1,+1]$，以及标识向量 $\boldsymbol{Y}\in\{-1,1\}^m$，则根据 KTA 准则，核和标识向量之间可以写为

$$\boldsymbol{A}(\boldsymbol{K},\boldsymbol{Y})=\frac{\langle \boldsymbol{K},\boldsymbol{YY}'\rangle_{\mathrm{F}}}{\sqrt{\langle \boldsymbol{K},\boldsymbol{K}\rangle_{\mathrm{F}}\langle \boldsymbol{YY}',\boldsymbol{YY}'\rangle_{\mathrm{F}}}} \tag{2}$$

若有矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$，则上述的 F 范数为 $\langle \boldsymbol{P},\boldsymbol{Q}\rangle_{\mathrm{F}}=\sum_{i,j}p_{ij}q_{ij}=\mathrm{Tr}(\boldsymbol{PQ})$，$\boldsymbol{A}(\boldsymbol{K},\boldsymbol{Y})$ 能够集群化的度量，而 $\langle \boldsymbol{K},\boldsymbol{YY}'\rangle_{\mathrm{F}}=\sum_{y_i=y_j}k(x_i,x_j)-\sum_{y_i\neq y_j}k(x_i,x_j)$ 类内和类间的距离。KTA 准则能够给出集中在期望值附近的点，并且给出的核矩阵接近正确值的概率也更大，同时根据准则可以预测使用该核矩阵可能带来的错误概率，$\boldsymbol{A}(\boldsymbol{K},\boldsymbol{Y})$ 的值越大表明选择的核矩阵对于所给定的分类目标越适宜。

因为 $\langle \boldsymbol{K},\boldsymbol{YY}'\rangle_{\mathrm{F}}$ 表示了类内和类间的距离，因此可以定义可分离的评判标准

$$J=\langle \boldsymbol{K},\boldsymbol{YY}'\rangle_{\mathrm{F}}=\sum_{y_i=y_j}k(x_i,x_j)-\sum_{y_i\neq y_j}k(x_i,x_j) \tag{3}$$

对于优化的高斯径向基核，则可以将该准则改写成为

$$J(\sigma)=\sum_{y_i=y_j}\exp\left(-\frac{|x_i-x_j|^2}{\sigma^2}\right)-\sum_{y_i\neq y_j}\exp\left(-\frac{|x_i-x_j|^2}{\sigma^2}\right) \tag{4}$$

需要求解的目标函数是使得上式最大化，设函数 $R(\sigma)=-J(\sigma)$，因此需要求出使得 $R(\sigma)$ 最小的 σ

$$\sigma_{\mathrm{opt}}=\arg\min_{\sigma}R(\sigma)=\sum_{y_i\neq y_j}\exp\left(-\frac{|x_i-x_j|^2}{\sigma^2}\right)-\sum_{y_i=y_j}\exp\left(-\frac{|x_i-x_j|^2}{\sigma^2}\right) \tag{5}$$

为了得到全局的最小的 σ_{opt}，可以使 $\frac{R(\sigma)}{\sigma}=0$，求解该式得到最适合的参数；在求解 $\frac{R(\sigma)}{\sigma}=0$ 时，由于求解出非线性方程的解非常困难，故可采用经典的牛顿法计算，这是一个典型的迭代优化过程。

2　软间隔 MKL 模型

基于硬间隔的 SVM 算法通常假定样本是一定可分的[9-10]，不允许有采集的误差，所以容易出现过拟合的情况。为了使得 SVM 能够在实际应用中有效，通常的做法是引入一个松弛变量，以缓解过拟合。基于这个原理，在 MKL 中可以采用软间隔的框架模型。对于给定点的 M 个基核函数 $\boldsymbol{K}=[K_1,K_2,\cdots,K_M]$ 以及给定采集的样本集 $T=\{(x_i,y_i)\,|\,i=1,\cdots,n\}$，定义松弛变量为目标间隔 τ 和对偶的目标函数 SVM$\{K_m,\alpha\}$ 之间的差值。则对于每个基核有

