计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(1):225−230 [doi: 10.15888/j.cnki.csa.006153]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

GMM-UBM 和 SVM 在说话人识别中的应用①

李　荟[1], 赵云敏[2]

[1](东北石油大学 计算机与信息技术学院, 大庆 163318)
[2](大庆油田第一采油厂, 大庆 163318)

摘　要: 针对说话识别领域短语音导致的训练数据不充分的问题, 选择能够突出说话人个性特征的 GMM-UBM 作为基线系统模型, 并引入 SVM 解决 GMM-UBM 导致的系统鲁棒性差的问题. 选择不同的核函数对 SVM 的识别性能有较大的影响, 针对多项式核函数泛化能力较强、学习能力较差与径向基核函数学习能力较强、泛化能力较差的特性, 对两种单核核函数进行线性加权组合, 以使组合核函数兼具各单核的优点. 仿真实验结果表明, 组合核函数 SVM 的识别率和等错误率明显优于不引入 SVM 的 GMM-UBM 的基线系统及其它三个单核函数, 并在不同信噪比情况下也兼顾了系统识别准确率与鲁棒性.

关键词: 说话人识别; GMM-UBM; SVM; 组合核函数

引用格式: 李荟,赵云敏.GMM-UBM 和 SVM 在说话人识别中的应用.计算机系统应用,2018,27(1):225–230. http://www.c-s-a.org.cn/1003-3254/6153.html

Application of GMM-UBM and SVM in Speaker Recognition

LI Hui[1], ZHAO Yun-Min[2]

[1](Institute of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)
[2](No.1 Oil Production Plant, PetroChina Daqing Oilfield Company, Daqing 163318, China)

Abstract: Aiming at the problem that training data is insufficient due to little training data in speaker recognition system, this paper adopts GMM-UBM as the background model which can identify the characteristics of the target speaker. And SVM is introduced to solve the problem of poor robustness of the system caused by GMM-UBM. It has much influence on SVM identification performance with different kernel functions. Aiming at the Characteristics of Polynomial kernel with good generalization ability and poor earning ability and Gaussian kernel with good earning ability and poor generalization ability, it structures a new combination kernel function which combines the advantages of each single kernel function by linear weighted method. The experimental results show that the recognition rate and Equal Error Rate of the combination kernel is more ideal than other kernel functions. And it achieves satisfactory recognition rate and robustness in the situations of different signal-to-noise ratio.

Key words: speaker recognition; GMM-UBM; SVM; combination kernel function

1　引言

说话人识别是一项根据说话人的语音参数来区分说话人身份的技术, 广泛地应用于语音拨号、安全控制、电话银行、司法鉴定、语音导航等方面[1]. 但在实际应用中, 系统的识别性能受到短语音、背景噪声干扰、信号引起的信号畸变等多种因素的影响, 其中短语音导致的训练数据不足是较为常见且较为突出的问题. GMM-UBM 模型能够有效地解决训练数据不充分的问题, 但它导致的问题是系统鲁棒性差, SVM 利用帧特征向量在空间分布的高斯混合的均值进行识别,

① 收稿时间: 2017-04-07; 修改时间: 2017-04-26; 采用时间: 2017-05-08; csa 在线出版时间: 2017-12-22

能显著提高系统的鲁棒性能, 而且 SVM 还能有效地解决小样本、低维线性不可分等实际问题. 但应用 SVM 对说话人进行识别, 重点就是选择合适的核函数, 为了提高性能, 这里根据单核核函数的特性不同构造了一种组合核函数. 因此, 本文选用 GMM-UBM 为基线系统模型, 在此基础上应用 SVM 组合核函数作为分类器进行分类.

2 GMM-UBM 基线系统模型

高斯混合模型 (GMM) 利用多个高斯分布的加权混合来描绘说话人的特征空间分布[2], 因此, 混合度越高, 识别性能越好, 当然所需的训练语音也会越多. 但在很多实际应用中, 有些训练语音比较短, 这些有限的训练语音无法很好地代表说话人所有可能的发音情况, 因此, 训练得到的模型也无法很好地表征说话人的特征, 这种情况使 GMM 识别的性能较差.

