组合核函数 SVM 在说话人识别中的应用①

吕洪艳, 刘 芳

(东北石油大学 计算机与信息技术学院, 大庆 163318)

摘 要: 针对说话人识别实际应用中训练数据不足的问题, 选取 GMM-UBM 作为基准系统模型, 用 EigenVoice 对其作自适应, 应用泛化能力较强的多项式核函数和学习能力较强的径向基核函数进行线性加权组合后的组合核函数进行模型参数优化, 并用多重网格搜索法确定核函数的最优参数, 采用 DAG 方法实现 SVM 核函数的多元分类. 在仿真实验中评估了线性核、多项式核、径向基核以及组合核函数, 实验结果表明, 在采用正确的参数前提下, 在不同的多分类策略、自适应时间、信噪比和不同的说话人数量的情况下, 组合核函数的识别性能明显都优于其它三个单核函数.

关键词: 说话人识别; 组合核函数; SVM; GMM-UBM

Application of Combination Kernel Function SVM in Speech Recognition

LV Hong-Yan, LIU Fang

(Institute of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: In the problems of practical application, GMM - UBM is adopted as the background model when the training data is insufficient in speaker recognition system. EigenVoice is used as adaptation ways, then it structured a new combination kernel function combined with homogeneous polynomial kernel with good generalization ability and radial basis kernel function with good earning ability by linear weighted method to optimize model parameter. The optimal parameters of kernel function are determined through the multiple grid search method. DAG method is adopted to realize multivariate classification of SVM kernel function. Then the linear kernel, homogeneous polynomial kernel, radial basis kernel function and combination kernel function are evaluated in the experiments. The experimental results show that the identify performance of the combination kernel is more ideal than that of other kernel functions in the different classification strategy, different adaptive time, different signal-to-noise ratio and different number of speakers.

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

1 引言

说话人识别是一种根据说话人的语音对其进行区分的技术, 广泛地应用于司法鉴定、语音拨号、数据库访问、电话购物等方面. 但在实际应用中, 由于存在训练数据不充分、短话音、声音模仿、噪声干扰等问题, 导致说话人识别系统的性能不高[1]. 其中, 训练数据不足是最为常见而且对识别性能影响较大的问题, 也是此领域一直研究的重要课题.

目前, 针对训练数据不充分问题, 主要通过说话人自适应方法和模型参数估计算法来解决. 在说话人自适应的方法中, 模型参数自适应因其能够充分刻画说话人的特性, 识别效果较好而得到了广泛的应用. 模型参数估计算法能够在模型参数自适应的基础上进行参数优化, 进一步提高识别性能. 模型参数估计算法中常用的是最大似然估计 MLE 方法, 但由于 MLE 依据的是与实际不符的假设参数分布进行优化, 所以性能不佳[1]. 针对这一问题, 基于结构风险最小化的原理的 SVM 得到应用, 而且它还能有效地解决小样本、非线性不可分、高纬度等说话人识别领域的实际问题. 选择不同的核函数对 SVM 的识别性能有较大的

① 收稿时间:2015-09-15;收到修改稿时间:2015-10-30

影响, 本文结合单核函数的不同特性来构造一种组合核函数, 以期能够提高说话人识别性能.

2 基于SVM的说话人识别过程

运用SVM进行说话人识别, 包括训练阶段和测试阶段. 具体框架如图 1 所示. 在训练阶段, 输入大量训练语音信号, 经过预处理后得到信号帧信号, 特征提取后得到帧特征向量, 这些帧特征向量以 GMM-UBM 为基线系统模型经过自适应后形成超向量, 它们直接作为SVM分类器的输入并进行参数优化, 用优化后的特征参数建立训练样本模式库. 在测试阶段, 输入的待测语音信号同样经过预处理、特征提取、GMM-UBM 作为基线系统进行自适应、SVM 分类器后得到的特征参数与训练样本模式库所有的参考模型进行模式匹配, 最后输出判决结果.

