

基于模型混淆度的模型组合算法研究

吴娅辉¹ 刘刚¹ 郭军¹

摘要 传统的声学模型训练算法如最大似然估计 (Maximum likelihood estimation, MLE), 在训练时只考虑了模型自身而没有考虑模型之间的相互影响. 为了进一步提升模型的识别效果, 区分性训练算法被提出. 本文在最小音素错误 (Minimum phone error, MPE) 区分性训练算法的基础上提出一种基于模型间混淆程度进行模型组合的算法: 针对单混合分量模型, 依据模型间混淆程度对 MLE 和 MPE 的模型进行加权组合; 针对多混合分量模型, 提出一种模型选择的算法来获取新的模型参数. 实验表明, 与 MPE 算法相比, 对单分量的情况, 该算法可以使系统的误识率相对降低 4% 左右; 对于多分量的情况, 该算法可以使系统的误识率相对降低 3% 左右.

关键词 模型加权, 区分性训练, 最小音素错误, 最大似然估计
中图分类号 TP391.4

Research on Model Combination Based on Model Confusion

WU Ya-Hui¹ LIU Gang¹ GUO Jun¹

Abstract Traditional training methods such as maximum likelihood estimation (MLE) do not consider discriminative relation between acoustic models, so some models are apt to obscure each other. In order to raise the differentiation degree between models, discriminative training criteria are proposed. This paper mainly introduces the new discriminative training method, the minimum phone error (MPE), and proposes a new model combination method. For the single mixture model, it weights the models of MPE and MLE based on the confusion of them. For the multimixture model, a model selection method is proposed. Experiments demonstrate that this method achieves consistent improvement over the conventional MPE training and the relative error reduction is about 4% for the single mixture model and about 3% for the multi mixture model.

Key words Model weighting, discriminative training, minimum phone error (MPE), maximum likelihood estimation (MLE)

声学模型训练是语音识别的一个关键环节. 传统的训练方法采用最大似然估计 (Maximum likelihood estimation, MLE) 准则, 这种方法假设训练数据足够多并且服从一定分布等条件^[1], 这在实际中常常是不能满足的. 此外, MLE 是一种生成性的训练算法, 即只用类内训练样本学习模型参数, 没有考虑类与类之间的相互影响, 因此不同模型之间的区分性得不到最大化.

与生成性训练相对应的是区分性训练. 早期的区分性训练算法主要是最大互信息估计 (Maximum mutual information estimation, MMIE)^[2]. 由于算法比较复杂, 开始时很难将其应用于大词汇连续语音识别. 后来提出了基于 Lattice 的 MMIE 训练算法^[3], 有效地解决了这一问题. MMIE 的实现可以分

为两部分: 1) 增加相应特征对模型的相似度, 这和 MLE 是一样的; 2) 要降低竞争句子的特征对模型的相似度, 这是区分性训练与 MLE 的差异之处, 即训练时不仅考虑正确句子对模型的相似度, 而且考虑了竞争句子对模型的影响, 尽量增加模型间的区分性.

1992 年 Juang 等^[4] 提出最小分类错误 (Minimum classification error, MCE) 准则, 该准则直接对损失函数最小化进而实现最小化分类错误的目的. 这种方法在训练过程中同时考虑了类外的信息, 增加了类间的区分性.

最小音素错误 (Minimum phone error, MPE) 训练准则最早由 Povey 提出^[5], 国外在这方面做了大量的相关研究^[5-11], 但是国内在这方面的研究还处于起步阶段. MPE 与 MMIE 比较相似, 两者的目标都是最大化所有句子的识别率, 不同的是 MPE 从音素层次入手, 这与识别的目标更加接近, 因此它一般较 MMIE 有更好的效果.

