第 36 卷 第 8 期 计 算 机 工 程 2010 年 4 月
Vol.36 No.8 Computer Engineering April 2010
·人工智能及识别技术· 文章编号：1000—3428(2010)08—0194—03 文献标识码：A 中图分类号：TP391

NAP 序列核函数在话者识别中的应用

邢玉娟[1]，李 明[2]
(1. 甘肃联合大学理工学院，兰州 730000；2. 兰州理工大学计算机与通信学院，兰州 730000)

摘 要：针对话者识别系统中特征向量不定长和交叉信道干扰等问题，提出一种基于超向量的扰动属性投影(NAP)核函数。该函数是一种新型的序列核函数，使支持向量机能在整体语音序列上分类，移除核函数空间中与话者识别无关的信道子空间信息。仿真实验结果表明，该函数可有效提高支持向量机的分类性能和话者识别系统的识别准确率。

关键词：扰动属性投影；高斯混合模型超向量；话者识别；支持向量机

Application of NAP Sequence Kernel Function in Speaker Verification

XING Yu-juan[1], LI Ming[2]
(1. School of Science and Engineering, Gansu Lianhe University, Lanzhou 730000;
2. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730000)

【Abstract】 For the sake of solving the problem of variable-length feature vectors and channel impact which existed in speaker verification, a novel kernel function based on Gaussian Mixture Model(GMM) supervector, called Nuisance Attribute Projection(NAP) mapping KL divergence linear kernel function, is proposed in this paper. This function can not only be in the interest of enabling Support Vector Machine(SVM) to classify on whole audio sequences, but also has the benefit that channel subspace, which causes variability, is removed in kernel space. By doing so, the classification performances of SVM and verification accuracy of system are improved excellently. Simulation experimental results demonstrate the effectiveness of this kernel function.

【Key words】 Nuisance Attribute Projection(NAP); Gaussian Mixture Model(GMM) supervector; speaker verification; Support Vector Machine(SVM)

1 概述

在文本无关话者识别[1]领域，高斯混合模型(Gaussian Mixture Model, GMM)被学者公认为一种成功的话者识别模型，获得了广泛的应用。然而 GMM 算法的训练准则是使似然度最大，而使非分类错误最小，因此，不能产生性能最佳的识别模型。

由 Vapnik 等人提出的基于统计理论的支持向量机(Support Vector Machine, SVM)以其结构简单、具有全局最优性以及较好的泛化能力等优点，已被成功地应用于话者识别中。核函数是 SVM 中的关键部分，复杂的内积运算使用核函数替代，从而降低了计算复杂度，避免了“维数灾难”。然而，不管是内积的计算还是核函数的计算，SVM 处理的都是定长向量。虽然在语音数据的预处理过程中，每一帧语音都转换为定长向量，但是语音序列整体上是不定长的，这使 SVM 只能在帧向量[2]上进行分类。而对于语音数据的每一帧来说，它所包含的判别信息很少，甚至有些帧不包含任何判别信息(如静音帧和噪音帧)，因此，SVM 并不能获得理想的分类效果。同时在话者识别过程中，语音不可避免地受到周围环境的影响，诸如背景噪声、信道畸变等，这样使话者的训练语音和测试语音不在同一信道上，从而产生信道的交叉干扰，导致系统的识别性能显著下降。

为了解决 SVM 不能在整体语音序列上进行分类的问题，同时克服信道交叉干扰的影响，使系统在信道交叉干扰情况下依然能够稳健地进行识别，本文提出一种基于 GMM 超向量的扰动属性投影(Nuisance Attribute Projection, NAP)核函数，并将其应用于话者识别系统。

2 基于 NAP 序列核函数的话者识别系统

在话者识别中，由于语音数据规模庞大，整体语音序列不定长，且容易受到噪声以及信道的干扰，导致直接应用 SVM 进行话者识别不能得到更好的识别效果。如何使 SVM 可以在整体语音序列上进行分类，以及如何消除信道的干扰是话者识别研究的关键问题。在对高斯混合模型超向量深入研究的基础上，本文提出一种基于 NAP 映射的 KL 散度线性核函数，并将其应用于话者识别系统，系统框图如图 1 所示。