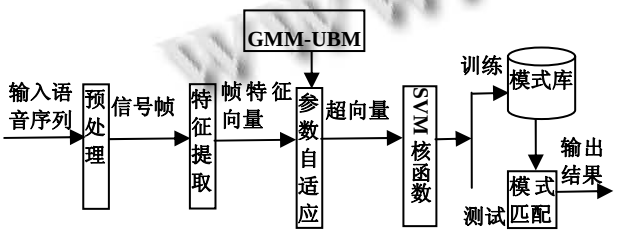

图 1 基于SVM的说话人模型识别系统框

(1) 预处理. 预处理主要包括采样与量化、预加重和分帧加窗三个阶段[2]. 依据一定的带宽和采样率对输入的模拟信号进行采样, 使之转换为数字信号. 再通过预加重数字滤波器实现加重处理, 使语音信号包含大量的个性特性, 以利于特征提取. 由于语音信号典型的非平稳特性, 需要加窗分帧处理, 这里采用在说话识别领域应用比较广泛的交替分段方法和汉明窗进行分帧加窗, 以体现信号的短时平稳特征.

(2) 特征提取. 特征提取的目的是在保留样本特征信息的基础上舍去语义内容. 常用的方法有线性预测倒谱系数 LPCC、线性预测系数 LPC 及梅尔(Mel)频率倒谱系数MFCC等. 其中MFCC系数模拟了人耳对不同频率语音的感知特性, 能够取得良好的识别性能, 并得到了广泛的应用, 这里也选用 MFCC 系数[3]. 经过处理的信号帧会统一成相同维数矩阵的帧特征向量. 一般来说, MFCC 系统维数越大, 识别率越高, 但识别所用时间越长.

(3) GMM-UBM 基线系统模型

在说话人识别的实际应用中, 传统高斯混合模型(GMM)在复杂背景下会由于训练语音不足, 使系统识别性能较差, 而通用背景模型(UBM)可以有效地解决这个问题. 运用最大期望(EM)算法对 GMM 模型训练得到GMM-UBM模型[4]. UBM是一个与说话人无关的高阶的GMM. 通常UBM可达到1024~4096个混合度. 它是由大量样本(通常数百人)、均衡(性别比例均衡)、长时间(至少1小时)语音训练得到的. 由于训练样本大, 所以可以认为 UBM 是所有说话人特征参数的并集. 这样若有未覆盖到的发音情况, 便可以用与说话人无关的特征分布来近似描述, 提高了系统的识别性能. 另外它可以处理不定长的语音序列, 因此, 这里选择 GMM-UBM 作为基线系统模型.

(4) 自适应. 模型参数自适应的目的有两个; 一是优化GMM-UBM的某些参数; 二是将不等长的语音序列转换为定长的超向量, 一组超向量可以表示每个样本的语音, 这些特定长度的超向量可以直接作为 SVM 核函数的输入. 自适应的方法主要有最大后验概率(MAP)算法、最大似然回归(MLLR)、MAP\MLLR和基于本征音(EigenVoice)算法. EigenVoice 算法是一种快速自适应算法, 能够实现在较少的数据下得到良好的识别性能. 在说话识别应用中, 已有研究表明, EigenVoice 能得到较其他三种自适应更为理想的识别性能, 因此这里选用 EigenVoice 作模型参数自适应.

3 基于组合核函数SVM的参数调整

经过自适应优化的超向量可以直接作为 SVM 核函数的输入. SVM 核函数的作用是再次对输入的超向量进行优化, 通过选用合适的核函数、核函数参数及多分类策略进行判别, 以提高系统的识别性能.

