DOI:10.13229/j.cnki.jdxbgxb2013.02.033

基于多核 SVM-GMM 的短语音说话人识别

林 琳,陈 虹,陈 建,金焕梅

(吉林大学 通信工程学院,长春 130022)

摘 要:运用多个核函数的线性组合构造多核空间,在多核空间上设计了基于支持向量机的说话人分类器,实现短语音说话人识别。多核映射能够解决单核映射核函数及其参数选择的难题,增加说话人的可区分性,提高分类器的性能。算法中结合了高斯混合模型(GMM),并以 GMM 超向量作为说话人的最终特征参数进行仿真实验。实验表明,在短语音和两种噪声环境中,基于多核 SVM-GMM 的短语音说话人识别算法较 SVM-GMM 算法能得到更好的识别性能和鲁棒性。

关键词:通信技术;说话人识别;短语音;多核支持向量机;高斯混合模型超向量

中图分类号:TN912.3 **文献标志码**:A **文章编号**:1671-5497(2013)02-0504-06

Speaker recognition with short utterances based on multiple kernel SVM-GMM

LIN lin, CHEN Hong, CHEN Jian, JIN Huan-mei

(*College of Communication Engineering, Jilin University, Changchun* 130022, *China*)

Abstract: A linear combination of several kernels is used to construct multiple kernel space. In multiple kernel space, Support Vector Machine (SVM) classifiers are designed to identify speakers with short utterances. Multiple kernel mapping can solve the problem of single kernel mapping, such as the selection of kernel function and parameters. Besides, multiple kernel mapping can increase discriminative power among different speakers and improve the performance of classifiers. In simulation experiment, Gaussian Mixture Model (GMM) was used to get GMM supervector as speakers' final feature parameters. Experiment results show that under the condition of short utterances and two noisy environments, the performance and robustness of the multiple SVM-GMM speaker recognition algorithm are better than that of SVM-GMM algorithm.

Key words: communication; speaker recognition; short utterances; multiple kernel SVM; Gaussian mixture model supervector

说话人识别是一种以说话人语音对说话人进行区分,从而进行身份鉴别与验证的技术。为了达到令人满意的效果,大多数说话人识别系统在建立话者模型时仍然需要较长的语音文本和大量的训练数据,尽管可以利用各种算法来减少系统的识别时间,达到实用化,但是对于那些只能获得少量说话人语音数据的应用场合,这些系统就无能为力了。因此,利用短语音文本以及尽可能少

收稿日期:2012-05-10.

基金项目:吉林省科技发展计划项目(201101032);高等学校博士学科点专项科研基金项目(20090061120042).

作者简介:林琳(1979a2),女,讲师,博士. 研究方向:语音信号处理,模式识别. E-mail:lin_lin@jlu.edu.cn

通信作者:陈建(1977a2),男,讲师,博士. 研究方向:数字信号处理,阵列信号处理. E-mail:chenjian@jlu.edu.cn

的训练数据建立有效的说话人模型实现高性能的说话人识别，更具有现实意义[1-2]。

近年来，支持向量机（Support vector machine，SVM）已经成为说话人识别领域的主流方法，尤其适合小样本数据条件下的数据分类问题[3]。由于核函数的引入，支持向量机能够有效地解决非线性的分类问题，但核函数及其参数的选择尚未获得理论上的支持。随着多核学习（Multiple kernel learning，MKL）[4-7]的提出，多个核函数映射的特征空间更适合于对分布复杂的说话人语音特征参数进行分类。因此，本文将多核学习引入支持向量机中，运用多个核函数的线性组合构造多核空间，将说话人的特征参数映射到高维的多核空间中，使说话人的特征参数得到更好的表达。同时，结合高斯混合模型（Gaussian mixture model，GMM），以 GMM 超向量作为说话人的最终特征参数，进行多核支持向量机分类器的设计，进一步提高短语音说话人识别系统的性能。

1　多核支持向量机

1.1　多核支持向量机的结构

多核学习方法利用多个核函数代替单一核函数，增强了决策函数的可解释性，能够获得比单核模型或单核机器组合模型更优的性能。

假设存在 M 个核函数 $K_m(\boldsymbol{x}_i,\boldsymbol{x}_j)$，$m=1,\cdots,M$；$i,j=1,\cdots,n$，定义多核函数

$$K(\boldsymbol{x},\boldsymbol{x}')=\sum_{m=1}^{M}d_mK_m(\boldsymbol{x},\boldsymbol{x}') \tag{1}$$

