第37卷 第14期 计 算 机 工 程 2011年7月
Vol.37 No.14 Computer Engineering July 2011

·人工智能及识别技术· 文章编号：1000—3428(2011)14—0208—03 文献标识码：A 中图分类号：TP18

基于保局部核 RVM 的说话人识别方法

郑泽萍，王万良，郑建炜

(浙江工业大学计算机科学与技术学院，杭州 310023)

摘 要：针对说话人语音特征随音量、情绪、健康等因素变化呈现出的复杂分布结构，提出一种基于保局部核相关向量机(RVM)的说话人识别方法。在 RVM 模型所采用的高斯核函数中引入相似度因子，以保留数据局部结构，构成保局部核 RVM 模型。在模型训练过程中采用快速算法以避免大型矩阵逆操作，减少计算量，可适用于大样本场合。应用结果表明，该方法能加快测试速度，提高分类精度。

关键词：说话人识别；保局部核；相关向量机；高斯核函数；类内相似度

Speaker Recognition Method Based on RVM Using Locality Preserving Kernel

ZHENG Ze-ping, WANG Wan-liang, ZHENG Jian-wei

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

【Abstract】Taking account of the complex structure of the speech features, which is affected by the change of volume, emotion, health and other factors, a new method for Speaker Recognition(SR) based on Relevance Vector Machine(RVM) using locality preserving kernel is proposed. RVM using locality preserving kernel introduces intra-class similarity into Gaussian kernel function to keep the data set's neighborhood structure, and is applied into SR. For the purpose of avoiding the inverse matrix operation and applying to a larger sample, the new method uses a fast algorithm for training. Experimental results show that the new classifier model speeds up the test speed and improves the classification accuracy.

【Key words】Speaker Recognition(SR); locality preserving kernel; Relevance Vector Machine(RVM); Gaussian kernel function; intra-class similarity

DOI: 10.3969/j.issn.1000-3428.2011.14.070

1 概述

目前说话人匹配模型主要包括基于概率统计的隐马尔科夫模型(Hidden Markov Model, HMM)和高斯混合模型(Gaussian Mixture Model, GMM)。这 2 个模型统称为生成性模型，在识别率和鲁棒性方面具有一定优势，但需要大量训练样本优化模型参数。文献[1]在基于统计学习理论的结构风险最小化原则上提出支持向量机(Support Vector Machine, SVM)模型。该模型已证实能满足高识别率要求，受到说话人识别研究者的认可。

但 SVM 本身的一些局限性，如：核函数受梅西定理的限制，需设置惩罚因子，有效向量(支持向量)数目基本上随训练数据集的规模线性增长，输出为一个点估计而不是概率形式等，影响了模型识别性能。

文献[2]在贝叶斯框架的基础上提出了相关向量机(Relevance Vector Machine, RVM)，RVM 弥补了上述 SVM 的缺点，具有强大的泛化能力，得到的模型更加稀疏，从而测试时间更短，提高了匹配模型性能。因此，本文在说话人识别中采用基于 RVM 的模型，并采用常用的高斯核函数化解说话人识别这个非线性问题。

语音样本数据本身受音量、情绪、健康等影响，呈复杂分布结构，因此，本文将考虑类内相似度特征，用来保持局部结构，并将其引入到核函数。在构建匹配模型中，如何根据具体样本数据集选择恰当的核函数参数是一个重要问题。目前常用的方法包括十字交叉验证法、拟牛顿法、遗传算法等[3]。十字交叉法精度高，但计算量大；拟牛顿法，虽然避免了人为经验不足的缺点，但它受初始值影响大；遗传算法使用方便、鲁棒性强，但受初始值、适应度函数的影响较大。为解决核参数调整问题，保持样本集的局部特征，提高说话人识别模型的识别率，本文引入类内相似度。同时为适应大样本训练情况，本文在模型训练中采用快速算法，从而避免矩阵的逆操作，减少计算量。

2 RVM 分类模型

RVM 应用广泛，可以解决回归问题，如时间序列预测、物体的跟踪等，也可以用于分类问题，如人脸识别、神经活动分类、汽车车型识别等。本文利用 RVM 分类原理将其应用于说话人辨别来提高识别率。