总的来说, 区分性训练算法更侧重于模型的分类边界, 而生成性算法侧重于模型的内部描述, 虽然区分性训练算法整体上可以取得优于 MLE 的实验结果, 但是当我们观察各个识别结果时可以发现, 对

收稿日期 2008-06-18 收修改稿日期 2008-09-18
Received June 18, 2008; in revised form September 18, 2008
国家高技术研究发展计划 (863 计划) (2006AA010102, 2007AA01Z417), 国家自然科学基金 (60705019) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA010102, 2007AA01Z417), and National Natural Science Foundation of China (60705019)
1. 北京邮电大学模式识别与智能系统实验室 北京 100876
1. Pattern Recognition and Intelligent System Laboratory, Beijing University of Posts and Telecommunications, Beijing 100876
DOI: 10.3724/SP.J.1004.2009.00551

于一些识别结果 MLE 仍然会优于区分性训练. 以 MPE 训练模型为例, 例如对 863 测试集中 F50A486 句子: “自治区党委和政府要求有关地区”, MLE 的识别结果为: “自治区党委和政府要求有关地区与”, MPE 的识别结果为: “飞区党委和政府要求有关地区与”, 这里 MLE 出现了一个插入错误, 而 MPE 出现了一个替代错误和一个插入错误, 这样的例子在识别集合中还有很多, 这说明, 经过 MPE 的训练, 可以使总体上模型间的区分性有了提高, 但是对于一些具体模型来说, MLE 的模型仍会比 MPE 更加适合. 在实际中二者各有优势, 如果可以将二者结合应该可以得到更好的效果.

基于上面的分析, 本文提出一种基于模型混淆程度进行 MPE 和 MLE 模型组合的算法, 该算法直接从模型一级进行加权优化, 调整模型的参数, 进一步提高了模型的鲁棒性. 本文第 1 节介绍了 MPE 目标函数及其更新算法, 第 2 节介绍了本文提出的基于模型混淆程度进行两种模型组合算法的实现, 最后给出了实验结果和今后的研究方向.

1 MPE 目标函数

设训练语料为 $\mathbf{O} = \{\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_R\}$, 则 MPE 目标函数可表示为

$$F_{\text{MPE}}(\lambda) = \sum_{r=1}^R \sum_s P_\lambda(s|\mathbf{O}_r) A(s, s_r) \quad (1)$$

式中 λ 表示声学模型, $P_\lambda(s|\mathbf{O}_r)$ 为句子 s 的后验概率: $P_\lambda(s|\mathbf{O}_r) = \frac{P_\lambda(\mathbf{O}_r|s)P(s)}{\sum_{u \in P_{\text{Lat}}} P_\lambda(\mathbf{O}_r|u)P(u)}$, P_{Lat} 表示竞争句子的集合, 这里用 Lattice 表示, \mathbf{O}_r 表示第 r 句训练语句的声学特征, s_r 表示 \mathbf{O}_r 对应的正确语句, $A(s, s_r)$ 代表根据正确句子 s_r 计算出来语句 s 的音素正确率. 实际中一般采用下面的近似算法实现^[5]:

$$A(q) = \max_z \begin{cases} -1 + 2e(q, z), & \text{如果 } q, z \text{ 为相同音素} \\ -1 + e(q, z), & \text{如果 } q, z \text{ 为不同音素} \end{cases} \quad (2)$$

其中 q 是音素图上某段音素, z 是正确句子上的参考音素, $e(q, z)$ 是音素 z 被音素 q 重迭的比例.

MPE 的目标函数不容易直接采用梯度下降算法来求解模型参数, 因此必须采用弱势辅助函数的

方法, 通过辅助函数的最大化来实现其目标函数的最大化. 这里首先简要的介绍一下弱势辅助函数的概念.

$F(\lambda)$ 在 $\bar{\lambda}$ 处的弱势辅助函数 $H(\lambda, \bar{\lambda})$ 满足下式

$$\frac{\partial}{\partial \lambda} H(\lambda, \bar{\lambda})|_{\lambda=\bar{\lambda}} = \frac{\partial}{\partial \lambda} F(\lambda)|_{\lambda=\bar{\lambda}} \quad (3)$$

可见, 弱势辅助函数只要求辅助函数与目标函数在 $\bar{\lambda}$ 处有相同斜率, 因此未必能保证辅助函数收敛时目标函数一定收敛, 解决方法就是增加一个平滑函数来加快弱势辅助函数的收敛.

