基于核Fisher判别的分类器算法及其在语种识别中的应用研究

李晋徽 杨俊安 项要杰
(电子工程学院 合肥 230037) (电子制约技术安徽省重点实验室 合肥 230037)

摘 要 GMM与SVM的建模和识别性能具有较好的互补性,因此GMM-SVM在语种识别中得到广泛使用,以其为基础的GMM-MMI-SVM已成为语种识别的主流研究方法。但是SVM在判别时仅仅使用了训练样本中的一些特殊样本即支持向量,并没有使用全部样本,从而影响了系统识别性能的进一步提高。针对上述问题,提出一种基于核Fisher判别的分类算法——GMM-MMI-KFD。该算法的核心思想是用核Fisher准则(KFD)替代SVM分类准则,从语音片段中提取出特征向量序列,分别通过GMM-MMI分类器与GMM-KFD分类器进行判决打分。相对SVM,KFD更注重语音数据非线性分布的特点,并且将样本向高维空间H上投影后可以最大限度地增大类间距,减小类内距。实验数据表明,GMM-MMI-KFD方法在语种识别中具有更高的识别率。

关键词 语种识别,核Fisher判别,分类器融合,SVM,GMM-MMI
中图法分类号 TM344.1 **文献标识码** A

Novel Classifier Algorithm Based on Kernel Fisher Discriminant and its Application in Language Recognition

LI Jin-hui YANG Jun-an XIANG Yao-jie
(Electronic Engineering Institute, Hefei 230037, China)
(Anhui Key Laboratory of Electronic Restriction, Hefei 230037, China)

Abstract GMM and SVM have a good complementation on the modeling and recognition performance. Therefore, GMM-MMI-SVM has become a mainstream research method in language recognition. However, SVM only employs some special samples in the training samples, i. e. support vector, but doesn't use all samples. This affects further improvement of system's recognition performance. In order to solve this problem, an novel classification algorithm based on Kernel Fisher Discriminant(KFD) was proposed in this paper, called GMM-MMI-KFD. The core idea is the substitution of SVM with KFD, Extracting eigenvector sequence from voice segment, and then inputing them into GMM-MMI and GMM-KFD classifiers respectively, which judge them. Compared to SVM, KFD gets more emphasis on the characteristic of nonlinear distribution of voice data. Meanwhile, it can maximize between-class space and minimize within-class space after the projection of samples onto high-dimensional space. The experimental data shows that the GMM-MMI-KFD Classifier has higher recognition rate in language recognition.

Keywords Language recognition, Kernel fisher discriminant, Classifier fusion, SVM, GMM-MMI

1 引言

语种识别是通过分析处理一个语言片段来判别其所属语言的种类。随着全球一体化进程的加快,各种语言间的信息交互日趋频繁,语种识别作为基于语音识别技术的自动翻译和信息查询等系统的前端处理,获得了广泛的应用。

近年来,越来越多的人考虑到将不同的数学建模方式相结合能取得更好的结果。文献[1-4]将音素识别器结合支持向量机(Phone Recognition followed by Support Vector Machine, PRSVM)的方法用于说话人识别以及关键词识别,但这往往需要大量训练样本。文献[5,6]使用矢量量化(Vector Quantization, VQ)和隐马尔可夫模型(Hidden Markov Models, HMM)结合的方法进行说话人识别,但识别速度有待提高。文献[7-9]将生产型(generative)的高斯混合模型(Gaussian Mixture Model, GMM)与区分型(discriminative)的SVM模型相结合,对于语音数据来说,其分布呈非线性,SVM处理非线性问题的核心思想是通过一个非线性映射ϕ将原输入空间映射到一个新的高维特征空间H中,并转化成一个线性可分问题。SVM最优分类面的法向量w是具有强判别分析能力的投影方向,一般情况下,原始空间的样本投影到法向量w形成一维空间,具有非常好的可区分性,但SVM分类器在强调分类间隔最大的同时无法考虑类内散度尽可能小的问题,并且在用SVM进行判别时仅仅使用了训练样本中的一些特殊样本即支持向量,从而影响到系统的识别性能。