3.1 SVM 原理

SVM是由 Vapnik 等人在1996年提出的基于结构风险最小化原理的一种机器学习方法[5]. 基本思想是将训练样本映射到高维特征空间, 并在此空间构建一个最优分类超平面, 把两类样本正确分开, 且使两类样本间分类间隔最大. 具体步骤是先构造一个最优分类面, 再将最优分类面问题转化为求解最优化问题, 应用 Langrange 函数合并优化问题和约束, 再使用对偶理论, 得到分类优化问题, 即最优分类函数为:

$$f(x)=\mathrm{sgn}((w\cdot x)+b)=\mathrm{sgn}(\sum_{i=1}^{N_s}(a_i y_i(\phi(x_i)\cdot\phi(x_j)+b)) \quad (1)$$

假定训练样本集为 $(x_i, y_i), i=1,\cdots,n$，$x \in R^d$，$y \in \{-1,+1\}$，在公式(1)中，x_i 代表第 i 个样本，y_i 是类别标号，w 和 b 分别是超平面的法向量和偏移量，$\phi_i(x)$ 表示特征映射，α_i 为Lagrange系数，N_v 为支持向量数. 对于非线性问题，通过引入核函数 $K(x_i,x_j)$ 使高维空间的内积运算转化为原空间一个内积核函数计算，相应的判别函数为

$$f(x)=\mathrm{sgn}\left\{\sum_{i=1}^{N} a_i y_i K(x_i,x_j)+b\right\} \tag{2}$$

核函数的引入可以实现在不增加算法复杂度的同时实现了非线性算法，这就是 SVM.

3.2 SVM 核函数

在 SVM 中，常见的核函数有以下四种.

① 线性内积核函数

$$K(x_i \cdot x_j)=(x_i \cdot x_j) \tag{3}$$

② 多项式核函数

$$K(x_i \cdot x_j)=[(x_i \cdot x_j)+C]^q, q>0 \tag{4}$$

③ 高斯径向基核函数

$$K(x_i \cdot x_j)=\exp(-\|x_i-x_j\|^2/2\sigma^2) \tag{5}$$

④ Sigmoid 核函数

$$K(x_i \cdot x_j)=\tanh(v(x_i \cdot x_j)+\theta) \tag{6}$$

其中 q，σ，v，θ 为可调节参数，q 是多项式核函数的幂指数，$C \geq 0$ 是一个常数，在实际应用中通常令 $C=1$[6]. σ 是径向基核函数的宽度系数，v 是 Sigmoid 核函数的一个标量，θ 是其位移参数. 选取不同的核函数可以构建不同的 SVM. 目前，线性核函数是最简单的核函数，而 Sigmoid 核函数由于只有参数满足特定条件时，才是半正定的，所以在实际应用不多，这里也不考虑 Sigmoid 核函数.

不同类型的核函数表现出不同的特性，根据其特性不同，常分为局部性核函数和全局性核函数. 径向基核函数是典型的局部核函数，由其表达式可知，当 $\sigma \to 0$ 时，全部训练样本都能够正确分类，但它只对样本距离 σ 相当的小范围内的样本有效，当样本距离大于 σ 且逐渐增大时，它的核函数值会逐渐下降，而且下降的速度会越来越快，说明径向基核函数局部性较强，泛化能力较弱. 多项式核函数是典型的全局核函数，它作用范围较广，甚至对整个数据点都有影响，而且对于离测试点较远的数据仍然有较强的影响，泛化能力较强，但局部学习能力较弱.

3.3 组合核函数构建

应用 SVM 于语音识别的实际应用中，选择合适的核函数是关键. 而核函数的选择主要包括选择核函数和确定核函数的参数两部分内容. 目前，关于单核的构造及其参数选择的研究较多，但采用单核SVM进行语音识别的效果并不十分理想. 由于全局核函数具有较强的泛化能力，而局部核函数具有较强的学习能力，因此，这里将这两种核函数进行组合，使组合核函数兼具它们各自单核的优点，以提高识别性能[6].