2.1 RVM 分类器原理

在分类问题中，RVM 是一个二元分类器。设现有训练样本集 $\{(\boldsymbol{x}_n,t_n)\}_{n=1}^N$，$\boldsymbol{x}_n$ 是输入特征向量，$t_n\in\{0,1\}$ 是模型识别结果，其概率可表示为：

$$p(t_n\mid\boldsymbol{x}_i,\boldsymbol{\omega})=\sigma(y)^{t_n}[1-\sigma(y)]^{1-t_n} \tag{1}$$

其中，向量 $\boldsymbol{\omega}=[w_1,w_2,\mathrm{L}\ ,w_N]^{\mathrm{T}}$ 是权值参数。模型输出值的表达式为：

$$y(\boldsymbol{x}_i)=\sum_{j=1}^{N}w_iK(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\varphi}(\boldsymbol{x}_i) \tag{2}$$

基金项目：国家自然科学基金资助项目(61070043)

作者简介：郑泽萍(1985－)，女，硕士研究生，主研方向：语音识别，模式识别；王万良，教授、博士、博士生导师；郑建炜，博士研究生

收稿日期：2010-12-10 **E-mail：**zpzheng321@163.com

其中，$\boldsymbol{\varphi}(\boldsymbol{x}_i)=[K(\boldsymbol{x}_i,\boldsymbol{x}_1),K(\boldsymbol{x}_i,\boldsymbol{x}_2),\mathrm{L}\ ,K(\boldsymbol{x}_i,\boldsymbol{x}_N)]^{\mathrm{T}}$ 是对应训练样本 $\boldsymbol{x}_i$ 的核函数列；$\boldsymbol{\Phi}=[\boldsymbol{\varphi}(\boldsymbol{x}_1),\boldsymbol{\varphi}(\boldsymbol{x}_2),\mathrm{L}\ ,\boldsymbol{\varphi}(\boldsymbol{x}_N)]$ 是核矩阵；$\sigma(y)=1/(1+\mathrm{e}^{-y})$ 是连接函数，把 y 从 $(-\infty,+\infty)$ 映射到 $\{0,1\}$。则可以得到整个样本集的最终似然函数：

$$p(\boldsymbol{t}\mid\boldsymbol{\omega})=\prod_{n=1}^{N}\sigma(y)^{t_n}[1-\sigma(y)]^{1-t_n} \tag{3}$$

通过最大化似然比函数方法估计出的最优权值 $\boldsymbol{\omega}$，不可避免会导致过学习现象。为避免这一现象，对权值 $\boldsymbol{\omega}$ 加上一个先验的条件概率分布——student-t 分布。student-t 分布可分两级先验概率分布进行描述：

第 1 级是参数 $\boldsymbol{\omega}$ 的先验：

$$P(\boldsymbol{\omega}\mid\boldsymbol{\alpha})=\prod_{i=1}^{N}N(w_i\mid 0,\alpha_i^{-2}) \tag{4}$$

其中，$\boldsymbol{\alpha}=\mathrm{diag}(\alpha_1^{-2},\ \alpha_2^{-2},\mathrm{L}\ ,\alpha_N^{-2})$ 是模型相应参数 $\boldsymbol{\omega}$ 的超参数。

第 2 级是超参数 $\boldsymbol{\alpha}$ 的先验分布：

$$P(\boldsymbol{\alpha}\mid a,b)=\prod_{i=1}^{N}\alpha_i^{a-1}\mathrm{e}^{-b\alpha_i} \tag{5}$$

为减少模型训练的额外计算，一般选择 $a=b=0$。先验 student-t 分布具有重尾分布特性，使模型产生了稀疏性。

由于 $P(\boldsymbol{\omega}\mid\boldsymbol{t},\boldsymbol{\alpha})\propto P(\boldsymbol{t}\mid\boldsymbol{\omega})P(\boldsymbol{\omega}\mid\boldsymbol{\alpha})$，因此对模型添加先验 student-t 分布后，最终目标函数可表示为：

$$L=\ln\{p(\boldsymbol{t}\mid\boldsymbol{\omega})p(\boldsymbol{\omega}\mid\boldsymbol{\alpha})\}=\sum_{i=1}^{N}[t_i\ln\sigma_i+(1-t_i)\ln(1-\sigma_i)]-\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\omega} \tag{6}$$