式中：$d_m\geqslant 0$，$\sum_{m=1}^{M}d_m=1$；对于每个基本核函数 K_m 而言，K_m 可分别选取不同的核函数和自由参数，例如可以选取一组不同宽度的高斯核函数。

利用式(1)中的加权线性合成核代替传统支持向量机里的单个核函数，就构成多核支持向量机。图 1 给出了多核支持向量机示意图。

1.2　多核支持向量机的优化

单核支持向量机的判决函数可以表示为

$$f(\boldsymbol{x})=\sum_{i=1}^{n}y_ia_i^*K(\boldsymbol{x}_i,\boldsymbol{x})+b^* \tag{2}$$

式中：α_i^* 和 b^* 通过求解下面的优化问题得到。

$$\min_{w,b,\xi}\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_i\xi_i \tag{3}$$

约束于

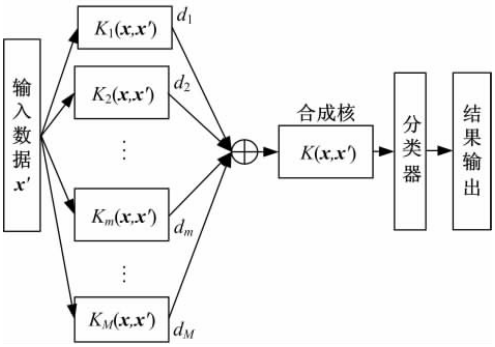

图 1　多核支持向量机结构图

Fig. 1　Multiple SVM construction

$$y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b)\geqslant 1-\xi_i,\xi_i\geqslant 0,\forall i \tag{4}$$

多核支持向量机的判决函数可以定义为

$$f(\boldsymbol{x})=\sum_{m=1}^{M}f_m(\boldsymbol{x})+b \tag{5}$$

式中：f_m 由核函数 K_m 相对应的再生核希尔伯特空间确定；M 为核函数的个数。

多核 SVM 原始问题可以定义为下面的凸优化问题

$$\min_{\{f_m\},b,\xi,d}\frac{1}{2}\sum_{m=1}^{M}\frac{1}{d_m}\|f_m\|^2+C\sum_i\xi_i \tag{6}$$

约束于

$$y_i\sum_{m=1}^{M}f_m(\boldsymbol{x}_i)+y_ib\geqslant 1-\xi_i,\ \xi_i\geqslant 0,\ \forall i \tag{7}$$

式中：$\sum_{m=1}^{M}d_m=1$，$d_m>0$，$\forall m$，权值 d_m 控制目标函数 f_m 的平方范数的比重，d_m 的 l_1 范数约束属于稀疏约束，限制了某些 d_m 等于 0，有利于稀疏基本核扩展。

本文通过迭代使用现有的支持向量机学习算法解决上述优化问题，分两步执行：

(1)假设 d 已知，仅考虑 f_m、b、ξ，优化式(6)。

(2)假设 f_m、b、ξ 固定，仅考虑 d，优化式(6)。

上述优化问题转化为如下形式：

$$\min_d J(d),\ d\geqslant 0,\sum_{m=1}^{M}d_m=1 \tag{8}$$

式中：

$$J(d)=\begin{cases}\min\limits_{f_m,b,\xi_i}\dfrac{1}{2}\sum\limits_{m=1}^{M}\dfrac{1}{d_m}\|f_m\|^2+C\sum\limits_i\xi_i\\ \text{subject to: } y_i\Big(\sum\limits_{m=1}^{M}f_m(\boldsymbol{x}_i)+b\Big)\geqslant 1-\xi_i\\ \xi_i\geqslant 0\end{cases} \tag{9}$$

很显然，对目标函数 $J(d)$ 的优化是典型的 SVM

优化问题,假设 $J(d)$ 可微,则采用梯度投影法求解式(8)的最小值优化问题。

设 d 已知,$J(d)$ 是典型 SVM 问题的目标值,$K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\sum_{m=1}^{M}d_mK_m(\boldsymbol{x}_i,\boldsymbol{x}_j)$,式(9)Lagrange 函数为

$$L=\frac{1}{2}\sum_{m=1}^{M}\frac{1}{d_m}\|f_m\|^2+C\sum_i\xi_i+\sum_i\alpha_i\left\{1-\xi_i-y_i\left[\sum_{m=1}^{M}f_m(x_i)+b\right]\right\}-\sum_i\nu_i\xi_i \tag{10}$$