根据核函数的构成条件，两个核函数之和仍然满足 Mercer 条件，因此这里对多项式核函数和径向基核函数进行线性组合，即:

$$K(x_i \cdot x_j)=\alpha[(x_i \cdot x_j)+1]^q+(1-\alpha)\exp(-\|x_i-x_j\|^2/2\sigma^2) \tag{7}$$

其中 $\alpha>0$，表示组合核函数中两种核函数的比例系数. 由于组合核函数兼具两核的优点，所以通过选择合适的参数，可以在一定程度上提高语音识别性能.

3.4 SVM 参数优化方法

需要确定的参数有惩罚因子 C、q、σ、及 α 四个参数. SVM 核函数参数优化的主要方法有交叉验证法、网格搜索法及遗传算法、粒子群算法、蚁群算法等一些群智能算法. 交叉验证法优点是简单、易于实现，缺点是精度不高. 网格搜索法优点是模型简单且可以同时搜索多个参数，缺点是当参数较多时精度不能满足要求，但多重网格搜索法能够在一定程度上解决这个问题. 群智能算法相对于以上两种方法来说，算法复杂度都要高得多，此外，遗传算法的优点是对目标函数要求不高，缺点是受初值影响较大. 粒子群的优点是收敛速度快，全局优化性能较好，缺点是精度不稳定，局部优化性能较差. 蚁群算法的优点是精度较高，缺点是收敛速度慢，并且鲁棒性差[6]. 综合考虑算法复杂度、参数精度、效率及参数个数等因素，这里选择多重网格搜索法. 网格搜索法的主要思路是选取合适的搜索范围和步长，再以确定的步长沿着各个参数方向生成网格，网格中的节点就是所有可能的参数组合. 多重网格搜索法从上一次网格寻优最优点开始，减小搜索步长，再次寻优，以此类推.

在说话人识别的实际应用中，不仅期望具有极高的识别率，同时还要具有较高的实时性. 本文中引入组合核函数的目的是提高系统的识别性能，但随着组合核函数中参数的增多，显然会增加算法的运算复杂

度, 针对这一问题, 从以下三方面加以说明: 一是应用多重网格搜法确认核函数参数. 多重网格搜索法的最大优势是模型简单, 易于实现, 可以同时搜索多个参数, 可以在一定程度上减少参数搜索的时间. 二是使用超向量作为SVM的输入. 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, 不需要将语音段切割为彼此独立的的帧, 再根据各个帧的决策值进行判定, 因此能够降低运算复杂度. 三是引入核函数解决非线性问题. 在SVM 中引入核函数能够实现将输入空间升维, 以求在高维空间中将问题变得线性可分, 这种方法的训练复杂度不受特征维数影响, 只取决于支持向量的个数, 因此, 核函数的引入巧妙地避免了高维空间中运算量大的问题. 因此, 以上方法可以在一定程度上减少算法复杂度, 提高算法实际应用的可行性.

3.5 SVM 多类分类方法选择

针对说话人识别的特定应用, 需解决SVM的多分类问题, 常用的多分类方法有“一对多”方法、“一对一”方法和有向无环图(DAG). “一对多”方法是先把某个类别的样本作为正类, 其他剩下的样本作为负类, 分类时将待分样本归为有最大决策函数输出值的那一类. “一对一”方法是在任意两类样本之间都构造一个SVM, 通过每个 SVM 将某一类样本和其它类别样本区分开. 采取“投票法”进行判别, 分类时将待分样本归为得票数最多的那一类. DAG 方法在训练阶段与“一对一”方法相同. 在测试阶段, 它将所有分类器建成一个有向无环图[2]. 待测样本分类时, 从顶部节点开始, 根据判别结果沿着无环图从顶层向下一层游走, 直到在底层找到待测样本的所属类别. 由于前两种方法误分率较高, 所以这里选择 DAG 策略.