其中，$\boldsymbol{\Omega}=\mathrm{diag}(a_0,a_1,\mathrm{L}\ ,a_N)$。使用高斯函数逼近参数 $\boldsymbol{\omega}$ 的后验概率 $p(\boldsymbol{\omega}\mid\boldsymbol{t},\boldsymbol{\alpha})$，则该高斯中心 $\mu=w_{MP}=\boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{t}$，其协方差为 $\boldsymbol{\Sigma}=(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Phi}+\boldsymbol{\Omega})^{-1}$，式中的 $\boldsymbol{B}=diag(b_1,b_2,\mathrm{L}\ ,b_N)$；$b_i=\sigma_i(1-\sigma_i)$。

2.2　训练算法

原始的 RVM 算法采用自上向下建立模型，即把所有基向量作为相关向量，随着参数不断迭代，逐渐削减相关向量达到最后的高度稀疏性。由于本文所用样本比较大，若采用此方法，计算量将相当巨大，因此本文采用快速边缘最大化算法[4]，即自下向上建立模型。

快速算法首先初始化一个空模型，然后在迭代过程中添加相关向量，并修改它们所占的比重，从而大幅减少计算量。

目标函数 L 的优化对象是模型参数 $\boldsymbol{\omega}$，而 $\boldsymbol{\omega}$ 的确定首先需要获取合理的超参数 $\boldsymbol{\alpha}$ 的先验值，采用最大化边缘似然函数对参数 $\boldsymbol{\alpha}$ 迭代更新，即：

$$L(\boldsymbol{\alpha})=\ln p(\boldsymbol{t}\mid\boldsymbol{a})=\ln\int p(\boldsymbol{t}\mid\boldsymbol{\omega})p(\boldsymbol{\omega}\mid\boldsymbol{a})=-\frac{1}{2}[N-\ln 2\pi+\ln|\boldsymbol{C}|+\boldsymbol{t}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{t}] \tag{7}$$

其中，$\boldsymbol{C}=\boldsymbol{B}+\boldsymbol{\Phi}\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}$。分解矩阵 $\boldsymbol{C}$ 得到：

$$\boldsymbol{C}=\boldsymbol{B}+\boldsymbol{\Phi}\boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}=\boldsymbol{B}+\sum_{n\neq i}a_n^{-1}\boldsymbol{\varphi}_n\boldsymbol{\varphi}_n^{\mathrm{T}}+a_i^{-1}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathrm{T}}=\boldsymbol{C}_{-i}+a_i^{-1}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathrm{T}} \tag{8}$$

其中，$\boldsymbol{C}_{-i}$ 表示 $\boldsymbol{C}$ 中第 i 个基向量削减后的矩阵。分解式(7)得到：

$$\begin{aligned}L(\boldsymbol{\alpha})&=-\frac{1}{2}[N\ln(2\pi)+\ln|\boldsymbol{C}_{-i}|+\boldsymbol{t}^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{t}-\ln a_i-\frac{(\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{t})^2}{\alpha_i+\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_i}+\ln(a_i+\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_i)]=\\&L(a_{-i})+\frac{1}{2}[\ln a_i-\ln(\alpha_i+s_i)+\frac{q_i^2}{\alpha_i+s_i}]=\\&L(a_{-i})+l(a_i)\end{aligned} \tag{9}$$

其中，$s_i=\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_{-i}$ 是稀疏因子，作为一个衡量基向量是否已在模型中存在的依据；因子 $q_i=\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{t}$ 是某一个基向量排除在模型外的误差调整尺度。$L(a_{-i})$ 表示消除第 i 个输入向量后的对数似然函数，并可以根据 a_i 得到惟一的最大值：

$$a_i=\frac{s_i^2}{q_i^2-s_i}\qquad \text{if}\ \ q_i^2>s_i \tag{10}$$

$$a_i=\infty\qquad \text{if}\ \ q_i^2\leqslant s_i \tag{11}$$

如果当前 $a_i=\infty$，而更新的 $a_i^{\mathrm{new}}<\infty$，则模型中添加 $\boldsymbol{\varphi}_i$，且 a_i 依式(10)赋值；如果 $a_i<\infty$ 且 $a_i^{\mathrm{new}}=\infty$，则将 $\boldsymbol{\varphi}_i$ 从模型中删除，并且设置 $a_i=\infty$；如果 $a_i<\infty$ 且 $a_i^{\mathrm{new}}<\infty$，则仍然保留 $\boldsymbol{\varphi}_i$，且依式(10)修正 a_i。

根据上述分析，快速算法的优化求解过程是针对每一个输入矢量进行添加、消除或参数重估操作，避免了进行大矩阵求逆运算，同时每一步迭代都保证边缘似然目标函数的递增，确保迭代的收敛性。