GMM-UBM 模型能够有效地解决 GMM 由于训练语音不足导致的问题. 通用背景模型 (UBM) 是一个高阶的 GMM, 通常能够达到 1024~4096 个混合度. 它由数百人、性别比例均衡、长时间的语音训练得到的模型, 使得 UBM 基本包括了所有说话人的特征参数. 这样, 短的语音未覆盖到的发音部分就可以用 UBM 中与说话人无关的特征分布近似描述, 降低训练语音短带来的影响, 继而提高系统识别性能. 但 GMM-UBM 在说话人应用中存在受信道影响较大的问题, 使系统的鲁棒性较差[3], 鉴于此, 这里用 GMM-UBM 为基线系统模型.

3 SVM

3.1 SVM 原理

SVM 是由 Vapnik 等人提出的基于统计学习理论和结构风险最小化原理的一种分类算法[4]. 基本思想是将低维空间无法线性可分样本映射到高维特征空间, 并构建一个最优分类面以达到使两类样本正确分开, 且类间间隔最大的结果.

给定训练样本集$X=\{(x_i,y_j),i=1,\cdots,n;j=-1,+1\}$, 其中 x_i 和 y_i 分别代表第 i 个样本和类别标号. 找出区分不同样本的最优分类面为:

$$y(x)=\mathrm{sgn}\{w\cdot\phi(x)+b\} \tag{1}$$

其中 x、w、b 和$\phi(x)$分别表示输入向量、权重系数、偏移量和特征映射. 这样可将式 (1) 表示成求解以下问题:

$$\begin{aligned}&\min_{w,b,\xi}\ \frac{1}{2}w^Tw+C\sum_{i=1}^{N}\xi_i,\\&\text{s.t.}y_i\left(w^t\cdot\phi(x_i)+b\right)\geq 1-\xi_i,\xi_i\geq 0,i=1,\ldots,N\end{aligned} \tag{2}$$

其中 C 和ξ_i分别是惩罚因子和松弛变量. 将以上问题应用 Lagrange 转化为对偶问题:

$$\begin{aligned}&\max Q(\alpha)=\max\sum_{i=1}^{N}\alpha_i-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_j\phi(x_i)\phi(x_j)\\&\text{s.t.}\sum_{i=1}^{N}\alpha_iy_i=0,0\leq\alpha_i\leq C,i=1,2,\ldots,N\end{aligned} \tag{3}$$

其中α_i为拉格朗日系数, 可得最优分类函数为:

$$f(x)=\mathrm{sgn}((w\cdot x)+b)=\mathrm{sgn}(\sum_{i=1}^{N_V}(a_iy_i(\phi(x_i)\cdot\phi(x_j)+b)) \tag{4}$$

解式 (4), 其中大于零的解所对应的样本 x_i 就称为支持向量. 在实际问题中, 低维空间的向量集总是难以线性划分, 而通常的解决办法就是将低维空间的向量集映射到高维空间以线性划分, 但导致的最大问题就是计算复杂度大大增加, 引入核函数可以有效地解决这个问题. 相应的判别函数为:

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{N}a_iy_iK(x_i,x_j)+b\right\} \tag{5}$$

3.2 SVM 核函数

常见的核函数有:

① 线性内积 (Linear) 核函数:

$$K(x_i\cdot x_j)=\left(x_i\cdot x_j\right) \tag{6}$$

② 高斯径向基 (Gaussian) 核函数:

$$K(x_i\cdot x_j)=\exp(-\left\|x_i-x_j\right\|^2/2\sigma^2) \tag{7}$$

其中σ是 Gaussian 核函数的宽度系数.

③ 多项式 (Polynomial) 核函数:

$$K(x_i\cdot x_j)=\left[\left(x_i\cdot x_j\right)+C\right]^d,d>0 \tag{8}$$

其中 d 是 Polynomial 核函数的幂指数, C 是一个常数, 实际应用中一般令 C=1[5].

④ 两层神经网络 (Sigmoid) 核函数:

$$K(x_i\cdot x_j)=\tanh(\left(v(x_i\cdot x_j\right)+\theta) \tag{9}$$

其中 v 和θ分别是 Sigmoid 核函数的一个标量及其位移参数, Sigmoid 核函数在实际应用中并不多见, 这里也不予考虑.