平滑函数的条件为

$$H^{\text{sm}}(\lambda, \bar{\lambda}) \leq H^{\text{sm}}(\bar{\lambda}, \bar{\lambda}) \quad (4)$$

因此, 平滑函数只要求其在 $\bar{\lambda}$ 处有最大值. 如果将弱势辅助函数与平滑函数相加, 可以实现在不改变弱势辅助函数条件的同时, 将其以 $\bar{\lambda}$ 为基点向两边压缩, 加快收敛. 虽然加上平滑函数仍无法保证弱势辅助函数一定收敛, 但实验证明该方法基本可以得到比较好的效果.

经过推导, MPE 的弱势辅助函数加上平滑函数后可以表示为式 (5) (见本页下方)^[5], 其中 $\bar{\lambda}$ 表示上一轮的模型参数; s_q 代表音素 q 的开始时间; e_q 代表音素 q 的结束时间; j 代表音素 q 第 j 个状态; $\gamma_{qj}^r(t)$ 代表在第 r 句训练语料中, 音素 q 第 j 个状态的高斯分布在时间 t 的占有概率; D_j 为平滑系数, 控制模型参数更新的程度, 一般 D_j 越小, 模型更新越快, 但是过小的值也可能造成系统不稳定; γ_q^{MPE} 称为最小音素错误的权重, 定义为

$$\gamma_q^{\text{MPE}} = \frac{\partial F_{\text{MPE}}}{\partial \log p(\mathbf{O}|q)} = \gamma_q^r(c_r(q) - c_{\text{avg}}^r) \quad (6)$$

式中 γ_q^r 代表所有经过音素 q 的句子在所有可能句子中的比例. $c_r(q)$ 代表所有经过音素 q 的句子的期望正确率. c_{avg}^r 代表所有可能句子的期望正确率. γ_q^{MPE} 的意义为, 如果音素 q 的 γ_q^{MPE} 大于零, 表示音素 q 在第 r 句训练语料中做的是正向训练, 并且此正数越大表示正向训练的程度越大; 反之, 如果音素 q 的 γ_q^{MPE} 小于零, 表示音素 q 在第 r 句训练语料中做的是负向训练, 并且此负数越小表示负向训练的程度越大. 这些值可以通过在 Lattice^[5] 上采用前向后向算法有效地计算.

$$g_{\text{MPE}}(\lambda, \bar{\lambda}) = \sum_{r=1}^R \sum_{u \in P_{\text{Lat}}} \sum_{j=1}^J \sum_{t=s_q}^{e_q} \gamma_q^{\text{MPE}} \gamma_{qj}^r(t) \log N(\mathbf{O}_r(t), \boldsymbol{\mu}_j, \sigma_j^2) + \sum_j - \frac{1}{2} \left(D_j \log(2\pi\sigma_j^2) + \frac{D_j(\bar{\mu}_j^2 + \bar{\sigma}_j^2) - 2D_j\bar{\mu}_j\mu_j + D_j\mu_j^2}{\sigma_j^2} \right) \quad (5)$$

由此辅助函数可以推导出模型参数更新公式, 扩展为具有 M_j 个高斯分布的情况, 其中 m 表示状态 j 中第 m 个高斯分布:

$$\boldsymbol{\mu}_{jm} = \frac{\boldsymbol{\theta}_{jm}^{\text{num}}(\mathbf{O}) - \boldsymbol{\theta}_{jm}^{\text{den}}(\mathbf{O}) + D_{jm}\bar{\boldsymbol{\mu}}_{jm}}{\gamma_{jm}^{\text{num}} - \gamma_{jm}^{\text{den}} + D_{jm}} \quad (7)$$

$$\sigma_{jm}^2 = \frac{\boldsymbol{\theta}_{jm}^{\text{num}}(\mathbf{O}^2) - \boldsymbol{\theta}_{jm}^{\text{den}}(\mathbf{O}^2) + D_{jm}(\bar{\boldsymbol{\mu}}_{jm}^2 + \bar{\sigma}_{jm}^2)}{\gamma_{jm}^{\text{num}} - \gamma_{jm}^{\text{den}} + D_{jm}} - \boldsymbol{\mu}_{jm}^2 \quad (8)$$