3　保局部特征的核函数

3.1　核函数的选择

不同的核函数对 RVM 分类模型的推广性能有重要影响，因此，如何选择恰当的核函数也是 RVM 的一个重要难题。常见的核函数包括高斯核、多项式(Polynomial)核、Sigmoid 核函数等。文献[5]提出了 NAP 序列核函数，使话者识别无关的信道子空间被移除，提高了识别准确率，但新核增加了计算复杂度。在常用核函数中，高斯核具有较好的学习能力，适用于高/低维、大/小样本集，并且多项式核是高斯核的特例[6]，在有些情况下，Sigmoid 核同高斯核具有相似的性能[7]。而且高斯核参数比常用核的参数少，可减轻核参数的训练压力。因此，本文采用应用最广泛的高斯核函数，其表达式为：

$$K_{ij}=\exp(-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{\sigma^2}) \tag{12}$$

其中，σ 值是经过多次验证确定的。可以用十字交叉法、拟牛顿法等。目前对选择核参数的研究还在不断改进，文献[8]提出利用变异函数估计高斯核函数宽度参数；文献[9]提出利用分段函数对参数 σ 进行选择的方法，该方法可以获得最优参数但需要预先指定参数范围。

3.2　保局部核

说话人样本数据集有一定的分布结构，因此，可利用类内相似度[10]来表达原有局部特征。根据 k-近邻法，相似度可表示为：

$$S_{ij}=\exp(-\frac{\|\bar{\boldsymbol{x}}_i-\bar{\boldsymbol{x}}_j\|^2}{\sigma^2}) \tag{13}$$

其中，σ 是一个经验值。为使相似度矩阵 $\boldsymbol{S}=\left(S_{ij}\right)$ 具有稀疏性，只考虑样本同类数据点的 k-近邻点相应的相似度，而非 k-近邻点的相似度为 0。即在样本空间中，样本 $\boldsymbol{x}_i$ 根据欧式距离寻找其最近邻的 k 个样本点，若与 $\boldsymbol{x}_i$ 是同类样本则记第 k 个近邻距离为 N_i^k，若 $\boldsymbol{x}_j(j\neq i)$ 与 x_i 的距离小于 N_i^k 且属于同类，则认为 x_j 属于 x_i 的近邻，取相似度如下：

$$S_{ij}=\exp(-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{N_i^k\times N_j^k}) \tag{14}$$

而其余 $S_{ij}=0$，一般 $k=7$ 时能获得较优的结果[11]。这种方法得到的类内相似度矩阵具有稀疏性，能大幅提升计算速度，同时保持了类内结构，提高了分类精度。

本文将类内相似度引入 RVM 的目标函数中，发挥其保持近邻样本点结构的作用，则式(2)调整为：

$$y(\boldsymbol{x}_i,\boldsymbol{\omega})=\sum_{j=0}^{N}w_iS_{ij}K(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{15}$$

分析上式可知，类内相似度的引入只是改变了核函数的值。观察高斯核表达式(12)，当类内相似度取式(13)时，两者在形式上是完全一致的。因此，式(2)调整为：

$$y(\boldsymbol{x}_i,\boldsymbol{\omega})=\sum_{i=1}^{N}w_iK^2(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{16}$$

则目标函数中相应的输出值表达式直接用式(16)替换即可。得到的目标函数表达式的基本框架并没改变，无论采用式(13)还是式(14)作为相似度计算方法，都保持了原 RVM 的所有优点：概率输出，解的稀疏性等。类内相似矩阵 $\boldsymbol{S}$ 是对称的，所以，从核矩阵角度看，高斯核矩阵的结构保持不变，主对角元素仍为 1，其他矩阵元素只是随类内相似度进行了调整。即本文根据相似度构造出一个新核函数，使训练样本数据局部结构得到恰当的保留，而零星分散的数据点影响力得到减弱，RVM 分类精度得到提升。并且模型的空间复杂度和时间复杂度都没增加，不需额外负担。

4 实验分析

本文实验采用笔者录制的语料库，共包含 20 个说话人，(10 位男性，10 位女性)。每个人以 1 个月为间隔分 3 个时期录取，每时期剪取 5 s 纯语音作为训练样本，共 15 s 时长。测试样本取自随机时间，时长仅为 1.5 s，每个人 20 个短语，共 400 个测试语音。由于 RVM 是二元分类器，因此本文采用“一对一”方法构造分类器，共需 $C_{20}^2=190$ 个子分类器。

