基于 GMM-UBM 和 GLDS-SVM的 英文发音错误检测方法

李宏言 [1] 黄 申 [1] 王士进 [1] 梁家恩 [1] 徐 波 [1,2]

摘 要 将语种和说话人识别的方法应用到英语发音错误检测系统, 提出一种基于广义线性区分序列支持向量机 (Generalized linear discriminant sequence based SVM, GLDS-SVM) 的发音错误检测方法. 主要创新点为: 1) 提出一种基于状态拼接的特征规整方案, 增强 SVM 对发音特征的建模能力; 2) 提出一种基于多模型融合的模型训练策略, 该策略可以更加充分地利用训练数据, 并在一定程度上解决了由于真实发音错误数据缺乏造成的正负样本不均衡的问题; 3) 将 GLDS-SVM 与基于通用背景模型 GMM (Universal background models based GMM, GMM-UBM) 的方法进行融合, 以进一步提高发音检错性能. GLDS-SVM 和 GMM-UBM 的融合系统在仿真测试集和真实测试集上的等错误率 (Equal error rate, EER) 分别达到 9.92 % 和 16.35 %. 同时, GLDS-SVM 在模型占用空间和运算速度方面均比传统径向基函数 (Radial basic function, RBF) 核方法具有明显优势.

关键词 计算机辅助语言学习, 自动发音错误检测, 支持向量机特征规整, 多模型融合策略

DOI 10.3724/SP.J.1004.2010.00332

Automatic Mispronunciation Detection for English Learners by GMM-UBM and GLDS-SVM Methods

LI Hong-Yan[1] HUANG Shen[1] WANG Shi-Jin[1]
LIANG Jia-En[1] XU Bo[1,2]

Abstract The paper proposes an efficient generalized linear discriminant sequence based SVM (GLDS-SVM) based mispronunciation detection method. Firstly, in order to enhance the ability of describing pronunciation characteristics, we introduce an improved SVM feature normalization scheme based on state-concatenated operation. Then, we propose a novel multi-model strategy for model training to make full use of samples and solve the problem of data unbalance caused by lack of the actual mispronunciation corpus. Finally, we combine GLDS-SVM with universal background models based GMM (GMM-UBM) to further improve the performance. The fused system by these two methods achieves 9.92 % and 16.35 % in equal error rate (EER) for simulation set and real set, respectively. Meanwhile, GLDS-SVM processes a higher computation speed and smaller model size than traditional radial basic function (RBF) kernel.

Key words Computer assisted language learning (CALL), automatic mispronunciation detection, support vector machine (SVM) feature normalization, multi-model fusion strategy

收稿日期 2009-03-19 录用日期 2009-10-21
Manuscript received March 19, 2009; accepted October 21, 2009
国家高技术研究发展计划 (863 计划) (2006AA010103) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA010103)

1. 中国科学院自动化研究所数字内容技术研究中心 北京 100190 2. 中国科学院自动化研究所模式识别国家重点实验室 北京 100190
1. Digital Content Technology Research Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 2. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

自动发音错误检测是计算机辅助语言学习 (Computer assisted language learning, CALL) 领域的重要内容, 其目标是给出发音的错误情况并提供反馈和改进意见. 在实际应用中, 发音错误检测问题的关键点可以体现在两个方面, 即发音质量特征的提取和检测方法的选用. 具体而言, 发音质量特征要求可以有效鲁棒地描述语音的发音质量, 而检测方法则要求对待测语音的发音正误进行准确地区分和判决.