3.3 SVM 核函数的特性

根据 SVM 核函数特性的不同, 可分为局部性核函数和全局性核函数.

由公式 (7) 可知, 当σ足够小时, Gaussian 核函数的分类性能较好, 当$\sigma \to 0$时, 分类性能达到最佳, 但当样本距离逐渐增大时, 它的值会逐渐下降且趋于零. 如图 1 所示, 当σ^2(图中用 p 表示) 分别取 0.1, 0.2, 0.5, 1, 测试点为 0.2 时, Gaussian 核函数值在测试点 0.2 附近较大, 离测试点较远时, 值会显著下降, 因此 Gaussian 核函数插值能力较强, 但泛化能力较差.

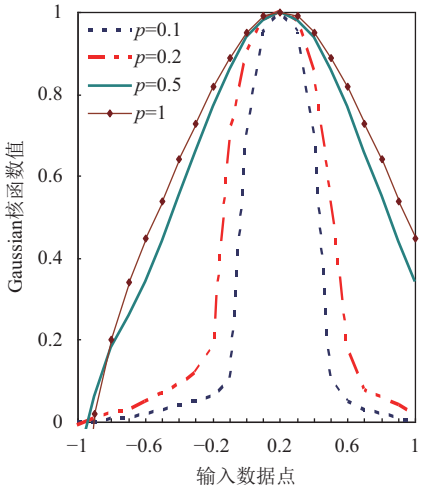

图 1 Gaussian 核函数特征曲线图

根据公式 (8), 当测试点取 0.2, 可得图 2. 可以看出, Polynomial 核函数对测试点附近以及较远的数据都有影响, 且相差不大, 可见全局核函数具有较强的泛化能力, 但局部学习能力较弱.

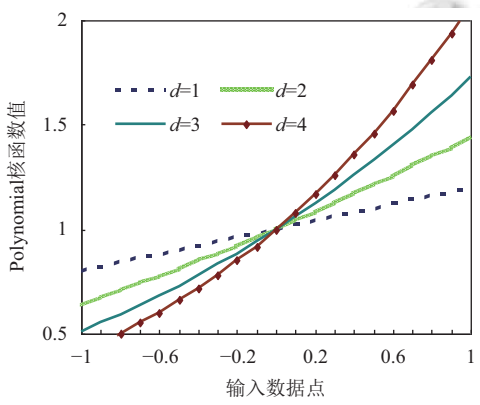

图 2 Polynomial 核函数特征曲线图

3.4 组合核函数的构建

选择的核函数是否合适直接影响着 SVM 的识别性能, 鉴于 Gaussian 核函数较强的局部学习能力和 Polynomial 核函数较强的全局泛化能力, 可将两种核函数进行线性组合, 使其充分发挥各自单核的优点.

由核函数的构成条件可知, 两个核函数的线性加权, 仍然满足 Mercer 条件, 组合后的核函数如公式 (10) 所示.

$$K(x_i \cdot x_j) = \alpha\left[\left(x_i \cdot x_j\right)+1\right]^d + (1-\alpha)\exp(-\left\|x_i - x_j\right\|^2/2\sigma^2) \tag{10}$$

其中$0 \le \alpha \le 1$, 表示加权系数, 表示两种核函数的比例系数, 根据不同的数据分布调整权系数α, 以达到组合核函数最佳的识别性能. 组合核函数的特征曲线图如图 3 所示. 可知, 当取合适的参数时, 组合核函数能够兼具良好的插值能力与良好的泛化能力.