首先根据话者的语音建立一个高斯混合通用背景模型(GMM-UBM)，然后只对均值向量采用最大后验概率(Maximum A Posteriori, MAP)自适应得到每个话者的高斯混合模型超向量。紧接着在超向量空间，计算话者 GMM 模型间 KL 散度距离得到 KL 散度核函数。对此 KL 散度核函数进行 NAP 映射处理，根据主成分分析法估计信道空间的方法，

作者简介：邢玉娟(1981－)，女，讲师、硕士研究生，主研方向：生物特征识别，智能信息处理；李 明，教授
收稿日期：2009-10-20 **E-mail：**xyj19811010@126.com

移除核函数空间中的信道干扰子空间，生成一种新的 NAP 映射核函数。将该核函数直接应用到 SVM 中进行话者的识别。

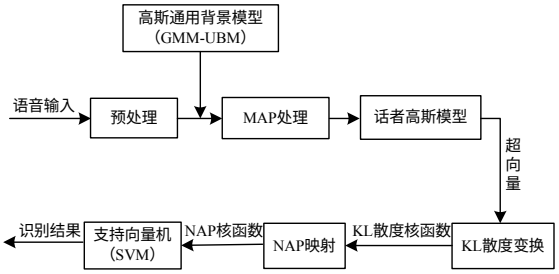

图 1　基于 NAP 核函数的话者识别系统框图

2.1　高斯混合模型超向量

超向量是文献[3]提出的一种新的特征向量，在其生成过程中，GMM-UBM 的建立是关键。在本文中的话者识别系统中，采用所有注册者的语音数据，根据 EM 算法训练建立 GMM-UBM。假设得到的 GMM-UBM 为

$$g(\boldsymbol{x})=\sum_{i=1}^{M}\omega_i N(\boldsymbol{x};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i) \tag{1}$$

其中，$N(\boldsymbol{x};\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)$ 为语音向量 $\boldsymbol{x}$ 的高斯密度函数，ω_i, $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$ 分别是第 i 个高斯分量的混合权重、均值向量和协方差矩阵。由于在模型的计算过程中涉及到协方差矩阵的逆，而矩阵的求逆运算非常耗时，因此本文采用对角化的协方差矩阵形式。

对于话者的 d 维输入语音序列 $X=\{x_1, x_2, \cdots, x_n\}$，GMM-UBM 只对均值向量进行 MAP 自适应得到每个话者的高斯混合模型超向量 $\boldsymbol{\mu}=[\mu_1, \mu_2, \cdots, \mu_d]^{\mathrm{T}}$。在这种情况下，所有话者的高斯模型具有相同的协方差矩阵和权重向量。超向量的生成过程如图 2 所示。

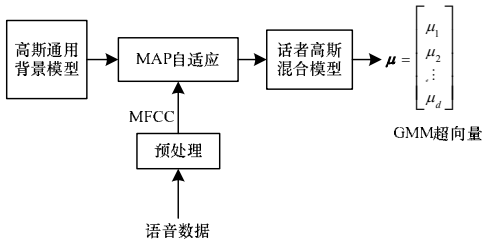

图 2　GMM 超向量的形成过程

由图 2 可知，GMM 超向量的生成可看成是输入语音向量映射到某一高维向量，且每个话者的 GMM 超向量的长度是固定的。

2.2　KL 散度线性核函数

KL 散度线性核函数[3]是通过对话者高斯混合模型之间的 KL 散度距离进行分析，得到的一种新的语音序列核函数。由于每个话者高斯混合模型的协方差矩阵和权值向量相同，因此反应模型间 KL 散度距离的只有超向量。假设话者 S_a 和 S_b 对应的高斯混合模型为 g_a 和 g_b，则两者模型间的对称 KL 散度为

$$D(g_a,g_b)=D(g_a\|g_b)+D(g_b\|g_a)=$$
$$\int_{R^n} g_a(x)\mathrm{lb}\left(\frac{g_a(x)}{g_b(x)}\right)\mathrm{d}x+\int_{R^n} g_b(x)\mathrm{lb}\left(\frac{g_b(x)}{g_a(x)}\right)\mathrm{d}x \tag{2}$$