考虑到求解最优分类面算法的性质,在高维空间中只需进行内积运算,所以采用Kernel技术而不必了解ϕ的具体形式。Mika[10]和Baudat[11]提出了核方法,将Fisher判别(Kernel Fisher Discriminant, KFD)方法进一步拓展到非线性情

到稿日期:2012-12-10 返修日期:2013-03-01 本文受国家自然科学基金(60872113)资助。
李晋徽(1988—),硕士生,主要研究方向为语种识别,E-mail:hhlee88223@yahoo.com.cn;杨俊安 博士生导师。

形，该方法同样引入一个非线性映射 ϕ，在新的高维特征空间 H 中采用 Fisher 线性判别方法，实现相对于原输入空间的非线性判决，并且将样本向 H 上投影后可以最大限度地增大类间距，减小类内距。本文研究了语种识别的主流研究方法 GMM-MMI-SVM，针对其中 SVM 用于判别时的不足，即仅仅使用了训练样本中的一些特殊样本即支持向量，并没有使用全部样本，从而影响了系统识别性能的进一步提高，提出了一种基于核 Fisher 判别的分类算法——GMM-MMI-KFD。该算法的核心思想是用核 Fisher 准则替代 SVM 分类准则，从语音片段中提取出特征向量序列，分别通过 GMM-MMI 分类器与 GMM-KFD 分类器进行判决打分。相对 SVM，KFD 更注重语音数据非线性分布的特点，并且将样本向高维空间 H 上投影后可以最大限度地增大类间距，减小类内距。实验数据表明，GMM-MMI-KFD 方法在语种识别中具有更高的识别率。

2 基于 GMM-MMI-SVM 的语种识别方法

GMM-MMI-SVM 基于两种模型建模方法的差异，将不同分类器相融合，具有较好的互补性，其系统框架图如图 1 所示。

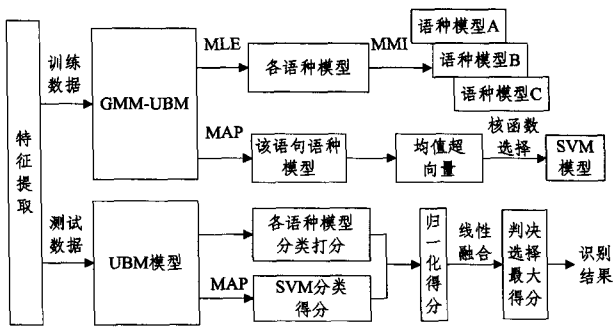

图 1 GMM-MMI-SVM 系统框图

训练阶段经 SDC(Shifted Delta Cepstra)[12] 特征提取后分别进入 GMM-MMI 和 GMM-SVM 两个系统。传统 GMM 训练是基于最大似然估计(Maximum likelihood estimation, MLE)的方法，没有充分考虑到模型间的相互影响，识别结果不是很理想。最大互信息量准则(Maximum Mutual Information, MMI)是目前语种识别领域中主流的方法[13]，输入语音与语种模型之间的互信息反映了两者的似然度，其值越大表示越相似。GMM-MMI 系统通过 MLE 准则[12] 训练出一个 UBM 模型，以 UBM 模型为基础，从中自适应出各语种的 GMM 模型作为初始模型，进一步采用 MMI 准则继续训练，迭代多次后得到最终的 MMI 模型。GMM-SVM 系统中，每条语音采用最大后验概率(Maximum a posteriori, MAP)准则，从通用背景模型 Universal background model, UBM)中自适应得到这条语音的 GMM 模型，然后将所有高斯的每一维均值向量按照顺序排列起来，构成一个超矢量作为 SVM 的输入，并通过核函数的选择训练出 SVM 模型。