式中, γ_{jm}^{num} , $\boldsymbol{\theta}_{jm}^{\text{num}}(\mathbf{O})$, $\boldsymbol{\theta}_{jm}^{\text{num}}(\mathbf{O}^2)$ 表示正向统计信息, 代表的意义如下

$$\gamma_{jm}^{\text{num}} = \sum_{r=1}^R \sum_{u \in P_{\text{Lat}}} \sum_{t=s_q}^{e_q} \gamma_{qjm}^r(t) \max(0, \gamma_q^{\text{MPE}}) \quad (9)$$

$$\boldsymbol{\theta}_{jm}^{\text{num}}(\mathbf{O}) = \sum_{r=1}^R \sum_{u \in P_{\text{Lat}}} \sum_{t=s_q}^{e_q} \gamma_{qjm}^r(t) \max(0, \gamma_q^{\text{MPE}}) \mathbf{O}_r(t) \quad (10)$$

$$\boldsymbol{\theta}_{jm}^{\text{num}}(\mathbf{O}^2) = \sum_{r=1}^R \sum_{u \in P_{\text{Lat}}} \sum_{t=s_q}^{e_q} \gamma_{qjm}^r(t) \max(0, \gamma_q^{\text{MPE}}) \mathbf{O}_r(t)^2 \quad (11)$$

γ_{jm}^{den} , $\boldsymbol{\theta}_{jm}^{\text{den}}(\mathbf{O})$, $\boldsymbol{\theta}_{jm}^{\text{den}}(\mathbf{O}^2)$ 表示负向统计信息. 其表达式与正向的相似, 只是将上标 num 换为 den, 将 γ_q^{MPE} 换为 $-\gamma_q^{\text{MPE}}$.

最后为了提供更好的模型估计, I-平滑技术 (I-Smoothing)^[5] 被加入到 MPE 辅助函数中. I-平滑函数实际上是以最大似然估计所计算的统计信息与最小音素错误法则估测的统计信息作线性内插实现的.

2 模型混淆度加权算法

由前面的分析可以得到, MPE 训练出的模型在整体上优于 MLE, 但是模型组内的每一项不一定都优于 MLE 的结果, 因此我们希望将两种模型进行组合, 使得新的模型有更好的识别效果, 而模型组合的方法则是基于模型之间的混淆度.

对于训练出的一组隐马尔可夫模型 (Hidden Markov model, HMM), 我们可以把里面的每一个模型的均值向量都看作是一组语音特征, 那么考察该特征在其他 HMM 模型下的输出概率, 则该值可以衡量出该模型对其他 HMM 造成的混淆, 我们将各个模型在某一个模型下的所有输出概率进行累加, 则该累加值可以衡量该模型受其他所有模型混淆的程度, 并且该值越大, 说明该模型越容易被混淆, 那么其权值应该越小. 这里我们直接从声学模型的角度计算模型之间的混淆程度而不采用混淆网络进行计算, 这样可以避免混淆网络计算的结果依赖于语

料的特性, 从而使混淆的情况只依赖于声学模型本身. 具体实现算法如下.

设 $p(\mathbf{x}|\mathbf{A})$ 是特征向量 \mathbf{x} 在模型 A 下的高斯分布概率密度函数:

$$p(\mathbf{x}|\mathbf{A}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_A)^T \Sigma_A^{-1}(\mathbf{x} - \boldsymbol{\mu}_A)\right]}{(2\pi)^{\frac{k}{2}} |\Sigma_A|^{\frac{1}{2}}} \quad (12)$$

其中, $\boldsymbol{\mu}_A$ 为模型 A 的均值向量, Σ_A 为其对应的协方差矩阵, 本文中只考虑其为对角阵的情况. 模型 B 在 A 下的输出概率分布可计算为

$$p(\boldsymbol{\mu}_B|\mathbf{A}) = \frac{\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_B - \boldsymbol{\mu}_A)^T \Sigma_A^{-1}(\boldsymbol{\mu}_B - \boldsymbol{\mu}_A)\right]}{(2\pi)^{\frac{k}{2}} |\Sigma_A|^{\frac{1}{2}}} \quad (13)$$

根据上面的分析, A 模型被混淆的情况表示为

$$\lambda_A = \frac{p(\boldsymbol{\mu}_A|\mathbf{A})}{\sum_{B \in H} p(\boldsymbol{\mu}_B|\mathbf{A})} \quad (14)$$

其中, H 表示所有的 HMM 模型. 分子表示模型自身的均值在本模型上的输出, 分母表示所有的模型在该模型上的输出, 即混淆的程度, 该比值越大, 说明模型的特性越好, 因此权值也就越大.