文献 [1] 利用 formant 特征并使用高斯混合模型 (Gaussian mixture model, GMM) 分类器对中文韵母质量进行评估. 文献 [2] 通过线性采样 (Linear-sampled) 的方法将不同长度的韵母共振峰轨迹归整为 12 帧, 并使用基于径向基函数 (Radial basic function, RBF) 核的支持向量机 (Support vector machine, SVM) 分类器进行质量评价. 通常, SVM 只能处理固定长度的输入向量, 因此在发音检错中, 必须将长度不等的语音段转换为维数统一的特征. 显然, 基于采样的特征归整方式会造成特征信息的部分丢失, 因此本文提出两种针对 SVM 输入特征归整的改进方案. 此外, SVM 也可以用于得分域的分类预测, 如文献 [3] 将后验得分向量作为 SVM 的输入特征, 对易混淆的中文音素进行发音检错, 取得了较好效果. 另外, SVM 核函数的选择至关重要, 为此, 本文将在说话人和语种识别领域取得成功应用的广义线性区分序列 (Generalized linear discriminant sequence, GLDS) 核引入到发音检错, 并提出一种基于多模型融合的模型训练策略.

1 Baseline 系统

1.1 系统概述

本文的发音错误检测系统的结构框图如图 1 所示. 输入语音首先进行预处理和分帧, 然后提取语音帧的特征参数, 而强制对齐部分的作用是提供输入语音中的各个音素的边界信息, 此处在已知语音脚本的前提下利用隐马尔科夫模型 (Hidden Markov models, HMM) 的维特比 (Viterbi) 算法解码出对应的音素序列, 并通过在词典中扩展发音变异的方式有效提高切分准确性. 在获取各音素边界信息的基础上, 将特征转换为各分类器要求的形式, 并通过运算得到一系列置信度得分. 由于各检测方法具有不同的结构, 因此具有互补性, 此时将结果进行融合可以进一步提高性能. 最后, 通过与阈值的比较来判决音素的发音正误, 本文对不同音素采用不同的判决阈值.

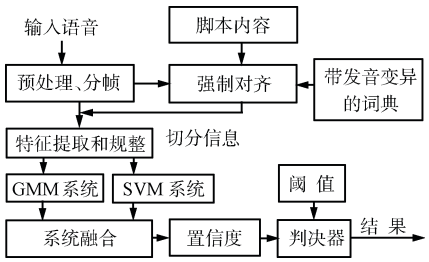


图 1 发音错误检测系统的结构框图

Fig. 1 Architecture of the whole detection system

1.2 GMM-UBM 发音错误检测方法

GMM 是特征分类中最基本的分类方法之一, 用于对特征进行概率分布建模. 本文将音素的发音质量得分描述为该音素各帧特征对应的 GMM 对数后验概率的平均值, 即

$$score(phn_i) = \frac{1}{e_i - s_i + 1} \sum_{t=s_i}^{e_i} \log p(phn_i|\boldsymbol{x}_t) \quad (1)$$

$$p(phn_i|\boldsymbol{x}_t) = \frac{p(\boldsymbol{x}_t|gmm_i)P(phn_i)}{\sum\limits_{k\in V} p(\boldsymbol{x}_t|gmm_k)P(phn_k)} \quad (2)$$

其中, $\boldsymbol{x}_t$ 为特征帧向量, s_i 和 e_i 分别为音素的起始帧和终止帧, gmm_i 为音素 phn_i 对应的 GMM 模型, V 代表元音集或辅音集, $p(\boldsymbol{x}_t|gmm_k)$ 为特征帧 $\boldsymbol{x}_t$ 对于模型 gmm_i 的输出概率, $P(phn_k)$ 为先验概率, 此处认为所有音素的先验概率相同.

每个音素训练一个 GMM 模型, 为了提高模型训练速度和模型精度, 首先分别训练出元音集和辅音集对应的通用背景模型 (Universal background models, UBM), 然后在 UBM 的基础上利用贝叶斯自适应算法训练得到各音素对应的 GMM 模型. 使用 UBM 是语种和说话人识别中常用的做法, 实验表明比单独训练效果更好[4], 因此本文将 GMM-UBM 作为发音错误检测的基线系统.