测试阶段测试数据经 UBM 模型和 MMI 模型求对数似然值，相减后的值作为 GMM-MMI 得分，同时经 UBM 模型和 SVM 分类器进行打分并作为 GMM-SVM 的得分。将两类系统的得分归一化后进行线性融合，输出的最大得分为识别结果。

但是此系统中的 SVM 分类器在强调分类间隔最大的同时无法考虑类内散度尽可能小的问题。由于 SVM 在处理非线性问题时，计算的复杂度不再取决于空间维数，而是取决于样本数，尤其是样本中的支持矢量数，因此在进行判别时仅仅使用了训练样本中的某些特殊样本，从而影响到系统的识别性能。综合上述不足，本文提出了一种基于核 Fisher 判别的分类算法——GMM-MMI-KFD。

3 基于 GMM-MMI-KFD 的语种识别

3.1 KFD 准则

核 Fisher 判别的核心思想是通过一个非线性映射 ϕ 将原始空间线性不可分的样本映射到高维空间 H，转变成高维空间的线性可分情况，在新的特征空间 H 中使用 Fisher 线性判别，找出使类间散度最大而类内散度最小的投影方向进行分类，其中非线性映射通过核函数内积运算来完成。在语种识别中，每个语种都是一个基于 Fisher 准则的最优映射方向，在这个映射方向上目标语种和其他语种可以被正确地区分。若 x 为原始特征向量，则 $\phi(x)$ 可看作空间 H 的训练样本，其中 N 个语种可被分为 N 类。

H 中的目标函数为：

$$J(w)=\frac{w^{\mathrm{T}}s_b^{\phi}w}{w^{\mathrm{T}}s_w^{\phi}w} \tag{1}$$

式中，s_b^{ϕ} 和 s_w^{ϕ} 是相应的在空间 H 中的类间离散度矩阵和类内离散度矩阵，w 是投影方向。

$$s_b^{\phi}=(m_1^{\phi}-m_2^{\phi})(m_1^{\phi}-m_2^{\phi})^{\mathrm{T}} \tag{2}$$

$$s_w^{\phi}=\sum_{i=1}^{2}\sum_{x\in X}(\phi(x)-m_i^{\phi})(\phi(x)-m_i^{\phi})^{\mathrm{T}} \tag{3}$$

$$m_i^{\phi}=\frac{1}{N_i}\sum_{j=1}^{N_i}\phi(x_j^i),\ i=1\cdots N \tag{4}$$

由于空间 H 的维数很高，直接求解不太现实，因此采用了核函数的方法而不涉及到具体的非线性运算。根据核函数的理论，任何一个目标函数的解 w 都可以用特征空间中元素的线性组合表示：

$$w=\sum_{i=1}^{N}a_i\phi(x_i) \tag{5}$$

将式(4)和式(5)相乘并利用核函数 $K(x_j,x_k^i)$ 代替相应的点积$\langle\phi(x_j),\phi(x_k^i)\rangle$得：

$$w^{\mathrm{T}}m_i^{\phi}=\frac{1}{N_i}\sum_{j=1}^{N}\sum_{k=1}^{N_i}a_jK(x_j,x_k^i)=a^{\mathrm{T}}\mu_i \tag{6}$$

式中，$\mu_i=\frac{1}{N_i}\sum_{k=1}^{N_i}K(x_j,x_k^i)$ 由式(2)和式(6)可得

$$w^{\mathrm{T}}s_b^{\phi}w=a^{\mathrm{T}}Ma \tag{7}$$

$$w^{\mathrm{T}}s_w^{\phi}w=a^{\mathrm{T}}Pa \tag{8}$$

式中，$M=(\mu_1-\mu_2)(\mu_1-\mu_2)^{\mathrm{T}}$，$P=P_1+P_2$，$P_i=K_iK_i^{T}-N_i(\mu_i\mu_i^{\mathrm{T}})$，$i=1,2$。$K_i$ 为核函数矩阵，