同理可以计算 H 中所有模型被混淆的权值. 上面的计算都是针对 MPE 训练后的模型进行的, 为了表示方便, 我们将其重写为

$$\lambda_{A, \text{MPE}} = \lambda_A \quad (15)$$

另外, 还需要计算 MLE 对应的权重向量. MLE 权值的计算和 MPE 类似, 不同的是对 MLE 中各个模型计算时都是选出该模型, 然后将该模型替换 MPE 中对应的模型, 在 MPE 的模型空间上计算 MLE 的权重, 这是因为我们认为整体上 MPE 模型比 MLE 模型更加完善, 因此 MPE 所占的比重也要高于 MLE 的比重, 这种方法计算出的 MLE 的权重将更能代表模型加权算法中 MLE 的混淆程度.

MLE 的权重计算公式可表示为

$$\lambda_{A, \text{MLE}} = \frac{p_{\text{MLE}}(\boldsymbol{\mu}_A|\mathbf{A})}{\sum_{B \in H} p_{\text{MLE}}(\boldsymbol{\mu}_B|\mathbf{A})} \quad (16)$$

将对应的权重进行归一化

$$\alpha_{A, \text{MPE}} = \frac{\lambda_{A, \text{MPE}}}{\lambda_{A, \text{MPE}} + \lambda_{A, \text{MLE}}} \quad (17)$$

$$\alpha_{A, \text{MLE}} = \frac{\lambda_{A, \text{MLE}}}{\lambda_{A, \text{MPE}} + \lambda_{A, \text{MLE}}} \quad (18)$$

则新模型的均值和方差可表示为

$$\boldsymbol{\mu}_A = \alpha_{A, \text{MPE}} \boldsymbol{\mu}_{A, \text{MPE}} + \alpha_{A, \text{MLE}} \boldsymbol{\mu}_{A, \text{MLE}} \quad (19)$$

$$\sigma_A^2 = \alpha_{A, \text{MPE}} \sigma_{A, \text{MPE}}^2 + \alpha_{A, \text{MLE}} \sigma_{A, \text{MLE}}^2 \quad (20)$$

上面的分析是对单混合分量下的两个模型进行加权组合得到新的模型参数,与此类似,如果我们采用退化处理即在权值计算时直接将其进行二值化处理,将权值比较大的模型的权重直接设为1,而权值相对较小的模型的权重表示为0,则得到了一种模型选择的算法.在后面的实验中我们将给出两种算法下的实验的结果,并将其与MPE模型的识别结果进行对比.

对于多混合分量的问题,由于加权算法会受到多混合分量各项不对应的限制,模型加权很难实现,因此我们采用上面提到的简化的模型选择的方法来实现.

多混合分量下,模型B在A下的输出概率分布为

$$p(B|A) = \sum_{i=1}^M w_{B_i} p(\mu_{B_i}|A) \quad (21)$$

其中, $p(\mu_{B_i}|A)$ 表示模型B的各混合分量在模型A上的输出概率分布,它的计算公式与式(13)类似,只是将A模型扩展为多混合分量的情况, M 表示混合分量的总个数, w_{B_i} 表示模型B中第 i 个混合分量的权值.