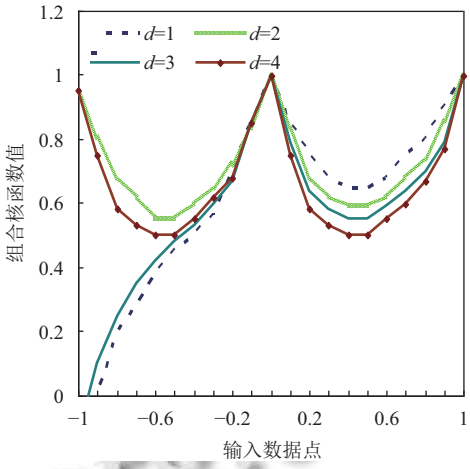

图 3 组合核函数特征曲线图

3.5 SVM 参数优化方法

SVM 核函数的参数优化方法主要有网格搜索法、交叉验证法和蚁群算法、遗传算法等智能算法. 与其它算法相比, 网格搜索法能实现并行操作, 因此效率较高, 但缺点是精度不高[6], 多重网格搜索可以在一定程度上提高参数精度. 鉴于组合核函数中参数较多, 综合考虑参数的精度与效率, 这里选取多重网格搜索来优化参数.

网格搜索法的主要思路是先确定搜索范围和步长, 再按照确定的步长沿每个参数方向生成网格, 得到的网格中的节点即构成可能的参数组合. 在上次网格寻优最优点的基础上, 减小搜索步长, 并再次寻优, 就是多重网格搜索. 如要确定参数 C 与 d, 首先设定参数 C 的范围为$C \in [C_1, C_2]$, 搜法步长为C_s, 参数 d 的范围

为$d \in [d_1, d_2]$, 搜法步长为d_s, 然后针对每对参数$[C', d']$进行训练. 多重网格搜索法是完成一次网格搜索后得到一组最优的参数组合$[c'', d'']$, 再对$[c'', d'']$附近一定范围内实现更为细致的一次网格搜索, 以提高参数优化的精度.

4 GMM-UBM 和 SVM 组合核函数的说话人识别过程

图 4 为运用 UBM-SVM 组合核函数进行说话人识别的框架图, 基于 UBM 的 SVM 组合核函数的识别过程从整体上包括训练和测试两个阶段. 如图 4 所示, 一是训练阶段, 输入训练语音信号, 这些信号经过预处理后形成信号帧, 经过特征提取后形成帧特征向量, 它们是以 GMM-UBM 作为基线模型经过参数自适应后形成的定长超向量, 这些超向量可以直接作为 SVM 组合核函数分类器的输入, 在此基础上并进行参数优化, 根据优化后的特征参数就可以建立训练样本模式库. 二是测试阶段, 输入的测试语音信号同样经过预处理、特征提取、GMM-UBM 为基线模型进行自适应、SVM 组合核函数分类几个过程, 将得到的特征参数与训练过程得到的样本模式库里的所有参考模型进行匹配, 即可输出判决结果.

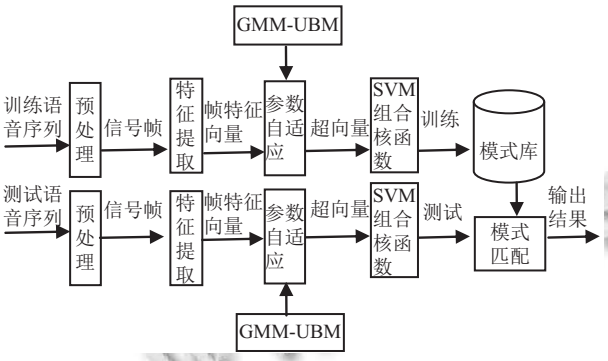

图 4 基于 GMM-UBM 和 SVM 组合核函数的说话人模型识别

5 实验结果及分析

5.1 数据来源

本实验采用自建语音库, 正常情况下, 选取 400 个说话人 (200 男 200 女) 进行录音, 时间为 5-6 分钟/人. 训练语音选取每个说话人录音的前 4 分钟, 从 400 人中随即选择 20 人的后 50 s 作为测试语音, 使训练语音与测试语音之间不重叠. 对所得数据进行处理, 预加重系数为 0.97, 分析窗选用宽度为 32 ms 的汉明窗, 帧长为 25 ms, 步长为 10 ms, 选取 16 维的 MFCC 系数以及其 16 维一阶差分. 自适应方法选为 EigenVoice, 维数取为 10, 段间隔为 5 s, 这里自适应时长取 10 s.

5.2 性能评价指标

识别率 (正确识别率) 是系统识别性能最为直观的评价指标, 但对于一个实际说话系统来说, 错误拒绝率 FRR 和错误接受率 FAR 也是两个重要的性能评价指标.