$(K_i)_{jk}=K(x_j,x_k^i)$，$i=1,2;j=1,2,\cdots,N;k=1,2,\cdots,$

N；x_k^i 表示第 i 类第 k 个样本点。式(1)可变为：

$$J(a)=\frac{a^{\mathrm{T}}Ma}{a^{\mathrm{T}}Pa} \tag{9}$$

根据广义 Rayleigh 熵并忽略比例因子得 $a=P^{-1}(\mu_1-\mu_2)$，特征空间 H 中任一向量 $\phi(x)$ 在 Fisher 判定最优方向上的投影为：

$$w^{\phi}(x)=\sum_{i=1}^{N}a_iK(x_i,x) \tag{10}$$

选择合适的阈值可得在新的特征空间 H 中的分类判别函数为：

$$f(x)=\mathrm{sgn}[w^{\mathrm{T}}\phi(x)+b]=\mathrm{sgn}[\sum_{i=1}^{N}a_iK(x_i,x)+b] \tag{11}$$

在 KFD 中，核函数的选择十分重要，其设计得不合适直接影响到分类效果，目前较常用的核函数有线性核函数、多项式核函数、高斯径向基核函数等。

3.2 GMM-KFD 在语种识别中的应用

语种识别中，目标间的声学矢量空间重叠比较严重，传统的方法只能借助于提高 GMM 的混合数来提高系统性能，而混合数的提高则会增加运算的开销，KFD 结合 Fisher 区分性训练的优点，可以帮助减小目标之间的重叠性，加大模型间的区分性。语种识别是一种典型的多分类问题，常用的思想是将多类问题分解成多个二类问题，主要方法有一对一(OAO)法、一对余(OAR)法。OAO 法是在 K 类训练样本中构造所有可能的 2 类分类器，每个分类器仅在 K 个类别中的 2 类训练样本上训练，共构造 C_n^2 个分类器，OAR 法的判别数为 K，由于每个判别函数都需所有样本参与训练，因此计算量很大。本文将采取 OAO 法，设计一个基于核 Fisher 判别的多类分类器解决 GMM-KFD 分类问题。KFD 分类器结构如图 2 所示。

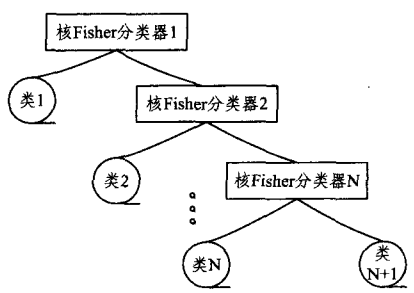

图 2 基于核 Fisher 的多分类器

GMM-KFD 是将训练出的 UBM 和 KFD 分类器结合，由于 KFD 使用了所有的训练样本，而不是少量称为“支持向量”的样本，因此 KFD 在分类准确度上优于 SVM，本文用 GMM-KFD 替代 GMM-SVM 系统。GMM-KFD 框图如图 3 所示。

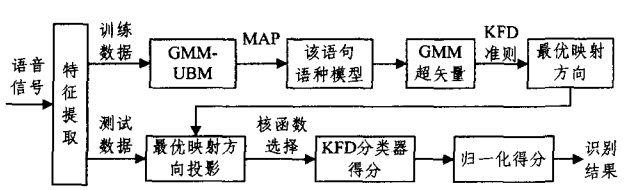

图 3 GMM- KFD 系统框图

训练数据经 SDC 特征提取后采用 MAP 准则，从 UBM 中自适应得到这条语音的 GMM 模型，之后将所有高斯的每一维均值向量按照顺序排列起来作为输入，由 KFD 准则得到最优映射方向。测试数据根据最优映射方向计算出投影，并在 KFD 分类器选择出合适的核函数后对其打分，最后归一化得出结果。