$$\xi_m=\tau-\mathrm{SVM}\{K_m,\alpha\}\ \forall\, m=1,\cdots,M \quad (6)$$

由于引入的核的松弛变量所产生的损失可以用一个函数表示，$z_m=l(\xi_m)$，$l(\cdot)$为一般的损失函数，因此考虑铰链损失函数($l(\xi_m)=\max(0,\xi_m)$)，通过对于每个基核定义核的松弛变量因子，提出基于软间隔的 MKL 框架模型。当取铰链损失函数，要求解的目标函数如下

$$\begin{aligned}&\min_{\tau,\alpha\in A,\xi_m} -\tau+\theta\sum_{m=1}^{M}\xi_m\\ &\text{s.t.}\quad \mathrm{SVM}\{K_m,\alpha\}\geqslant\tau-\xi_m,\xi_m\geqslant 0\end{aligned} \quad (7)$$

该目标函数是使得类间隔 τ 最大化，同时考虑对于 M 个基核函数引入松弛因子所带来的误差。变量 θ 起到了平衡由 ξ_m 所带来的损失度和最大间隔的作用。上述铰链损失软间隔的 MKL 可以表示为

$$\min_{\mu\in M_1}\max_{\alpha\in A} J(\mu,\alpha) \quad (8)$$

式中：$J(\mu,\alpha)=-\frac{1}{2}\sum_{m=1}^{M}\mu_m(\alpha\cdot y)'K_m(\alpha\cdot y)$，$M_1=\{\mu|\sum_{m=1}^{M}\mu_m=1,0\leqslant\mu_m\leqslant\theta\}$，注意到上面的形式和硬间隔的 MKL 一致，唯一的区别是 μ 的取值范围不同，要求 μ 不能超过正则化的参数 θ，这样可以防止极端大值出现在组合系数里面，风险函数同样被参数 θ 所控制，MKL 的解也与该参数 θ 有关，通过参数平衡了最大分界面和 μ 之间的关系。

对于基于铰链损失的软间隔，式(8)可以转换为如下形式的原始问题

$$\begin{aligned}&\min_{\mu\in M_1,f_m,b,p,\xi_i}\frac{1}{2}\sum_{m=1}^{M}\frac{\|f_m\|^2_{H_m}}{\mu_m}+C\sum_{i=1}^{n}\xi_i-p\\ &\text{s.t.}\quad y_i\sum_{m=1}^{M}f_m(x_i)+b\geqslant p-\xi_i,\ \xi_i\geqslant 0\end{aligned} \quad (9)$$

上述原始问题是一个凸的线性约束问题，可以通过基于块下降算法进行求解。

3 实验

3.1 实验数据库

语音情感数据库是研究语音情感必需的研究基础，具有极为重要的意义。目前国际上流行的语音情感数据库有 AIBO 语料库、VAM 数据库、丹麦语数据库、柏林数据库、SUSAS 数据库等。但是各个语音库标准不一，而且涉及版权的问题，中文的语音情感数据库也较少。

为此，本文所采用的数据库是自行录制的。录制人员包含 20 个年龄在 20～30 周岁的身体健康的成年人，其中男女比例为 1:1，要求录音者分别用高兴、悲伤、愤怒、害怕、中性 5 种情感朗读录音材料。同时为了避免外界干扰，录音者单独参与录制，录制环境为消音室。录音的标准是采用采用采样频率为 48 kHz，编码方式为 PCM，采用单声道 16 bit 量化。

为了保证所采集的情感数据库的可靠性，选取了 10 名未参与录音的人员进行听辨的实验，对于录制的 2 000 条语句，采用以下方式进行筛选：(1)超过 70%的实验者认为该语音是自然的语音；(2)超过 80%的实验者认为该语音能够表达对应类型的情感。经过筛选，一共保留了1 232条短句，各种情感的语句组成如表 1 所示。