2　GLDS-SVM 发音错误检测方法

2.1　SVM 输入特征的规整

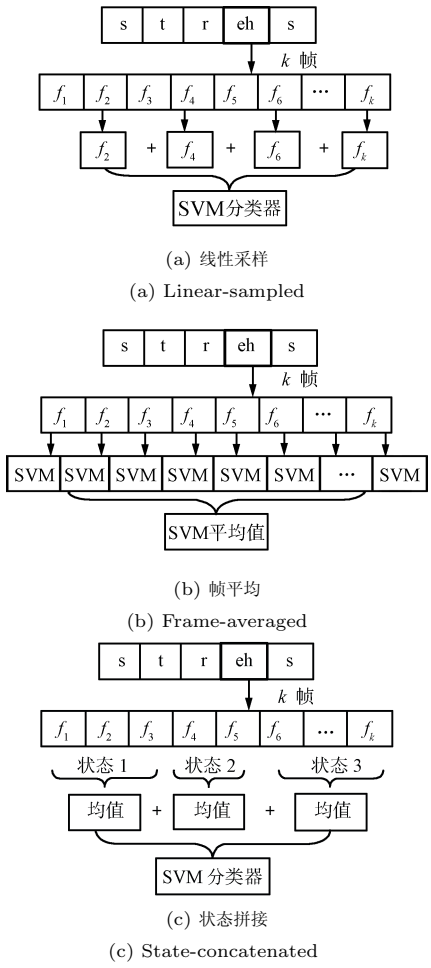
与 HMM 可以处理时序变量不同, SVM 只能处理固定长度的输入向量, 因此利用 SVM 进行发音诊断必须将长度 (帧数) 不等的语音段转换为维数统一的特征. 图 2 (a) 给出了文献 [2] 中使用线性采样的特征规整方式的原理图 (以 “stress” 词为例), 此做法实际上只利用了部分语音帧信息.

为了充分利用语音段的特征信息, 本文针对 SVM 对输入特征的要求及结合语音信号的特点, 引入了两种特征处理方案: 一种是基于帧平均的方法 (Frame-averaged), 即直接将每一帧的特征作为 SVM 的输入向量, 然后对各帧的 SVM 输出结果进行平均, 如图 2 (b) 所示; 另一种方法是基于状态拼接的方案 (State-concatenated), 即利用 HMM 进一步确定语音段的状态序列, 将各个状态对应的平均特征向量进行拼接, 形成一个维数固定的复合特征向量, 如图 2 (c) 所示.

语音的时序性决定了在对语音进行声学建模的时候必须考虑频域特征的时间变化, 线性采样方案虽然在一定程度上实现了特征的时序描述, 但对音素段的各特征帧的变化规律没有进行深入挖掘, 采样得到的特征帧不能充分描述整个音素段的特性 (采样帧数太少则信息丢失严重, 而采样帧数太多则显著增加了特征维数), 而帧平均的方案由于直接对各个孤立的语音帧进行操作, 没有体现出对特征时序变化的描述.

文献 [5] 利用 SVM 替代 HMM 中的 GMM 进行概率输出以提高语音识别的性能, 并指出可以将各音素段看作由 3 个区域 (Section) 构成, 而这 3 个区域分别对应 HMM 中的 3 个状态 (Three-state), 前后 2 个状态看作为音素段的开始变化和结束变化的区域, 中间的状态看作为音素段稳定变化的区域, 在假定各状态内部相对稳定的情况下, 可以使用各状态对应的平均特征帧向量来作为对各状态的描述, 而将各状态的平均特征进行拼接就可以有效刻画特征的时序变化. 此做法充分利用了 HMM 的结构特点, 并很自然地为 SVM 构造了维数统一的输入向量.

显然, 基于帧平均的方案简单易行, 但其缺点是建立在各帧特征相互独立的基础上, 因此降低了对语音时序变化的描述能力. 与之比较, 基于状态拼接的方案实质上是对语音的音素段进行整体建模 (Segmental modeling), 加强了时序描述的力度, 但在各状态的平均特征进行拼接的过程中, 需要事先通过 HMM 解码出对应的状态序列, 因此运算量有所增加.