FRR = 错误拒绝的次数/类内测试的总次数 (11)

FAR = 错误接受的次数/类间测试的总次数 (12)

但以上两个指标互相矛盾, 因此, 综合考虑两个指标, 一般采用二者相等时的错误率作为衡量标准, 称为等错误率 EER. 这个值在一定程度上能够反映系统的鲁棒性.

因此本实验采用识别率和等错误率两个指标作为评价模型分类性能的标准, 综合评价系统识别的准确率与鲁棒性.

5.3 参数确定

应用多重网格搜索法进行参数寻优, C、σ、d 和α分别在[0, 1000]、[0, 10]、[1, 20]、[0, 1]范围内进行多次网格寻优. 得到的最优参数如下: Linear 核参数 C=128, Gaussian 核函数参数 C=16、σ=0.2, Polynomial 核参数 C=32、d=2, 组合核函数参数 C=64、d=4, σ=0.25、α=0.2.

5.4 实验结果及其分析

实验一. 在混合度不同情况下, 比较 GMM 与 GMM-UBM 基线系统的识别性能, 实验结果见表 1.

表 1 不同混合度情况下 GMM 与 GMM-UBM 识别性能对比

基线系统模型	混合度	识别率(%)	EER(%)
GMM	16	71.4	16.6
	64	73.2	15.4
	256	76.6	13.2
	512	78.5	12.8
GMM-UBM	256	80.1	11.8
	512	83.2	11.2
	1024	85.4	10.4
	2048	88.6	9.5

实验结果表明, 随着混合度的增加, GMM 与 GMM-UBM 的识别率与 EER 都有所改善. 通常情况下 GMM-UBM 混合度都比较高, 即使同为 256 和 512 的情况下, GMM-UBM 的识别率也分别高于 GMM

3.5% 和 4.7%, 但 GMM-UBM 的 EER 不低, 即使随着混合度增加 EER 会下降, 但系统复杂性会增加.

实验二. 综合考虑系统复杂性与识别性能要求, 选取 GMM-UBM 混合数为 1024, 比较 SVM 选取不同核函数的识别性能.

由表 2 可知, 引入 SVM 核函数后, Gaussian 核、Polynomial 核和组合核的识别性能都优于 GMM-UBM 不引入 SVM 的基线系统模型. 可见, 引入 SVM 核函数不仅能提高系统的鲁棒性, 同时也能提高系统的识别率. 另外, 在以上核函数中, 组合核函数的识别性能最好, 它的识别率分别优于 Linear 核、Gaussian 核和 Polynomial 核 10.6%、7.3% 和 5.4%, EER 也优于其它三个单核.

表 2 不同 SVM 核函数识别性能对比

核函数	参数	识别率(%)	EER(%)
Linear	C=128	83.2	10.2
Gaussian	C=16、σ=0.2	88.4	7.8
Polynomial	C=32、d=2	86.5	9.1
组合核	C=64、d=4, σ=0.25、α=0.2	93.8	5.6

实验三. GMM-UBM 混合数为 1024, 人工添加白噪声, 得到信噪比不同的语音, 比较不同核函数的识别性能实验结果见图 5 和图 6.