表 1 中文语音情感数据库语句构成

Table 1 Constitution of Chinese speech emotion database

情感类别	高兴	悲伤	愤怒	害怕	中性
语句数量	238	254	245	261	234

3.2 实验步骤与参数

本文所采用的改进的多核方法的语音情感识别算法如图 1 所示，该方法分为两个步骤：(1)根据语音情感特征样本使用高斯径向基核函数，并得到最优高斯核函数的参数 σ_{opt}；(2)根据优化的 σ 构造不同的基核并加入其他类型的基核，按照软间隔的 MKL 模型，得到基于多核学习的分类器方程。本实验选用的基核数为 22 个。

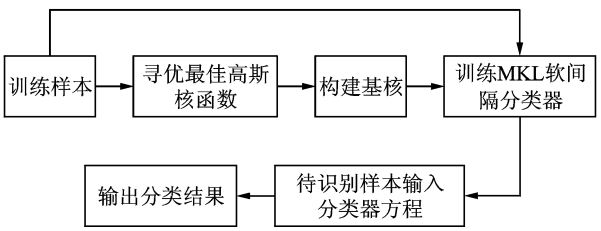

图 1 改进的多核方法的语音情感识别算法框图

Fig. 1 Diagram of speech emotion recognition based on modified MKL algorithm

在进行语音识别前，首先对每句语音信号进行分帧，其中帧长为 25 ms，50%的帧间交叠。分帧以后提取特征，包括 12 阶的 Mel 例谱函数，基音、共振峰以及这些特征类别的一阶、二阶差分，共得到 144 维的情感特征向量。将一部分样本选择为训练样本，另一部分选为测试样本。基于 SVM 原理，构建“一对一”的分类器。对于 5 种感情的语音，构建 10 个超平面作为分类器，分别是“愤怒—害怕”“愤怒—悲伤”“愤怒—中性”“愤怒—高兴”“害怕—悲伤”“害怕—中性”“害怕—高兴”“悲伤—中性”“悲伤—高兴”“中性—高兴”，分类判决时采

用投票法。

每次实验中,将样本集随机分为两类:一类是训练样本,另外一类作为测试样本,实验重复进行多次。为了统一每个类别的样本都取 200 句,组成语音样本库。实验中从每个样本中选取不同比例的数据 R 作为训练样本,然后将剩下的样本作为待识别的样本,最后得到识别结果。

3.3　实验结果与讨论

表 2 所示为不同样本下多核学习的识别率。由表可知,随着训练样本数量的增加,平均识别正确率也是增加的,但是当样本比例达到 70%以后,分类器的性能已经稳定,识别率趋于稳定。同时随着训练样本的增加,分类器的训练时间也越来越久,对于实时性要求较高的系统,很难满足要求。根据最高的识别效率,得到其中一次实验中每个具体情感类别的识别结果如表 3 所示。

表 2　不同比例训练样本 MKL 的识别正确率

Table 2　Correct recognition rate of MKL under different proportions of training samples

训练样本比例 R	训练样本数量	识别正确率/%
20	20	86.0
40	60	89.9
60	100	90.8
70	140	91.2
90	180	91.2

表 3　改进核方法语音情感识别结果

Table 3　Speech emotion recognition results of the improved kernel algorithm

	愤怒	害怕	悲伤	中性	高兴
愤怒	0.91	0.00	0.02	0.06	0.01
害怕	0.00	0.90	0.03	0.06	0.01
悲伤	0.00	0.00	0.99	0.01	0.00
中性	0.03	0.04	0.07	0.86	0.00
高兴	0.08	0.01	0.00	0.02	0.89

由表 3 可知,基于改进核方法语音情感识别的效果比较明显。结果显示该方法对于悲伤,愤怒等情感的识别正确率较高,对于其他情感的识别率相对较低,而总体的识别率达到了 90%以上,具有良好的分类效果。为了说明改进核方法的情感识别的性能,将与前面介绍的几种方法进行对比,训练样本取 30 的情况下,比较神经网络[11]、模糊支持向量机[12-13]和改进多核学习方法的平均识别率,结果如下表 4 所示。由表可知,在相同的训练样本下,从 3 种算法的识别结果中可看出,3 种方法中,改进核方法效果最佳,神经网络法最差。