(a) 线性采样

(a) Linear-sampled

(b) 帧平均

(b) Frame-averaged

(c) 状态拼接

(c) State-concatenated

图 2　SVM 输入特征规整方案

Fig. 2　SVM feature normalization scheme

2.2　SVM 后验概率形式的估计

对于模式识别问题, 一般通过后验概率来衡量分类的不确定性. 然而, SVM 只能提供分类面的间隔距离来间接描述类别间的区分程度. 因此, 将 SVM 应用于发音错误检测的关键问题之一就是需要给出分类面间隔距离和后验概率之间的映射关系 (此处是指映射为后验概率的形式, 而一般认为 SVM 的输出没有包含概率方面的信息). 本文综合计算复杂度和估计错误率等因素, 选取 Sigmoid 来描述 SVM 距离输出和后验概率的映射关系[6], 即

$$p(d) = \frac{1}{1 + \exp(\gamma \times d)} \quad (3)$$

其中, γ 为 Sigmoid 曲线的陡峭度参数, 此处取 -0.5, 如图 3 所示.

2.3　基于 GLDS 核的发音错误检测

近来, 基于广义线性区分序列核的 SVM 方法在说话人识别和语种识别领域得到广泛应用[7], GLDS 核在运算速度和模型占用空间方面具有明显优势. 本文提出一种基于 GLDS 核的发音错误检测方法, 图 4 简要给出了 GLDS-SVM 的算法流程.

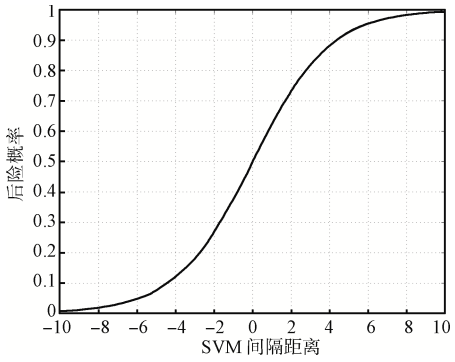


图 3 SVM 后验概率的 Sigmoid 估计

Fig. 3 Sigmoid estimation for SVM posterior probability

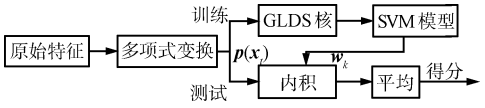


图 4 GLDS-SVM 发音错误检测的流程框图

Fig. 4 Block diagram of GLD-SVM detection system

首先, 原始特征向量通过多项式变换为高维特征 $\boldsymbol{p}(\boldsymbol{x}_t)$, 对于 d 维特征, 当多项式取 q 阶时, 转换后的特征维数为 C_{d+q}^{q}. 在测试过程中, 音素 phn_i 的帧特征 $\boldsymbol{x}_t$ 对应的得分通过 $\boldsymbol{p}(\boldsymbol{x}_t)$ 和模型 $\boldsymbol{W}_i$ 的内积获得, 然后以各帧得分的平均值作为音素 phn_i 的置信度得分, 即

$$score(phn_i) = \frac{1}{e_i - s_i + 1} \sum_{t=s_i}^{e_i} \boldsymbol{W}_i^{\mathrm{T}} \boldsymbol{p}(\boldsymbol{x}_t) \tag{4}$$

在训练过程中, 首先进行广义线性分类器训练, 训练算法可以近似描述如下

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \left[\sum_{k=1}^{N_{\mathrm{pos}}} |\boldsymbol{w}^{\mathrm{T}} \boldsymbol{p}(\boldsymbol{x}_k) - 1|^2 + \sum_{k=1}^{N_{\mathrm{neg}}} |\boldsymbol{w}^{\mathrm{T}} \boldsymbol{p}(\boldsymbol{y}_k) - 0|^2 \right] \tag{5}$$

其中, $\boldsymbol{w}^*$ 为音素对应的最优模型, $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{N_{\mathrm{pos}}}$ 代表训练用的正样本, $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{N_{\mathrm{neg}}}$ 代表负样本, 正负样本的输出值分别为 "1" 和 "0". 接下去, 在最小均方错误准则 (Mean-squared error, MSE) 下构造 GLDS 核函数, 即

$$k_{\mathrm{GLDS}}(\boldsymbol{x}_t, \boldsymbol{y}_t) = \boldsymbol{p}(\boldsymbol{x}_t)^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{p}(\boldsymbol{y}_t) \tag{6}$$