通过对称 KL 散度计算得到的距离矩阵不是一个正定矩阵，不满足 Mercer 条件，因此，采用对称 KL 散度的近似上限：

$$D(g_a\|g_b)\le D(\omega^a\|\omega^b)+\sum_{i=1}^{N}\omega_i^a D(N(x^a;\mu_i^a,\Sigma_i^a)\|N(x^b;\mu_i^b,\Sigma_i^b)) \tag{3}$$

由于话者模型的权重和协方差矩阵相等，对称 KL 散度的近似上限可简化为

$$D(g_a,g_b)\le\sum_{i=1}^{N}\omega_i(\mu_i^a-\mu_i^b)\boldsymbol{\Sigma}_i^{-1}(\mu_i^a-\mu_i^b)^t \tag{4}$$

式(4)中协方差矩阵采用对角化的形式，因此，模型间 KL 散度的上限实质上表示的是加权欧几里德距离。则 KL 散度线性核函数为

$$\underset{linear}{K}(S_a,S_b)=\sum_{i=1}^{N}\omega_i\mu_i^a\boldsymbol{\Sigma}_i^{-1}(\mu_i^b)^t=$$
$$\sum_{i=1}^{N}(\sqrt{\omega_i}\,\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\mu_i^a)(\sqrt{\omega_i}\,\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\mu_i^b)^t \tag{5}$$

令 $m_i^a=\sqrt{\omega_i}\,\boldsymbol{\Sigma}_i^{-\frac{1}{2}}\mu_i^a$，则

$$\underset{linear}{K}(S_a,S_b)=\sum_{i=1}^{N}m_i^a(m_i^b)^t \tag{6}$$

式(6)满足 mercer 条件，生成的 KL 散度线性核函数是将超向量通过 $\sqrt{\omega_i}\,\boldsymbol{\Sigma}_i^{-\frac{1}{2}}$ 映射到一个新的高维欧几里德空间来实现的。这是一种新型的序列核函数，它在替代内积运算的同时将不定长的向量转换成定长，使支持向量机可以在整体语音序列上进行分类。然而，在文本无关的话者识别过程中存在交叉信道干扰问题，导致系统性能的显著下降。为了克服这个问题，对生成的 KL 散度线性核函数进行 NAP 映射，在核函数空间移除引起信道干扰的与话者识别无关的子空间，从而生成一种新的 NAP 映射序列核函数。

2.3　NAP 映射

NAP 映射是由 Campbell W 等人提出的一种新的解决信道干扰问题的方法，其核心思想是在核函数空间移除与话者识别无关的信道子空间。如何估计信道子空间是 NAP 映射的关键问题。Campbell W 等人提出了一种通过求解最优化问题来估计信道子空间的方法，但这种方法涉及到多个矩阵的相乘及求逆，计算复杂度很高，导致系统实时性急剧下降。通过对超向量中信道变化的分析，本文采用主成分求解的方法估计信道空间。

假设系统共有 S 个注册的话者，每一个话者的超向量为 $\boldsymbol{\mu}(s)=\{\mu_1(s), \mu_2(s), \cdots, \mu_n(s)\}$, $s=1,2,\cdots,S$；$\overline{\boldsymbol{\mu}}(s)=\frac{1}{n}\sum_{i=1}^{n}\mu_i(s)$ 表示每个话者超向量的平均向量。如果话者的语音都受到信道的干扰，则话者超向量的平均向量可以认为是与信道无关的[4]，这样每个话者的超向量由信道干扰引起的偏移向量可计算如下：

$$\Delta\mu_i(s)=\mu_i(s)-\overline{\mu}(s) \tag{7}$$

通过对话者超向量的偏移变化的分析，可用类间离散度矩阵描述超向量受信道影响的偏移程度：

$$\boldsymbol{C}=\frac{1}{S\times n}\sum_{s=1}^{S}\sum_{i=1}^{n}\Delta\mu_i(s)(\Delta\mu_i(s))^t \tag{8}$$