4 实验结果及分析

这里采用自建语音库, 在普通情况下, 选取 40 个说话人(20 男 20 女)进行录音, 每人录音时间为 4-5 分钟. 其中训练语音由每个说话人的前 3 分钟录音构成, 测试语音由随机选择20个说话人的后40s的语音构成, 训练语音与测试语音不重叠. 对得到的语音数据进行处理, 预加重系数选为 0.97, 选用汉明窗宽度为 32ms, 帧移为 16ms, 选用 12 阶的 MFCC 系数及其一阶、二阶差分系数构成 36 维的特征向量. 选用 EigenVoice 进行模型参数自适应, 本征音维数取 10, 自适应阶段将测试语音分段, 段间隔为 5s, 最短为 5s, 最长 30s. 这里在无特殊说明的情况下自适应时长取 5s、多分类策略选用 DAG 进行实验.

根据选用的多重网格搜索法, C、q、σ和α分别在[0,1000], [1, 20], [0, 20], [0, 1]范围内进行网格寻优. 得到如下最优参数: 线性核参数惩罚因子 C=128, 多项式核参数 q=3、C=4, 径向基核函数参数σ=0.2、C=8, 组合核函数 C=64、σ=0.25、α=0.1、q=3.

实验一: 比较应用不同的多分类方法与不同核函数情况下的识别率, 得到实验结果表 1.

表 1 基于不同多分类方法不同核函数识别率对比

核函数	参数	SVMs(不引入核函数, 识别率为 0.735)		
		一对多	一对一	DAG
线性核	C=128	0.785	0.789	0.812
多项式核	q=2, C=2	0.832	0.836	0.864
径向基核	σ=0. 2, C=2	0.852	0.854	0.881
组合核	α=0.12, q=3, C=125, σ=0.2	0.906	0.910	0.932

由表 1 可知, 无论采用哪种核函数, DAG 方法的识别性能都高于“一对一”和“一对多”两种多分类方法 2～3 个百分点, 这是由于“一对一”和“一对多”分别存在多个得票数最多和多个最大决策函数输出值的情况, 这时容易发生误分, 使系统的识别率降低. 在不引入核函数情况下, 系统的识别率为 73.5%, 说明在各单核和组合核函数都选取最优参数的情况下, 引入 SVM 进行参数优化能提高系统的识别率. 同时可知, 组合核函数 SVM 的识别性能明显高于其他三个单核函数 SVM, 这是由于线性核函数适用于低维空间可分的情况, 但实际应用中绝大多数样本都是低维空间线性不可分的, 因此线性核函数的识别率最低. 而多项式核函数泛化能力强, 但局部学习能力弱, 径向基核函数局部性能相当优异, 但泛化能力弱, 由于说话人识别系统中输入的语音特征参数局部特性复杂, 其被估计实值函数的功率谱形状与径向基核函数较为匹配, 因此径向基核函数的识别率要高与多项式核函数. 对于组合核函数, 它继承了局部核的学习能力, 同时也融入了全局核的泛化能力, 同时由于输入语音特征参数的特性, 径向基核的比例要远大于多项式核函数, 使构造的组合核函数与被估计实值函数的功率谱形状最为匹配, 能够增加输入特征参数的可区分性, 所以在实验中也能得到最佳的识别性能.

实验二: EigenVoice 作自适应, 比较自适应时长不同时, 不同核函数的识别率情况, 实验结果见图 2.

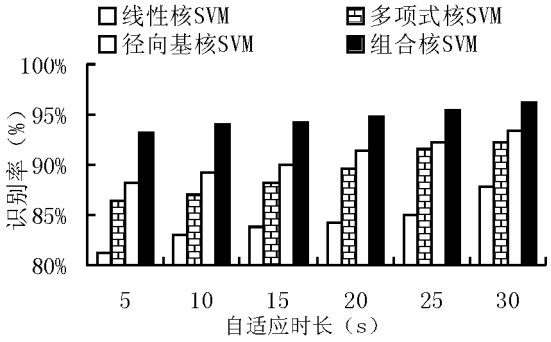

图 2　基于 EV 应用不同核 SVM 识别率对比

由图 2 可知, 从无论采用哪种核函数, 随着自适应时长的增长, 系统的识别率都会有所改进, 但组合组合数的识别率明显高于其他单核函数. 说明以 EigenVoice 作自适应, 在自适应时长不同的情况下, 组合核函数的识别性能也都优于其它三个单核 SVM.