其中, $\boldsymbol{R}$ 由式 (5) 的 $\boldsymbol{w}^*$ 分解得到. 此时, GLDS 核的输出可以表示为

$$g(\boldsymbol{p}) = \sum_{i=1}^{l} \alpha_i t_i k_{\mathrm{GLDS}}(\boldsymbol{p}_i, \boldsymbol{p}) + d \tag{7}$$

其中, t_i 为支持向量对应的输出值 (1 或 −1), 且满足 $\sum_{i=1}^{l} \alpha_i t_i = 0$, $\alpha_i > 0$. 将式 (6) 代入式 (7), 有

$$g(\boldsymbol{p}) = (\sum_{i=1}^{l} \alpha_i t_i \boldsymbol{R}^{-1} \boldsymbol{p}_i + \boldsymbol{d})^{\mathrm{T}} \boldsymbol{p} \tag{8}$$

此时, $\boldsymbol{d} = [d, 0, \cdots, 0]^{\mathrm{T}}$. 由此, 最终的 GLDS 模型可以写为

$$\boldsymbol{W} = \sum_{i=1}^{l} \alpha_i t_i \boldsymbol{R}^{-1} \boldsymbol{p}_i + \boldsymbol{d} \tag{9}$$

可见, GLDS 核可以将所有支持向量缩减为一个模型向量, 使得模型空间极大降低, 同时也使得识别测试过程的运算变得非常简单.

2.4 GLDS-SVM 的多模型融合训练策略

SVM 模型训练需要正负样本, 但对于发音错误检测来说, 在实际中很难获得大量错误发音数据来充当负样本, 因此训练样本不均衡是 SVM 应用于发音检错的主要困难之一.

为此, 本文使用当前音素的样本作为当前音素的正样本, 使用除了当前音素以外的元音或辅音音素的样本作为当前音素的负样本, 此处的音素样本均通过强制对齐获得. 上述操作使得音素对应的负样本数远远大于正样本数, 显然不利于 SVM 训练 (本文的 RBF-SVM 模型使用此方法训练获得).

进一步地, 本文针对 GLDS-SVM, 将各音素对应的负样本集随机划分为多个子集, 利用多个负样本子集训练获得多个模型, 即每个音素可以对应多个模型. 由于 GLDS-SVM 的输出得分是通过内积获得的, 因此测试样本与对应音素的多个模型的得分的平均值可以看作为测试样本与多个模型的平均模型的得分, 即该平均模型可以表示为

$$\boldsymbol{W}_k = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{W}_k^i \tag{10}$$

其中, $\boldsymbol{W}_k^i$ 为音素 phn_k 对应的第 i 个模型, N 为多模型个数. 多模型融合策略的流程如图 5 所示.

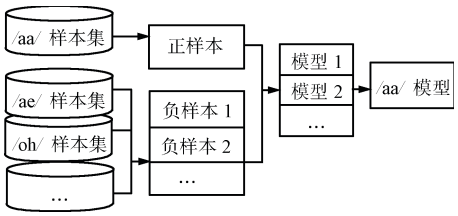


图 5 多模型融合训练策略

Fig. 5 Multi-model based training strategy

可见, 相比 RBF 核的模型训练, GLDS 核通过多模型融合的策略可以将训练集随机分成若干个子集, 而各子集训练的模型能够通过简单的平均融合方式形成最终的模型 (GLDS-SVM 的优势集中体现在只有 GLDS 核才可以做模型级的融合, 而这一点是 RBF 核无法达到的). 因此, 多模型融合的训练策略可以更加充分地利用训练数据, 并且加快模型训练速度 (对于 GLDS 训练, 多个并行小数据集的速度明显快于单个大数据集的速度).