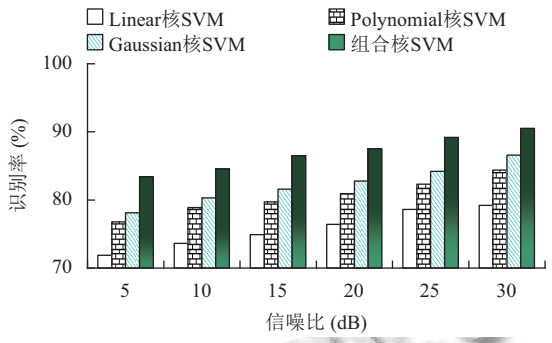

图 5 基于不同信噪比不同核 SVM 识别率对比

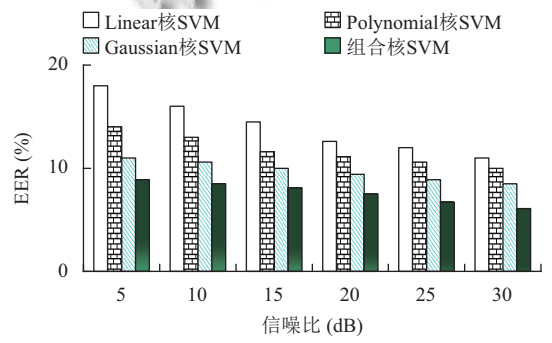

图 6 基于不同信噪比不同核 SVM 的 EER 对比

由图 5 和图 6 可知, 所有 SVM 核函数的识别性能都随着信噪比的减小而降低. 但对于给定的某一信噪比来说, 组合核函数的识别率要高其它核函数, EER 要低于其它核函数, 说明基于 GMM-UBM 基线系统的 SVM 组合核函数能够提高系统的识别率与鲁棒性.

实验四. 假定 20 个说话人, 选择不同的高斯混合数, 比较 SVM 选取不同的核函数的训练时长. 具体数据见表 3.

表 3 不同高斯混合数下, 不同 SVM 核函数训练时间比较

高斯混合数(N)	训练时长(s)			
	Linear	Gaussian	Polynomial	组合核
N=256	186	205	217	264
N=512	224	256	278	288
N=1024	311	329	337	356
N=2048	632	732	725	824
N=4096	1336	1556	1589	1682

由表 3 可知, 在不同的高斯混合数情况下, 组合核函数的运行时间比 Linear 核平均多 21%, 比 Gaussian 核平均多 10%, 比 Polynomial 核平均多 9%. 因为组合核参数最多, 其次是 Gaussian 核和 Polynomial 核, Linear 核参数最少, 运行时间与参数基本成正比. 组合核 SVM 的参数虽比 Gaussian 核和 Polynomial 核多, 但运行时间就多了 10% 左右, 主要原因有: 一是参数优化采用的是多重网格搜索法, 这种方法的最大优势是可以同时搜索多个参数, 在一定程度上能减少参数搜索的时间. 二是经过自适应后的超向量可以直接作为 SVM 的输入, 这样可以实现整体语音序列上进行分类, 因此能够降低运算复杂度. 综合考虑识别率、等错误率及运行时间, 组合核 SVM 是较理想的选择.

6 结语

针对训练数据不充分问题, 选取 GMM-UBM 为基准系统模型, 并应用 SVM 对其参数进行优化, 本文基于单核函数的特性, 构建具有良好的泛化能力与良好的学习能力的组合核函数. 在说话人识别的仿真实验中, 组合核函数表现出明显优于其它单核 SVM 的良好性能. 而且在信噪比不同、高斯混合数不同的情况下, 表现依旧不俗. 但由于组合核函数引入过多的参数, 增加了模型复杂度及系统运算时间. 模型参数自适应方法能够在一定程度上解决这个问题, 在模型参数自适应方法中基于特征音 EV 模型的变换方法由于能用少量的训练数据快速的调整模型以实现自适应得到广泛

的应用, 在此基础上再采用 SVM 组合核函数训练方法来弥补模型参数自适应方法的局限性, 能够弥补参数设置过多的问题, 但如何在保障识别正确率与系统鲁棒性的基础上减少参数设置依然是需要进一步研究的问题.

参考文献

1 王韵琪. 自适应高斯混合模型及说话人识别应用. 计算机系统应用, 2015, 24(6): 143–147.
2 翟玉杰. 基于 GMM-SVM 说话人识别的信道算法研究[硕士学位论文]. 长春: 吉林大学, 2015.
3 鲍焕军, 郑方. GMM-UBM 和 SVM 说话人辨认系统及融合的分析. 清华大学学报 (自然科学版), 2008, 48(S1): 693–698.
4 吕洪艳, 刘芳. 组合核函数 SVM 在说话人识别中的应用. 计算机系统应用, 2016, 25(5): 168–172.
5 栗志意, 张卫强, 何亮, 等. 基于核函数的 IVEC-SVM 说话人识别系统研究. 自动化学报, 2014, 40(4): 780–784.
6 刘群锋. 最优化问题的几种网格型算法[博士学位论文]. 长沙: 湖南大学, 2011.