采用与单混合分量相同的模型权值计算公式,使用模型选择算法,新模型的均值和方差可表示为

$$\mu_A = \begin{cases} \mu_{A,MPE}, & \text{若 } \alpha_{A,MPE} \geq \alpha_{A,MLE} \\ \mu_{A,MLE}, & \text{否则} \end{cases} \quad (22)$$

$$\sigma_A^2 = \begin{cases} \sigma_{A,MPE}^2, & \text{若 } \alpha_{A,MPE} \geq \alpha_{A,MLE} \\ \sigma_{A,MLE}^2, & \text{否则} \end{cases} \quad (23)$$

3 实验结果

实验的训练和测试语料分别采用863汉语连续语音数据库的男声数据和中国科学院自动化研究所的男声数据.语音数据的采样率为16kHz,为真实噪声环境下的连续朗读语音.在实验中我们使用39维的MFCC(Mel-frequency cepstrum coefficients)作为特征参数,包含能量参数,以及一阶差分和二阶差分参数.声学模型是基于决策树的参数共享的triphone模型.每个HMM模型包含3个实状态,状态转移从左至右自环无跳转.语言模型采用863语料和自动化研究所语料训练的bigram模型.采用的工具包为HTK 3.4.此外,在本实验中我们只在HMM捆绑后的状态一级上计算各个状态间的混淆程度,并且各级权值的计算只在本级状态上进行.

首先进行模型混合度为1的相关实验.图1给出了MPE和基于模型加权算法及模型选择算法下的系统的误识率比较图.图中横坐标代表不同的迭代次数下对应的MPE模型序号,MPE曲线中各点

表示该次迭代模型的误识率,其中“0”点的数值表示MPE迭代前即MLE的误识率,由图中可以计算出经过MPE训练后系统的误识率较MLE有大约6.8%的相对提高.模型加权和模型选择算法的曲线中对应的各个点分别表示该迭代次数下的MPE模型与MLE模型进行加权或模型选择后再训练得到的新模型的误识率.纵坐标表示系统字一级的误识率.

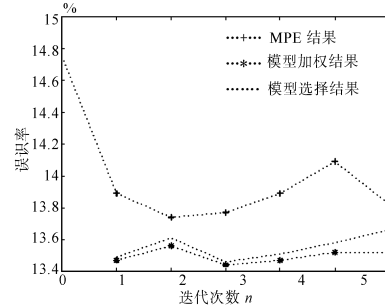


图1 单混合度下不同算法的误识率比较图

Fig. 1 Comparing the error rates with one mixture of different methods

从图1中可以看出,经过模型加权或者模型选择计算后,系统的识别率都有所提高,而且模型加权算法的性能要略优于模型选择算法的性能.同时可以发现,MPE第一次迭代后的模型与MLE的模型进行加权后的结果已经得到了系统最好的识别结果,这说明本文提出的算法可以加快系统的收敛.另外,实验发现,无论是模型加权算法还是模型选择算法都对那些本身识别效果比较差的模型的改进效果比较高,其效果要优于那些模型本身就比较好的情况下的改进结果.实验中改进最大的结果出现在第5次迭代,与MPE模型相比,系统的误识率相对降低了约4.1%.

图2给出了模型的混合分量为2时基于模型选择的实验结果与相应的MPE训练结果的对比.由图2中可以看出其与单分量下的实验有类似的变化趋势,经过MPE训练后系统的误识率较MLE相对降低了5.7%,当采用模型选择算法后,系统的误识

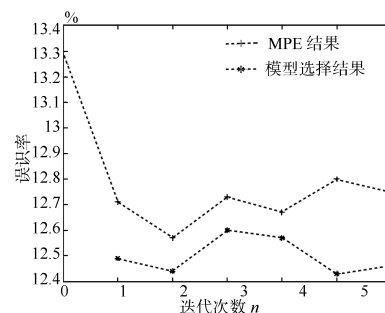


图2 2混合度下不同算法的误识率比较图

Fig. 2 Comparing the error rates with 2 mixture of different methods

率较 MPE 相对降低了 3% 左右。

图 3 给出了模型的混合度为 4 时对应的结果。经过 MPE 训练后系统的误识率较 MLE 只有相对 2% 的降低, 这说明随着混合度的提高, MPE 性能的提升越来越小, 这时两种训练算法的性能已经非常接近, 因此应用模型选择算法后, 新系统的性能与原 MPE 系统的性能很接近, 几乎没有明显的提高。