另外, 由于发音错误不一定就是发错成其他音, 也有可能是发音模糊或发音缺陷, 因此上述在 SVM 模型训练过程中, 把除了当前音素以外的其他音素作为对应负样本的做法也是折衷之举 (受限于真实错误样本不足).

2.5 系统融合

系统融合提高性能的关键在于多套系统之间具有较好的互补性[8], 即各系统性能相当, 同时其结果具有一定的差异性. 由于本文的 GMM 和 SVM 方法在系统建模上具有明显的互补性和差异性 (GMM 为产生式模型, 而 SVM 是典型的区分性模型, 同时基于状态拼接特征规整的 SVM 增强了对语音时序变化的描述能力), 因此两者的融合是很自然的.

由于非线性融合方法一般需要具有标注信息的开发集数据, 鉴于开发集数据有限, 本文直接使用如下的线性融合方

式

$$score = w_{\mathrm{GMM}} \times score_{\mathrm{GMM}} + w_{\mathrm{SVM}} \times score_{\mathrm{SVM}} \quad (11)$$

其中, $score_{\mathrm{GMM}}$ 和 $score_{\mathrm{SVM}}$ 分别为 GMM-UBM 和 GLDS-SVM 的后验得分 (转换到统一的得分域), 而 w_{GMM} 和 w_{SVM} 为对应的权重, 此处均取值 0.5.

3　实验分析

3.1　数据准备

为了检验算法的性能, 在收集的中国学生英语语音库上进行实验, 详情见表 1. 训练集和仿真测试集中的每个学生分别朗读 100 个词汇和 100 个短句, 而真实测试集的每个学生朗读 290 个容易错误发音的词汇并经过音素级错误标注. 对于仿真测试集, 由于真实发音错误数据很难获得, 并且音素级的错误标注非常复杂繁重, 因此本文通过变换脚本的方式进行模拟仿真, 以得到语音和脚本不匹配的发音错误 (发音错误的实质为语音和脚本的不匹配). 脚本变换是通过把标准发音对应的音素随机替换为容易混淆的其他音素来实现的.

表 1　实验用的数据集组成

Table 1　Speech corpus in the experiment

数据集	男生数	女生数	总数
训练集 (Train-set)	100	200	300
仿真测试集 (Simu-set)	28	50	78
真实测试集 (Real-set)	10	10	20
总　体	138	260	398

实验语音的采样频率为 16 KHz, 精度为 16 bit, 帧长和帧移分别为 25 ms 和 10 ms. 特征使用 39 维美尔频率倒谱系数 (Mel-frequency cepstrum coefficient, MFCC) 参数. 用于强制对齐的 HMM 模型使用 120 小时数据训练. GMM 模型的高斯混合数为 16. 对于 GLDS-SVM, 其多项式变换取 3 阶, 模型训练的正负样本比例控制在 1:1 左右. SVM 模型训练使用 Libsvm[9]. 错误接受率 (False acceptance rate, FAR) 和错误拒绝率 (False rejection rate, FRR) 被用来衡量系统的错误检测性能. 同时, 检测错误折衷曲线 (Detect error tradeoff, DET) 和等错误率 (Equal error rate, EER) 也一并提供.

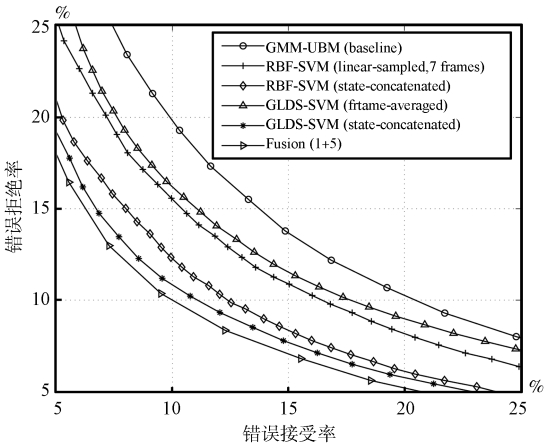
3.2　发音错误检测性能比较

图 6 分别给出了各种方法在 Simu-set 和 Real-set 测试集上的 DET 曲线, 而表 2 列出了相应的 EER. 可见, 相比 GMM 方法, SVM 方法在很大程度上提高了检测性能. 对于 RBF-SVM, 基于状态拼接的特征处理方案明显优于线性采样方案, 而对于 GLDS-SVM, 基于状态拼接的方案也明显优于帧平均方案. 同时, GMM 和 SVM 的融合可以进一步降低 EER.