实验先对语料库语音用端点检测进行静音切除，再用预加重滤波 $H(z)=1-0.95z^{-1}$ 对高频端信号进行增强，用 30 ms 汉明窗进行分帧处理，帧移 20 ms，语音特征参数采用 12 阶 Mel 倒谱系数(不包括 0 阶)，并进行一阶差分，共组成 24 维语音特征。

本文对识别精度相当高的 SVM 模型与 RVM 模型进行比较，两者均采用高斯核函数。实验结果如表 1 所示。可见，RVM 的识别率与 SVM 相当，但测试时间明显比 SVM 更短，即测试速度更快，适用于实时性要求高的场合。这主要是由于 RVM 的有效向量(相关向量)比 SVM 的有效向量(支持向量)少得多，即 RVM 模型更加稀疏。因此 RVM 分类器的性能更佳，比 SVM 更适用于说话人分类模型。

表 1 RVM 和 SVM 的性能比较

模型	测试时间/s	有效向量	识别率/(%)
RVM	0.13	105	98.75
SVM	0.35	478	98.75

本文提出了一种保局部核函数，为证实该新核在常用核中的优越性，在 RVM 模型中分别采用多项式核、Sigmoid 核、高斯核、保局部核来观察分类器的精确度，并且分为 5 个、10 个、20 个人进行分别训练测试，实验结果如表 2 所示。可见，在 4 种核函数对比测试中，随着人数的增加，基于保局部核的 RVM 模型精度最高，明显优于其他核函数。由此可证，实类内相似度的引入能更好地利用原数据分布结构这一先验信息保留样本数据的分布结构，从而更精确地对说话人进行识别。

表 2 RVM 采用不同核函数时的精确度比较 (%)

人数	精确度			
	多项式核	Sigmoid 核	高斯核	保局部核
5	96.5	97.5	100.00	100.00
10	90.5	95.0	99.50	99.75
20	86.0	92.5	98.75	99.25

5 结束语

本文提出一种保局部核 RVM 分类模型，并将其应用于说话人识别。相比于 SVM，RVM 模型能获得更大的稀疏性和更快的测试速度，同时也提高了匹配模型的泛化能力和学习能力。实验结果证明，保局部核的性能明显优越于其他核函数，保持了样本局部结构，提高了模型分类精度。同时本文在说话人识别中采用快速边缘最大化算法，有效缩短了训练时间，加快了训练速度。因此，本文提出的方法具有实用性。

参考文献

[1] Cortes C, Vapnik V. Support Vector Networks[J]. Machine Learning, 1995, 20(3): 273-295.

[2] Tipping M E. Sparse Bayesian Learning and the Relevance Vector Machine[J]. The Journal of Machine Learning Research, 2001, 1(6): 211-244.

[3] 董国君, 哈力木拉提·买买提, 余 辉. 基于RBF核的SVM核参数优化算法[J]. 新疆大学学报, 2009, 26(3): 355-358, 363.

[4] Tipping M E, Faul A. Fast Marginal Likelihood Maximisation for Sparse Bayesian Models[C]//Proc. of the 9th International Workshop on Artificial Intelligence and Statistics. Key West, Florida, USA: [s. n.], 2003.

[5] 邢丽娟, 李 明. NAP 序列核函数在话者识别中的应用[J]. 计算机工程, 2010, 36(8): 194-196.

[6] Keerthi S S, Linc J. Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel[J]. Neural Computation, 2003, 15(7): 1667-1689.

[7] Lin Hsuan-Tien, Lin Chih-Jen. A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-type Methods[D]. Taipei, China: Department of Computer Science and Information Engineering, National Taiwan University, 2003.

[8] 阎 辉, 张学工, 马云潜. 基于变异函数的径向基核函数参数估计[J]. 自动化学报, 2002, 28(3): 450-455.

[9] 武优西, 郭 磊, 柴 欣, 等. 基于优化算法的核函数参数选择的研究[J]. 计算机应用与软件, 2010, 27(1): 137-140.

[10] He Xiaofei. Locality Preserving Projections[D]. Chicago, USA: University of Chicago, 2005.

[11] Zelnik-Manor L, Perona P. Self-tuning Spectral Clustering[C]// Proc. of NIPS'05. Cambridge, Massachusetts, USA: MIT Press, 2005: 1601-1608.

编辑 金胡考