式(8)为类间离散度矩阵，求解 $\boldsymbol{C}v=\lambda v$ 的特征值 λ 和特征向量 $\boldsymbol{v}$，将特征值和特征向量按照由大到小的顺序排列，只取前 q 个特征向量。$\eta_i=\sum_i\Delta\mu_i(s)v_i^t$ 则是信道子空间中的一个向量，对其规整化 $\widetilde{\eta}_i=\eta_i/|\eta_i|$，可得 $P=\sum_i\widetilde{\eta}_i\widetilde{\eta}_i^t$ 表示超向量空间中受信道交叉干扰的信道子空间。确定信道子空间之后，就可以对 KL 散度线性核函数进行 NAP 映射。

对于 KL 散度线性核函数 $\underset{linear}{K}(S_a,S_b)=m^a(m^b)^t$，经过 NAP 映射后为

$$\underset{NAP}{K}(S_a,S_b)=Um_a(\boldsymbol{U}m_b)^t=m_a\boldsymbol{U}(m_b)^t=$$
$$m_a(I-P)(m_b)^t=m_a(m_b)^t-m_aP(m_b)^t \tag{9}$$

其中，$\boldsymbol{U}$ 是将核函数映射到无信道干扰空间的矩阵，且满足 $\boldsymbol{U}^2=\boldsymbol{U}$；$P$ 为将要移除的信道子空间。

将 $P=\sum_i \tilde{\eta}_i \tilde{\eta}_i^t$ 代入式(9)可得：

$$\begin{aligned} K_{NAP}(S_a,S_b) &= K_{linear}(S_a,S_b) - m_a\tilde{\eta}\tilde{\eta}^t(m_b)^t = \\ & K_{linear}(S_a,S_b) - m_a\tilde{\eta}(m_b\tilde{\eta})^t = \\ & K_{linear}(S_a,S_b) - K_{Chanelspace}(\widetilde{m_a},\widetilde{m_b}) \end{aligned} \tag{10}$$

上式明显满足 Mercer 条件，其中，$\widetilde{m_a}=m_a\tilde{\eta}$。从式(10)中可看出，经过 NAP 映射得到的新核函数是 KL 散度线性核函数移除与信道相关的子空间核函数部分，只保留与信道无关空间的核函数部分，从而将核函数映射到了一个抗信道干扰的空间，将其应用到基于 SVM 的话者识别系统中能在很大程度上提高系统的识别性能。

3 仿真实验及结果分析

为了测试本文所提出的 NAP 映射核函数应用到话者识别中的性能，采用 Matlab6.5 以及 SVM Toolbox 1.0 进行仿真实验。

3.1 语音库及参数设置

实验中采用自己录制的语音库，录音的人数为 53 人，其中男 26 人，女 27 人，采用普通的话筒进行录音，录音后的数据通过采样频率为 11 025 Hz，量化位数为 16 bit，单声道的 A/D 转换成数字信号存储。为反映话者的发声随时间变化，录入是间隔一段时间多次进行的，每人都录制了 10 个语音段，每个语音段为 30 s，随机选取 5 个语音段用于训练，剩余 5 个语音段用于测试。得到话者的语音段之后，对其进行预处理。采用一阶数字滤波器 $H(z)$=1-0.95z^{-1} 对语音信号预加重，采用帧长 30 ms、帧移 15 ms 为语音信号分帧和加汉明窗。每个话者的每个语音段共有 1 999 帧，对每帧数据提取 13 维的 MFCC 及它的一阶差分和二阶差分共同构成 39 维输入特征向量，则每个话者用于训练的输入语音向量共有 5×1 999=9 995 个，构成一个 39×9 995 的矩阵。采用所有 53 个人的语音数据通过 EM 算法训练一个 1 024 阶的高斯通用背景模型。话者的每个语音段根据均值 MAP 自适应处理生成 39 维的超向量作为 SVM 的输入。

3.2 实验结果比较

为了测试本文提出的新核函数的性能，将传统核函数如线性核函数、三阶多项式核函数以及径向基核函数[5]($\sigma=1.6$)和 KL 散度线性核函数、基于 NAP 映射的 KL 散度线性核函数分别应用到支持向量机中进行话者的识别测试，不仅验证了 KL 散度线性核函数的性能，同时也测试了 NAP 映射的有效性。仿真试验以等错误率(Equal Error Rate, EER)和最小决策代价函数(minimum Decision Cost Function, minDCF)值作为衡量分析实验结果的标准。表 1 为不同核函数应用到支持向量机中的 EER 和 minDCF 比较。