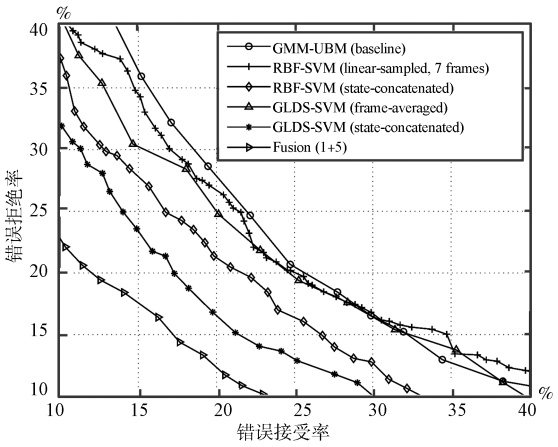
表 2　各检测方法的 EER 指标比较 (%)

Table 2　EER comparison for different methods (%)

方　法	Simu-set	Real-set
GMM-UBM (baseline)	14.32	23.40
RBF-SVM (linear-sampled, 7frames)	12.65	22.23
RBF-SVM (state-concatenated)	11.09	20.68
GLDS-SVM (frame-averaged)	13.05	22.31
GLDS-SVM (state-concatenated)	10.50	18.18
Fusion (1 + 5)	9.92	16.35



(a) Simu-set



(b) Real-set

图 6　各检测方法的 DET 比较

Fig. 6　DET for different methods

虽然区分性方法和基于状态拼接的特征处理方案比较有效, 但是在 Real-set 集上的检错性能仍然较低, 其根本原因在于缺乏大量用于分类器训练的带有细标的真实错误发音数据. 而本文提出的模型训练策略不失是一种解决正负样本不平衡问题的尝试, 其性能会在标注数据的不断积累下得到大幅度改善.

3.3　各检测方法的计算效率比较

各方法的计算效率见表 3. 由表 3 可见, 相对于 RBF 核, GLDS 核在速度提高的同时大幅度缩减了模型空间, 且各音素的模型大小统一, 因此 GLDS-SVM 更为实用. 而对于 GMM 方法, 由于本文采用的 GMM 混合项较少, 使得 UBM 的速度优势没有明显体现出来.

表 3　不同方法的计算效率比较

Table 3　Computation efficiencies of different methods

方　法	实时比 (X)	模型大小 (MB)
GMM-noUBM	0.018	1.44
GMM-UBM	0.012	1.44
RBF-SVM (state-concatenated)	0.022	138
GLDS-SVM (state-concatenated)	0.010	4.14

3.4 多模型训练策略中负样本集数的影响

对于 GLDS-SVM, 本文提出的多模型融合的模型训练策略受到负样本集数的影响. 由图 7 可见, EER 指标随着负样本集数的增加将逐渐下降, 当负样本集数为 10 的时候, EER 趋向稳定, 因此负样本集数设置为 10.