实验三: 人工加入白噪声, 得到不同信噪比的语音, 在此基础上对比不用的核函数的识别率的差异, 比较它们对于噪声的鲁棒性, 实验结果见图 3.

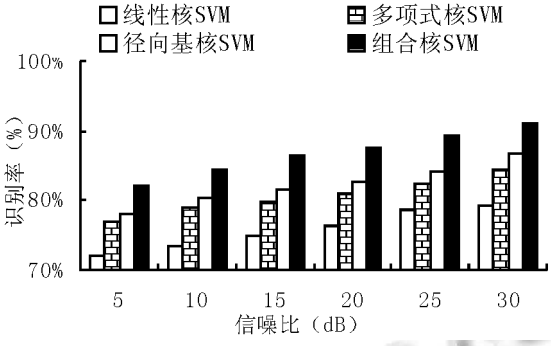

图 3　基于不同信噪比不同核 SVM 识别率对比

由图 3 可知, 系统的识别性能都因为噪音而下降, 无论采用哪些核函数, 信噪比越小, 识别性能越低. 但对于某种给定的信噪比情况下, 组合核函数的识别率要高于其他单核函数的识别率, 说明组合核函数 SVM 能够提高系统的鲁棒性.

实验四: 比较当说话人不同的情况下, 不同核函数的识别性能及其变化, 实验结果见图 4.

实验结果表明, 随着说话人人数的增长, 线性核、多项式核、径向基核和组合核 SVM 的识别率都有所下降, 但下降的幅度不大. 说明随着说话人人数的增加, 分类数增加了, 系统识别的复杂度加大了, 所以各核的识别率都有所下降. 下降幅度不大则说明 SVM 用于说话人识别是可行的, 即系统具有稳定性. 同时, 在给定的说话人人数的情况下, 组合核函数的识别率高于其它三个单核函数 1～3 个百分点, 再一次证明了组合核函数 SVM 能得到更好的识别性能.

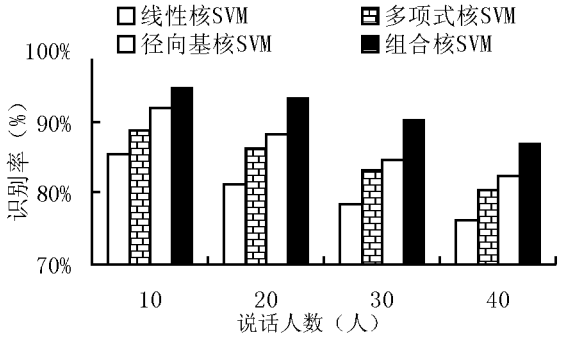

图 4　基于不同人说话人数不同核 SVM 识别率对比

4　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. 而且在自适应时长、信噪比、说话人数量不同的情况下, 表现依旧不俗. 但由于最优参数的确定对语音库具有依赖性, 所以对于差别较大的语音库, 参数如何变化还需要更多的实验来验证.

参考文献

1 吕洪艳,李荟.改进 MCE 训练算法在说话人识别中的应用.计算机系统应用,2015,24(6):143−147.
2 胡若华,改进的核函数算法及其在说话人辨认中的应用研究[硕士学位论文].北京:北京交通大学,2008.
3 胡政权.说话人识别中语音参数提取方法的研究[硕士学位论文].南京:南京师范大学,2013.
4 王秋雯.基于 GMM-UBM 的快速说话人识别方法[硕士学位论文].哈尔滨:哈尔滨工业大学,2011.
5 Theodoridis S, Koutroumbas K. 李晶皎等译.模式识别.北京:电子工业出版社,2006:45−48.
6 吕洪艳,杜鹃.基于 SVM 的不良文本信息识别.计算机系统应用,2015,24(6):183−187.