3.3 GMM-MMI 分类器与 GMM-KFD 分类器融合

信息融合是为了获得更多的信息，把不同性质和不同方法得到的信息进行组合，从而达到增强控制器控制能力、提高图像分析器的识别能力、增加跟踪器的精度以及分类器的决策能力等目的。信息融合通常可分为数据层融合、特征层融合和决策层融合 3 个层次，而决策融合是在信息表示的最高层次上进行融合处理。综合 GMM-MMI 与 GMM-KFD 不同分类方法，本文在决策层上对其进行融合，测试的语音片段提取特征后分别通过 GMM-MMI 分类器与 GMM-KFD 分类器进行判决打分。设 g_i 和 k_i 表示各语种为归一化的得分，其中 g_i 表示 GMM-MMI 系统的得分，k_i 表示 GMM-KFD 系统的得分，$i=1,2,\cdots,N$，其中 N 是语种的数量。

对 g_i 和 k_i 进行归一化，G_i 和 K_i 表示归一化后的得分：

$$G_i=g_i/\max(g_i) \tag{12}$$

$$K_i=k_i/\max(k_i) \tag{13}$$

将两者进行线性融合，其输出结果为最终判决：

$$R_i=wG_i+(1-w)K_i,0\leqslant w\leqslant 1 \tag{14}$$

判决依据 $i^*=\arg\max(R_i)$，w 为权重系数，w 从 0 逐渐变到 1，当 $w=0$ 时，单独以 GMM-KFD 为识别模型时的系统识别率；$w=1$ 时，单独以 GMM-MMI 为识别模型时的系统识别率，因此最优权重系数 w 的选择是 GMM-MMI-KFD 系统的关键。

4 实验结果及分析

实验语音库来自 NIST LRE 2007[14]，训练集 R 使用英语、法语、俄语、日语、韩语 5 个语种，包含 3 分钟至 6 分钟的语音片段共 1000 条，测试集 T 使用 10s 的语种片段 250 条且保证 $T\cap R=\varnothing$。特征提取方面采用 SDC 特征，即先提取 MFCC 特征参数，采样频率为 8kHz，各语种样本先经过预加重滤波器 $H(z)=1-0.97z^{-1}$，然后进行多帧平均，每帧长为 256 点，帧移 128 点，窗函数采用 Hamming 窗，得到前 7 阶系数(C0－C6)，MFCC 按(7,1,3,7)(N,d,P,k)扩展为 49 维特征，将 49 维 SDC 和 7 阶 MFCC 系数拼接起来，得到最终使用的 56 维 SDC 特征参数。通过 VTLN 正规、RASTA 滤波、VAD 检测、高斯化、倒谱域减均值(CMS)等预处理方法去除噪声和说话人影响等。

实验 1 在 KFD 中，核函数的选择非常重要，其设计得是否合适直接影响分类效果，目前较常使用的核函数有径向基核函数、3 阶多项式核函数、线性核函数。输入 SDC 特征数并根据 Fisher 准则得到最优映射方向，测试语料的特征参数通过最优映射方向计算出投影，最后使用 3 种典型的核函数

分别对其分类判决。其中，语种识别的评价指标由等概率错误(Equal Error Rate，EER)表示，结果如表 1 所列。

表 1　GMM-MMI- KFD 应用不同核函数的语种识别对比

核函数	EER(%)
径向基核函数	6.4%
三阶多项式核函数	8.6%
线性核函数	12.6%

由实验 1 可知，几种经典核函数对比测试中，径向基函数具有明显的性能优势，因此在后续的实验中，均采用径向基内核进行非线性分类。

实验 2　由 3.3 节可知，最优权重系数 w 的选择是 GMM-MMI-KFD 系统的关键。提取特征后分别与 UBM 模型和 MMI 模型进行对数似然率计算得分，同时，经过 Fisher 判定得到最优方向上的投影并使用分类判决函数对 Fisher 模型打分。对得分分别进行归一化后，按式(14)进行线性融合。将权系数 w 从 0 逐渐变大到 1，步长为 0.1，当 w 为 0 时，单独以 KFD 为识别模型时的系统识别率；而当 w 等于 1 时，单独以 GMM-MMI 为识别模型时的系统识别率。实验结果由图 3 所示，w 为 0.2 时融合效果为最优，因此以下 GMM-MMI-KFD 融合实验均采用最优参数 w=0.2。