对式(10)求偏导,令偏导数为 0,得到

$$①\ \frac{1}{d_m}f_m(\cdot)=\sum_i\alpha_iy_iK_m(\cdot,\boldsymbol{x}_i),\forall m$$

$$②\ \sum_i\alpha_iy_i=0$$

$$③C-\alpha_i-\nu_i=0,\forall i$$

转化为拉格朗日对偶问题

$$\max_{\alpha}\{-\frac{1}{2}\sum_i\sum_jy_iy_j\alpha_i\alpha_j\sum_{m=1}^{M}d_mK_m(\boldsymbol{x}_i,\boldsymbol{x}_j)+\sum_i\alpha_i\} \tag{11}$$

$$\text{with}\quad\sum_i\alpha_iy_i=0$$

$$C\geqslant\alpha_i\geqslant0,\forall i$$

其中,$m=1,\cdots,M$,解上述优化问题得到 α_i^*。

$$J(d)=-\frac{1}{2}\sum_i\sum_j\alpha_i^*\alpha_j^*\sum_{m=1}^{M}d_mK_m(\boldsymbol{x}_i,\boldsymbol{x}_j)+\sum_i\alpha_i^* \tag{12}$$

$$\frac{\partial J}{\partial d_m}=-\frac{1}{2}\sum_i\sum_jy_iy_j\alpha_i^*\alpha_j^*K_m(\boldsymbol{x}_i,\boldsymbol{x}_j),\forall m \tag{13}$$

利用梯度下降法对式(13)求解 d_m。

基于多核支持向量机优化算法的流程如下:

初始化 $d_m=\frac{1}{M}$,m=1,2,…,M;

While 未达到停止条件,do

使用 $K=\sum_{m=1}^{M}d_mK_m$ 的 SVM 计算 $J(d)$;

对 m=1,…,M 计算 $\frac{\partial J}{\partial d_m}$ 及下降方向 D;

置 $\mu=\underset{m}{\operatorname{argmax}}d_m$,$J^*=0$,$d^*=d$,$D^*=D$

While $J^*<J(d)$,do{更新下降方向}

$d^*=d$,$D^*=D$;

$\nu=\underset{\{m\mid D_m<0\}}{\operatorname{argmin}}-d_m/D_m$;$\gamma_{\max}=-d_\nu/D_\nu$

$d^*=d+\gamma_{\max}D$,$D_\mu{}^*=D_\mu-D_\nu$,$D_\nu^*=0$

使用 $K=\sum_{m=1}^{M}d_m^*K_m$ 的 SVM 计算 J^*;

end while

end while

2 多核 SVM-GMM 的短语音说话人识别系统

随着不同特征空间中数据的分布不同,支持向量机的性能很大程度上取决于核函数以及核参数的选择,然而至今这仍是一个难以解决的问题,同时单一核函数无法解决多个不同数据源的复杂问题。因此,本文将多核支持向量机应用到短语音说话人识别系统中,利用多核的组合空间对特征参数进行映射,并结合 GMM 提出了基于多核 SVM-GMM 的短语音说话人识别算法。其结构框图如图 2 所示。

从图 2 可以看出,基于多核 SVM-GMM 的说话人识别系统包含训练和识别两个阶段。

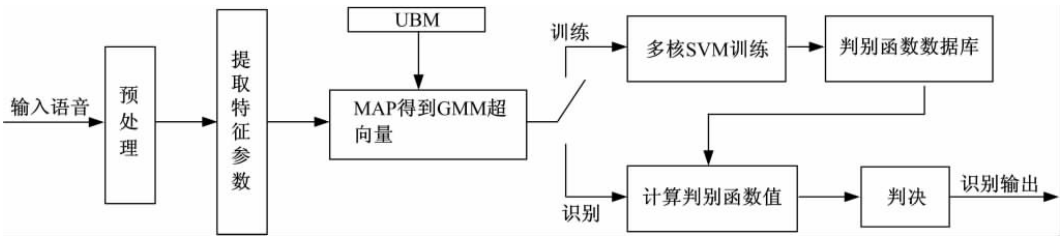

图 2 基于多核 SVM-GMM 的说话人识别系统

Fig. 2 Speaker recognition system based on multiple SVM-GMM

训练阶段:将所有说话人的语音数据依次输入,对每个说话人的语音数据进行预处理和特征提取,利用通用背景模型 UBM 进行 MAP 自适应,得到每个说话人的 GMM 超向量特征参数。以超向量特征参数作为输入样本进行多核 SVM 训练,采用一对一法实现多类分类。利用每两个说话人的特征参数训练一个多核 SVM 分类器,每个分类器对应一个判决函数,对于 k 个说话人,