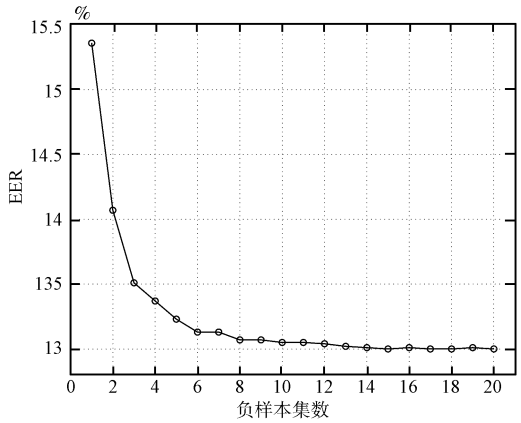


图 7 多模型训练策略中负样本集数对 EER 的影响

Fig. 7 EER curve with different numbers of negative sets

4 结论和展望

基于状态拼接的 SVM 输入特征规整方案优于帧平均和线性采样的策略. 对于 GLDS-SVM, 提出的多模型融合的模型训练策略充分利用了训练数据, 并在一定程度上解决了 SVM 模型训练过程中, 由于缺乏真实发音错误数据造成的正负样本不均衡的问题. GLDS-SVM 在速度和模型空间方面都相对 RBF-SVM 具有明显优势, 便于实际应用. 同时, 具有不同原理和结构的 GMM 和 SVM 方法通过融合可以进一步提高系统性能. 由于带细节标注的大规模真实数据的积累是一个长期过程, 相关内容会在后续工作中涉及.

References

1 Pan F P, Zhao Q W, Yan Y H. Mandarin vowel pronunciation quality evaluation by a novel formant classification method and its combination with traditional algorithms. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. Las Vegas, USA: IEEE, 2008. 5061−5064

2 Dong Bin, Zhao Qing-Wei, Yan Yong-Hong. Objective evaluation of vowels of standard Chinese pronunciation based on formant pattern. *Acta Acustica*, 2007, **32**(2): 122−128
(董滨, 赵庆卫, 颜永红. 基于共振峰模式的汉语普通话中韵母发音水平客观测试方法的研究. 声学学报, 2007, **32**(2): 122−128)

3 Jiang J, Xu B. Exploring the automatic mispronunciation detection of confusable phones for Mandarin. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. Taipei, China: IEEE, 2009. 4833−4836

4 Reynolds D A, Quatieri T F, Dunn R B. Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 2000, **10**(1-3): 19−41

5 Ganapathiraju A, Hamaker J E, Picone J. Applications of support vector machines to speech recognition. *IEEE Transactions on Signal Processing*, 2004, **52**(8): 2348−2355

6 Lin H T, Lin C J, Weng R C. A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, 2007, **68**(3): 267−276

7 Campbell W M, Campbell J P, Reynolds D A, Singer E, Torres-Carrasquillo P A. Support vector machines for speaker and language recognition. *Computer Speech and Language*, 2006, **20**(2-3): 210−229

8 Kittler J, Hatef M, Robert P W, Jiri M. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(3): 226−239

9 Chang C C, Lin C J. LIBSVM: a library for support vector machines [Online], available: http://www.csie.ntu.edu.tw/-cjlin/libsvm, September 11, 2007

李宏言 中国科学院自动化研究所博士研究生. 主要研究方向为计算机辅助语言学习. 本文通信作者. E-mail: hyli@hitic.ia.ac.cn
(**LI Hong-Yan** Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. His main research interest is computer assisted language learning. Corresponding author of this paper.)

黄 申 中国科学院自动化研究所博士研究生. 主要研究方向为计算机辅助语言学习. E-mail: shenhuang@hitic.ia.ac.cn
(**HUANG Shen** Ph. D. candidate at the Institute of Automation, Chinese Academy of Sciences. His main research interest is computer assisted language learning.)

王士进 中国科学院自动化研究所助理研究员. 主要研究方向为语音识别. E-mail: sjwang@hitic.ia.ac.cn
(**WANG Shi-Jin** Assistant professor at the Institute of Automation, Chinese Academy of Sciences. His main research interest is speech recognition.)

梁家恩 中国科学院自动化研究所助理研究员. 主要研究方向为语音识别. E-mail: jeliang@hitic.ia.ac.cn
(**LIANG Jia-En** Assistant professor at the Institute of Automation, Chinese Academy of Sciences. His main research interest is speech recognition.)

徐 波 中国科学院自动化研究所研究员. 主要研究方向为语音识别, 机器翻译和数字内容技术. E-mail: xubo@hitic.ia.ac.cn
(**XU Bo** Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers speech recognition, machine translation, and digital content technology.)