表 4　不同识别算法的情感识别结果对比

Table 4　Compare of emotion recognition results for different recognition algorithms

识别算法	识别正确率/%
神经网络法	76.3
模糊支持向量机	87.7
改进多核学习方法	90.6

4　结束语

本文在支持向量机理论的基础上,提出了对核方法的改进算法,并将该改进算法应用于汉语语音情感识别中。首先考虑对高斯核函数进行优化。根据可分离性的目标函数,采用牛顿法迭代求解到最优参数,同时利用该参数构造 MKL 的基核。在此基础上,利用松弛因子构造了基于软间隔的 MKL 目标函数以及优化求解算法,该算法能够同时求解分类器的方程和基核的组合系数。最后通过语音情感识别实验,验证了改进核方法的性能,证明了该算法的有效性。

参考文献:

[1] Ververidis D, Kotropoulos C. Emotional speech recognition: resources, features, and methods [J]. Speech Communication, 2006, 48(9):1162-1181.

[2] 王治平,赵力,邹采荣. 基于基音参数规整及统计分布模型距离的语音情感识别[J]. 声学学报, 2006, 31(1): 28-34.
Wang Zhiping, Zhao Li, Zou Cairong. Emotional speeeh recognition based on modified parameter and distance of statistical model of pitch [J]. Chinese Journal of Acoustics, 2006, 31(1): 28-34.

[3] Schuller B, Vlasenko B, Eyben F, et al. Acoustic emotion recognition: a benchmark comparison of performances [C]// Proceedings of 2009 IEEE Workshop on Automatic Speech Recognition & Understanding. Merano, Italy: IEEE Computer Society Press, 2009: 552-557.

[4] Lanckriet G, Cristianini N, Bartlett P, et al. Learning the kernel matrix with semidefinite programming [J]. The Journal of Machine Learning Research, 2004(5): 27-72.

[5] Bach F, Lanckriet G, Jordan M. Multiple kernel learning, conic duality, and the SMO algorithm [C]

// Proceedings of Twenty-First International Conference on Machine Learning. Banff, Canada: ACM Press, 2004: 41-48.

[6] Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis [M]. Cambridge University Press, 2004.

[7] Abbasnejad M, Ramachandram D, Mandava R. A survey of the state of the art in learning the kernels [J]. Knowledge and Information Systems, 2012, 31 (2): 193-221.

[8] Bach F. Consistency of the group lasso and multiple kernel learning [J]. The Journal of Machine Learning Research, 2008(9): 1179-1225.

[9] Xu Z, Jin R, Yang H, et al. Simple and efficient multiple kernel learning by group lasso [C] // Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: Omni Press, 2010: 1175-1182.

[10] Wu Z, Zhang H, Liu J. A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method[J]. Neurocomputing, 2014, 125:119-124.

[11] 余华，黄程韦，金赟，等. 基于粒子群优化神经网络的语音情感识别[J]. 数据采集与处理. 2011,26(1): 57-62.
Yu Hua, Huang Chengwei, Jin Yun, et al. Speech emotion recognition based on particle swarm optimizer neural network[J]. Journal of Data Acquisition and Processing, 2011,26(1):57-62.

[12] 赵力，黄程韦. 实用语音情感识别中的若干关键技术[J]. 数据采集与处理，2014,29(2):157-170.
Zhao Li, Huang Chengwei. Key technologies in practical speech emotion recognition[J]. Journal of Data Acquisition and Processing, 2014,29(2):157-170.

[13] 李嘉，黄程韦，余华. 语音情感的维度特征提取与识别[J]. 数据采集与处理，2012,27(3):389-393.
Li Jia, Huang Chengwei, Yu Hua. Dimensional feature extraction and recognition of speech emotion[J]. Journal of Data Acquisition and Processing, 2012,27 (3):389-393.

作者简介：奚吉(1977-)男，博士，研究方向：信息与信号处理，E-mail：xijie952611@gmail.com；赵力(1958-)男，教授，研究方向：语音信号处理等；左加阔(1985-)男，博士，研究方向：信号处理、认知无线电等。