需要训练 $k(k-1)/2$ 个 SVM 分类器，相应的判决函数为 $f_{i,j}(x)i=1,\cdots,k;j=(i+1),\cdots,k$。多核训练需要选定多个核函数，按照 1.2 节中介绍的多核 SVM 优化的算法流程计算各个核函数的权值，最终将得到的判决函数存储到数据库中。

识别阶段：输入待识别语音，对语音数据进行预处理和特征提取，然后对 UBM 模型进行 MAP 自适应，得到每个说话人的 GMM 超向量特征参数。设得到相应的 GMM 超向量为 $\boldsymbol{X}_t$，计算判决函数 $f_{i,j}(\boldsymbol{X}_t),i=1,\cdots,k;j=1,\cdots,k$ 其中，$f_{j,i}(\boldsymbol{X}_t)=-f_{i,j}(\boldsymbol{X}_t)$。

如果计算得到的判决函数值满足

$$f_{i,j}(\boldsymbol{X}_t)>0,\ j=1,\cdots,k \tag{14}$$

则待识别的说话人为第 i 个说话人。

3　计算结果与比较

实验 1　多核支持向量机与单核支持向量机的比较。

为了比较多核支持向量机和单核支持向量机的分类性能，采用 UCI 机器学习数据库[8] 中的 ionosphere 数据库进行分类实验，选取高斯核函数和多项式核函数，高斯核函数的参数 σ 分别取 0.5、1、2，多项式核函数的参数 p 分别取 1、2、3。实验结果如表 1 和表 2 所示。

表 1　单核 SVM 分类率

Table 1　Classification rate of single kernel SVM　%

高斯核函数			多项式核函数		
$\sigma=0.5$	$\sigma=1$	$\sigma=2$	$p=1$	$p=2$	$p=3$
16.48	31.84	33.52	17.61	18.18	17.61

表 2　多核 SVM 分类率

Table 2　Classification rate of multiple kernel SVM　%

高斯核函数				多项式核函数			
$\sigma=0.5$ $\sigma=1$	$\sigma=0.5$ $\sigma=2$	$\sigma=1$ $\sigma=2$	$\sigma=0.5$ $\sigma=1$ $\sigma=2$	$p=1$ $p=2$	$p=2$ $p=3$	$p=1$ $p=3$	$p=1$ $p=2$ $p=3$
15.47	10.23	10.23	10.23	11.36	13.86	14.20	10.80

表 1 为采用单个核函数进行分类的结果，从表 1 可以看出，高斯核函数在 时错误分类率最小为 16.48%；多项式核函数在 时错误分类率最小，为 17.61%。表 2 为多核支持向量机分类实验，分别对不同参数的核函数进行线性加权形成多核。从表 2 可以看出，对于高斯核函数而言，将 σ 分别为 0.5、1、2 的高斯核函数进行线性组合形成多核函数时，其错误分类率为 10.23%，而对于多项式核函数，将 p 分别为 1、2、3 的多项式核函数形成多核函数时，其错误分类率可以达到 10.80%。可见，多核支持向量机的错误分类率均低于单核支持向量机。如果将 σ 分别为 0.5、1、2 对应的高斯核函数和 p 分别为 1、2、3 对应的多项式核函数线性加权构成一个大的核函数，采用此核函数对上述数据进行分类实验，错误分类率可以达到 9.09%。由此可以得出，采用多核映射可以有效地降低错误分类率。

实验 2　基于多核 SVM 的说话人识别实验。

实验中采用 PKU-SRSC 语音数据库数据[9] 进行与文本无关的说话人辨认实验。从数据库中任选 50 个说话人（25 个男生和 25 女生），选择其中 15 个男生和 15 个女生的语音用来构建 UBM 模型，UBM 模型为 32 阶。另外 20 个说话人（包括男生 10 人，女生 10 人）进行说话人辨认实验。使用每个说话人的 10 次录音数据，其中说话人的每次录音间隔为一周，每个说话人的每次录音数据包括 20 个语音文件。选择其中第一次录音的部分语音作为训练语音，其余的所有语音作为识别语音，每个识别语音约为 1 s 左右。