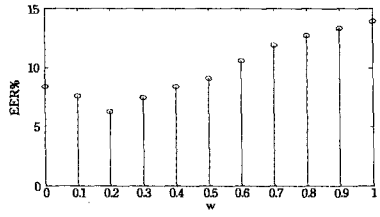

图 3　在不同权系数条件下融合系统的识别率

实验 3　应用 GMM-MMI、GMM-KFD、GMM-MMI-SVM 与 GMM-MMI-KFD 进行语种识别对比测试，GMM-MMI-SVM 与 GMM-MMI-KFD 采用相同的径向基核函数，实验结果如表 2 所列。

表 2　GMM-MMI-KFD 与其他语种识别方法对比

识别方法	EER(%)
GMM-MMI	13.8%
GMM-KFD	8.8%
GMM-MMI-SVM	7.6%
GMM-MMI-KFD	6.4%

由实验 3 可知，本文方法比 GMM-MMI 的错误率降低 6.4%，比 GMM-KFD 的错误率降低 2.4%，比 GMM-MMI-SVM 错误率降低 1.2%。

结束语　本文对多分类器融合的语种识别方法进行研究，提出一种基于核 Fisher 判别的分类算法——GMM-MMI-KFD。使用 KFD 分类器替代传统的 SVM 分类器，克服了 SVM 仅使用训练中特殊样本的缺点，且在计算样本最大类间距的同时也考虑到最小类内距，提高了分类效果。实验表明，本文方法在识别准确度方面优于 GMM-MMI、GMM-KFD、GMM-MMI-SVM 方法。

参 考 文 献

[1] Campbell W M, Campbell J P, Reynolds D A, et al. Phonetic Speaker Recognition with Support Vector Machines[C]// Advances in Neural Information Processing Systems. MIT Press, Cambridge, MA, 2004

[2] Richardson F S, Campbell W M. Language Recognition with Discriminative Keyword Selection [C]// Proc. of ICASSP 2008. Las Vegas, Nevada, U. S. A, 2008: 4145-4148

[3] Campbell W M, Richardson F, Reynolds D A. Language recognition with word lattices and support vector machines[C]// Proc of ICASSP. 2006, 11

[4] 金恬，宋彦，戴礼荣. 一种改进的 PRSVM 语种识别方法[J]. 小型微型计算机系统，2011，32(5)：1017-1020

[5] Revathi A, Venkataramani Y. Speaker independent continuous speech and isolated digit recognition using VQ and HMM[C]// International Conference on Communications and Signal Processing. Washington, DC: IEEE Computer Society, 2011: 198-202

[6] Zulfiqar A, Muhammad A, Martinez-Enriquez A M, et al. Text-Independent Speaker Identification Using VQ-HMM Model Based Multiple Classifier System[J]. Lecture Notes in Computer Science, 2010, 6438: 116-125

[7] 程杨. 基于多分类器的少数民族语种识别研究[D]. 昆明：云南大学，2012

[8] Torres-Carrasquillo P, Singer E, Gleason T, et al. The MITL LNISTLRE 2009 Language Recognition System[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. Dallas, T X, 2010: 4994-4997

[9] Campbell W. A Covariance Kernel For SVM Language Recognition[C]// IEEE International Conference on Acoustics, Speech, and Signal Processing. 2008

[10] Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels[C]// Proceedings of the IEEE International Workshop on Neural Networks for Signal Processing. Madison, USA, 1999: 41-48

[11] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach[J]. Neural Computation, 2000, 12(10): 2385-2404

[12] 徐颖. 语种识别声学建模方法研究[D]. 北京：中国科技大学，2011：13-19

[13] 付强. 基于高斯混合模型的语种识别的研究[D]. 北京：中国科技大学，2009：33-36

[14] The 2007 NIST Language Recognition Evaluation Plan[OL]. http://www.itl.nist.gov/iad/mig//tests/lre/2007/LRE07EvalPlan-v8b.pdf