表 1 不同核函数 EER 和 minDCF 比较

核函数	EER/(%)	minDCF
线性核函数	11.23	0.055 8
三阶多项式核函数	9.86	0.049 2
RBF 核函数	7.57	0.032 2
KL 散度线性核函数	5.33	0.020 1
NAP 映射 KL 散度线性核函数	4.56	0.014 3

从表 1 中可以明显地看出，KL 散度线性核函数和传统的线性核函数比较 EER 和 minDCF 都有大幅的下降，其中 EER 下降了 6.67%，minDCF 下降了 0.041 5；与目前最常用的 RBF 核函数相比，EER 的下降超过了 2%，minDCF 下降了 0.012 1。可见，序列核函数(KL 散度线性核函数)能有效地应用到 SVM 中进行语音序列的整体分类，大幅提高 SVM 的分类性能。基于 NAP 映射的 KL 散度线性核函数是 5 种核函数中性能最好的，和 KL 散度线性核函数比较 EER 下降了 0.77%，同时 minDCF 下降了 0.005 8，这表明基于 NAP 映射的 KL 散度线性核函数不仅具有 KL 散度线性核函数的优点，而且可以移除信道干扰信息，从而有效地提高话者识别系统的识别准确率。图 3 是 5 种核函数的 DET 曲线比较，倾斜向上的虚线是由错误拒绝率(FR)和错误接受率(FA)相等的点构成。

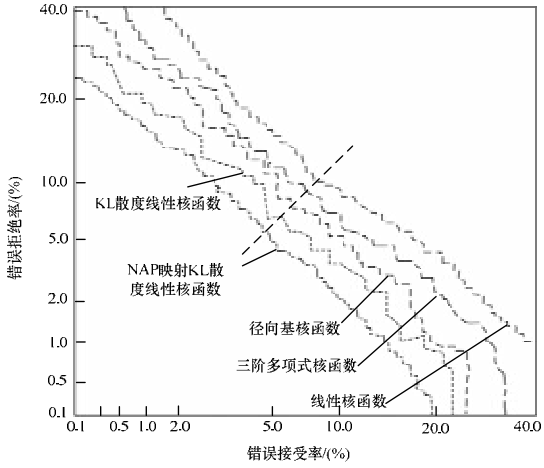

图 3 5 种核函数 DET 曲线比较

4 结束语

在深入研究 GMM 超向量的基础上，本文提出一种基于 NAP 映射的 KL 散度线性核函数。这是一种新型的序列核函数，不仅可以使得 SVM 在整体语音序列上进行分类，同时在核函数空间移除了与话者识别无关的信道子空间，有效地提高了话者识别系统的识别准确率。在 NAP 映射过程中，信道子空间的估计是关键，本文在研究了信道偏移的基础上，采用基于主成分求解的估计方法。仿真实验结果验证了这种新型序列核函数的有效性及可行性。但由于在进行 KL 散度变换和 NAP 映射的过程中，存在大量的矩阵运算，因此该核函数的计算复杂度较高，在后续的工作中将会针对这个问题进行研究，以期对核函数的求解过程进行简化。

参考文献

[1] 钱 博, 唐振民, 李燕萍, 等. 基于背景噪声估计的话者识别算法[J]. 计算机工程, 2008, 34(14): 14-16.

[2] Wan V, Rends S. Speaker Verification Using Sequence Discriminant Support Vector Machines[J]. Speech and Audio Processing, 2005, 13(2): 203-210.

[3] Campbell W M, Sturim D E, Reynolds D A, et al. SVM Based Speaker Verification Using a GMM Super Vector Kernel and NAP Variability Compensation[C]//Proc. of ICASSP'06. Toulouse, France: IEEE Press, 2006: 97-100.

[4] Vogt R, Sridharan S. Explicit Modeling of Session Variability for Speaker Verification[J]. Computer Speech and Language, 2008, 22(1): 17-38.

[5] 罗 瑜, 李 涛, 王丹琛, 等. 支持向量机中核函数的性能评价策略[J]. 计算机工程, 2007, 33(19): 186-187.

编辑 金胡考