在进行训练和识别时，对每个说话人的语音信号进行预处理，包括端点检测、预加重和分帧处理。采用汉明窗分帧，窗宽 256 个采样点，窗移 128 个采样点。语音信号经预处理后提取 MFCC 特征参数，包含 20 维 MFCC 系数和 20 维一阶动态 MFCC 系数进行组合，去掉第一维 MFCC 系数，将剩下的 39 维作为说话人的第一特征参数矩阵。利用上述提取的第一特征参数矩阵，采用 MAP 自适应算法，对 UBM 模型进行自适应，得到 GMM 超向量作为说话人最终的特征参数。MAP 自适应参数 $rp=10$，每 1 s 语音对 UBM 自适应得到一个超向量。选取不同的核及其线性组合进行实验，惩罚系数 $C=10$。由于 KL 核函数是针对 GMM 超向量提出的一种核函数，综合考虑 GMM 模型训练的协方差矩阵和权重的影响，实验中选取 KL 核函数及常用的高斯核函数来考察核函数对系统性能的影响。实验结果如表 3 所示。

从表 3 可以看出，KL 核函数对 GMM 超向量的分类性能较高斯核函数好，采用 KL 核函数和高斯核函数加权核进行分类，系统的误识率较低，随着核个数的增加，误识率下降，当核函数为 KL 核函数、$\sigma=0.1$ 的高斯核函数和 $\sigma=0.05$ 的高斯核函数的线性加权核时，系统误识率达到 1.5%。

表3 核函数对系统性能的影响

Table 3 Impact of kernel function to system performance

核函数	误识率/%
KL核	2.50
高斯核(σ=0.1)	19.5
高斯核(σ=0.05)	20.0
高斯核(σ=0.1)+高斯核(σ=0.05)	19.5
KL核+高斯核(σ=0.1)	2.80
KL核+高斯核(σ=0.05)	2.90
KL核+高斯核(σ=0.1)+高斯核(σ=0.05)	1.50

实验3 多核SVM-GMM模型和SVM-GMM模型的比较。

为了比较多核SVM-GMM模型和SVM-GMM模型[10]的系统性能,实验选取2～10 s的训练语音进行说话人辨认实验。SVM-GMM模型选取KL核函数,多核SVM-GMM的核函数为KL核函数、$\sigma=0.1$的高斯核函数和$\sigma=0.05$的高斯核函数的线性加权核,惩罚系数$C=10$。实验结果如图3所示。

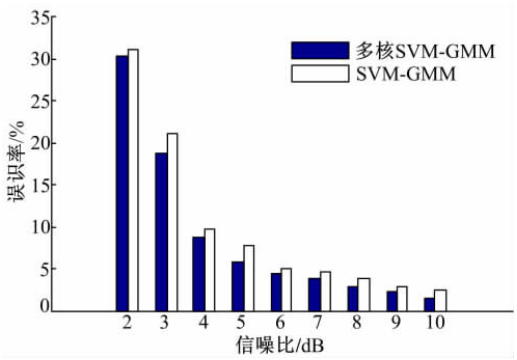

图3 不同训练语音下两个分类器的系统误识率

Fig. 3 System error rate of the two classifications with different training data

从图3可以看出,随着训练数据的增加,SVM-GMM分类器和多核SVM-GMM分类器的误识率均呈减小趋势,多核SVM-GMM分类器的误识率优于SVM-GMM分类器。这是由于多个核函数构成的多核空间能够增加输入特征参数的可区分性,提高SVM的分类正确率,从而降低了短语音说话人识别系统的误识率。

实验4 噪声环境下不同模型鲁棒性比较。

为了比较多核SVM-GMM分类器和SVM-GMM分类器的噪声鲁棒性,采用NOISEX-92噪声库中的高斯白噪声和Pink噪声进行实验。训练语音为10 s的纯净语音,选取不同信噪比的语音进行测试。SVM-GMM模型选取KL核函数,多核SVM-GMM的核函数为KL核函数、$\sigma=0.1$的高斯核函数和$\sigma=0.05$的高斯核函数的线性加权核,惩罚系数$C=10$。图4和图5分别为高斯白噪声和Pink噪声环境下两个分类器的系统误识率。

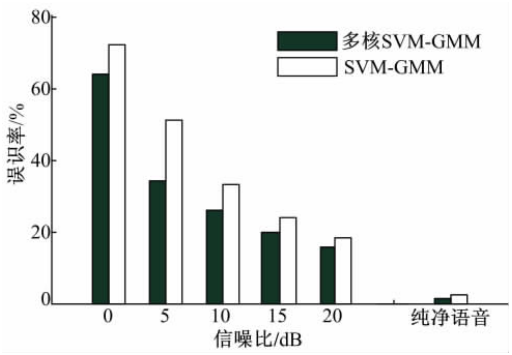

图4 高斯白噪声下不同分类器的系统误识率

Fig. 4 System error rate of the two classifications under Gaussian white noise condition