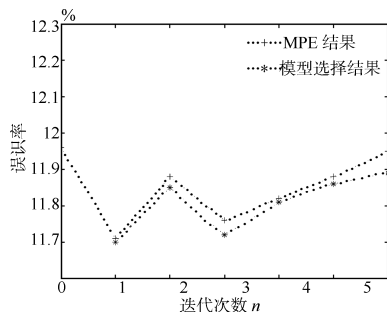


图 3 4 混合度下不同算法的误识率比较图

Fig. 3 Comparing the error rates with 4 mixture of different methods

4 结论

本文首先介绍了近年来提出的一种比较新的区分性训练算法: 最小音素错误训练算法, 并在此基础上针对区分性训练算法与生成算法一些互补的特性, 对不同混合分量的情况, 提出一种基于模型间混淆程度进行模型组合的模型加权算法和模型选择算法, 其组合的依据为该模型在整个模型空间上的混淆情况, 其混淆程度越小, 权值就越大。实验表明, 该算法可以提高模型最终的识别率。未来的研究方向包括如何进一步提高 MPE 系统的性能, 如何更有效地提取不同的加权信息, 以及如何向多混合度进行加权的扩展等。

References

- 1 Kapadia S. Discriminative Training of Hidden Markov Models [Ph. D. dissertation], Cambridge University Engineering Department, USA, 1998
- 2 Bahl L R, Brown P F, de Souza P, Mercer R. Maximum mutual information estimation of hidden Markov model parameters for speech recognition. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. New York, USA: IEEE, 1986. 49–52
- 3 Povey D, Woodland P C. Improved discriminative training techniques for large vocabulary continuous speech recognition. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. Salt Lake City, USA: Institute of Electrical and Electronics Engineers Inc, 2001. 45–48
- 4 Juang B H, Katagiri S. Discriminative learning for minimum error classification. *IEEE Transactions on Signal Processing*, 1992, 40(12): 3043–3054
- 5 Povey D. Discriminative Training for Large Vocabulary Speech Recognition [Ph. D. dissertation], Cambridge University, USA, 2004

- 6 Huang J, Povey D. Discriminatively trained features using fMPE for multi-stream audio-visual speech recognition. In: Proceedings of the 9th European Conference on Speech Communication and Technology. Lisbon, Portugal: International Speech and Communication Association, 2005. 777–780
- 7 Povey D, Kingsbury B, Manqu L, Saon G, Soltau H, Zweig G. fMPE: discriminatively trained features for speech recognition. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia, USA: IEEE, 2005. 961–964
- 8 Zheng J, Stolcke A. Improved discriminative training using phone lattices. In: Proceedings of the 9th European Conference on Speech Communication and Technology. Lisbon, Portugal: International Speech and Communication Association, 2005. 2125–2128
- 9 Huang H, Zhu J. Minimum phoneme error based filter bank analysis for speech recognition. In: Proceedings of IEEE International Conference on Multimedia and Expo. Toronto, Canada: IEEE, 2006. 1081–1084
- 10 Zhang B, Matsoukas S. Minimum phoneme error based heteroscedastic linear discriminant analysis for speech recognition. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia, USA: IEEE, 2005. 925–928
- 11 Povey D, Woodland P C. Minimum phone error and I-smoothing for improved discriminative training. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. Orlando, USA: IEEE, 2002. 105–108



吴娅辉 北京邮电大学博士研究生。2005 年获燕山大学通信工程硕士学位。主要研究方向为语音信号处理、模式识别。本文通信作者。

E-mail: annie.real.wu@gmail.com

(WU Ya-Hui Ph. D. candidate at Beijing University of Posts and Telecommunications. She received her

master degree in communication engineering from Yanshan University in 2005. Her research interest covers speech signal processing and pattern pattern. Corresponding author of this paper.)



刘刚 北京邮电大学副教授。主要研究方向为语音信号处理、模式识别。

E-mail: liugang@bupt.edu.cn

(LIU Gang Associate professor at Beijing University of Posts and Telecommunications. His research interest covers speech signal processing and pattern pattern.)



郭军 北京邮电大学教授。主要研究方向为模式识别、网络技术。

E-mail: guojun@bupt.edu.cn

(GUO Jun Professor at Beijing University of Posts and Telecommunications. His research interest covers pattern pattern and network.)