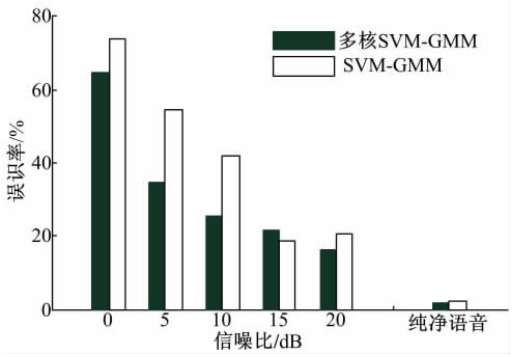

图5 Pink噪声下不同分类器的系统误识率

Fig. 5 System error rate of the two classifications under Pink noise condition

由图4和图5可以看出,在白噪声和Pink噪声环境中进行实验,不同信噪比及纯净语音条件下,基于多核SVM-GMM的系统都能得到较SVM-GMM系统更低的误识率,可见,基于多核映射的SVM-GMM分类器具有更好的鲁棒性。由仿真结果可以看出,在两种噪声条件下,随着信噪比的降低,两个说话人识别系统的误识率都会逐渐增加,尤其当信噪比为0 dB时,两个系统的误识率均达到60%以上,可见,尽管多核SVM-GMM系统对噪声具有一定的鲁棒性,但在低信噪比条件下,其性能受噪声影响仍然很严重,因此,克服噪声对系统性能的影响是很有必要的。

4　结束语

提出了一种基于多核 SVM-GMM 的短语音说话人识别算法，运用多个核函数的线性组合构造多核空间，设计了基于多核支持向量机的说话人分类器。算法结合 GMM 模型，以 GMM 超向量作为说话人的特征参数输入多核 SVM 进行训练，选取 KL 核函数和常用核函数的线性加权核作为 SVM 的多核映射空间。通过对 ionosphere 数据和说话人识别的仿真实验，验证了多核 SVM 的有效性。在短语音说话人识别系统的仿真实验中，提出的基于多核 SVM-GMM 的短语音说话人识别算法能够在较短训练语音和两种噪声环境中得到较 SVM-GMM 算法更好的识别性能和鲁棒性。

参考文献：

[1] Yang Yao-yuan, Chen Wei, Lu Yu-dong, et al. Research of speaker identification based on little training data[C]// Proceeding of the 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 2004.

[2] Jayanna H S, Mahadeva Prasanna S R. Multiple frame size and rate analysis for speaker recognition under limited data condition[J]. IET Signal Processing, 2009, 3(3): 189-204.

[3] Mak Man-Wai, Rao Wei. Utterance partitioning with acoustic vector resampling for GMM-SVM speaker verification[J]. Speech Communication, 2011, 53(1): 119-130.

[4] Zien A, Ong C S. Multiclass multiple kernel learning[C]// Proceedings of the 24th International Conference on Machine Learning. New York, USA: ACM, 2007.

[5] Tian Xi-lan, Gasso Gilles, Canu Stéphane. A multiple kernel framework for inductive semi-supervised SVM learning[J]. Neuro-computing, 2012, 90(1): 46-58.

[6] Chen Zhen-yu, Li Jian-ping, Wei Li-wei, et al. Multiple-kernel SVM based multiple-task oriented data mining system for gene expression data analysis [J]. Expert Systems with Applications, 2011, 38 (10): 12151-12159.

[7] Wu Zheng-peng, Zhang Xue-gong. Elastic multiple kernel learning[J]. Acta Automatica Sinica, 2011, 37(6): 693-699.

[8] University of California Irvine. UCI Machine Learning Repository[EB/OL]. http://archive.ics.uci.edu/ml

[9] 吴玺宏. 一个面向说话人识别的汉语语音数据库[EB/OL]. http://nlpr-web.ia.ac.cn/english/irds/chi-nese/sinobiometricspdf/wuxihong.pdf

[10] Campbell W, Sturim D E, Reynolds D A. Support vector machines using GMM supervectors for speaker verification[J]. IEEE Signal Processing Letters, 2006, 13(5): 308-311.