Journal of Computer Applications
计算机应用,2014,34(1):265-269
ISSN 1001-9081
CODEN JYIIDU
2014-01-10
http://www.joca.cn

文章编号:1001-9081(2014)01-0265-05
doi:10.11772/j.issn.1001-9081.2014.01.0265

鉴别性最大后验概率声学模型自适应

齐耀辉[1,2,3*],潘复平[2],葛凤培[2],颜永红[1,2]
(1.北京理工大学 信息与电子学院,北京 100081;
2.中国科学院声学研究所 中国科学院语言声学与内容理解重点实验室,北京 100190;
3.河北师范大学 物理科学与信息工程学院,石家庄 050024)
(*通信作者电子邮箱 qiyaohui@hccl.ioa.ac.cn)

摘 要:为了更加准确地估计最小音素错误最大后验概率(MPE-MAP)自适应算法中的先验分布中心,使自适应后的声学模型参数更为准确,从而提高系统的识别性能,分别采用最大互信息最大后验概率(MMI-MAP)自适应和基于最大互信息准则与最大似然准则相结合的H-criterion最大后验概率(H-MAP)自适应估计先验分布中心,提出了基于最大互信息最大后验概率先验的最小音素错误最大后验概率(MPE-MMI-MAP)和基于H-criterion最大后验概率先验的最小音素错误最大后验概率(MPE-H-MAP)算法。任务自适应实验结果表明,MPE-MMI-MAP和MPE-H-MAP算法的自适应性能均优于MPE-MAP、MMI-MAP和最大后验概率(MAP)自适应方法,分别比MPE-MAP相对提高3.4%和2.7%。

关键词:最大后验概率;鉴别性最大后验概率;最大互信息;最小音素错误;声学模型自适应
中图分类号:TN912.3 **文献标志码**:A

Discriminative maximum a posteriori for acoustic model adaptation

QI Yaohui[1,2,3*], PAN Fuping[2], GE Fengpei[2], YAN Yonghong[1,2]
(1. *College of Information and Electronics, Beijing Institute of Technology, Beijing* 100081, *China*;
2. *Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences, Beijing* 100190, *China*;
3. *College of Physics Science and Information Engineering, Hebei Normal University, Shijiazhuang Hebei* 050024, *China*)

Abstract: For Minimum Phone Error based Maximum A Posteriori (MPE-MAP) adaptation, in order to accurately estimate the center of prior distribution and to improve the recognition performance, the Maximum Mutual Information based MAP (MMI-MAP) adaptation and H-criterion, which was the interpolation of MMI and Maximum Likelihood (ML) criterion, based on MAP (H-MAP) adaptation were used for the estimation of the center of prior distribution, which led to MMI-MAP prior based MPE-MAP (MPE-MMI-MAP) and H-MAP prior based MPE-MAP (MPE-H-MAP). The experimental results of task adaptation show that the two proposed methods both can obtain better recognition performance than MPE-MAP, MMI-MAP and MAP adaptation. MPE-MMI-MAP and MPE-H-MAP can obtain 3.4% and 2.7% relative improvement over MPE-MAP respectively.

Key words: Maximum A Posteriori (MAP); Discriminative MAP (DMAP); Maximum Mutual Information (MMI); Minimum Phone Error (MPE); acoustic model adaptation

0 引言

训练环境与识别环境的不匹配是自动语音识别系统性能下降的主要原因之一。基于模型层的自适应算法利用有限的自适应数据对模型参数进行调整,逐渐将模型参数变换到实际环境,从而来提高识别系统的性能。

基于模型层的自适应方法通常分为三大类[1]:基于最大后验概率(Maximum A Posteriori, MAP)的方法、基于变换的方法和基于说话人聚类的方法。基于MAP的方法认为模型参数是符合某种先验分布的随机变量,将先验知识和从自适应数据中得到的知识结合起来估计模型参数,避免了自适应数据估计的错误。该方法有很好的渐进性,当自适应数据不断增加时,自适应效果将稳步提高。MAP算法是基于贝叶斯决策理论的,随着鉴别性准则在声学模型训练上表现出的优异性能,出现了将鉴别性准则与贝叶斯决策理论相结合的鉴别性最大后验概率自适应方法,例如最大互信息最大后验概率(Maximum Mutual Information based MAP, MMI-MAP)[2-3]、最小音素错误最大后验概率(Minimum Phone Error based MAP, MPE-MAP)[4-5]。基于变换的方法假设声学模型参数在自适应前后存在某种函数映射关系,利用自适应数据估计

收稿日期:2013-07-16;**修回日期**:2013-09-15。
基金项目:国家自然科学基金资助项目(10925419, 90920302, 11161140319, 91120001)。
作者简介:齐耀辉(1978-),女,河北石家庄人,讲师,博士研究生,主要研究方向:大词表连续语音识别; 潘复平(1977-),男,安徽阜阳人,副研究员,博士,主要研究方向:大词表连续语音识别、发音质量自动评估; 葛凤培(1982-),女,河北保定人,助理研究员,博士,主要研究方向:大词表连续语音识别、发音质量自动评估; 颜永红(1967-),男,江苏无锡人,研究员,博士生导师,博士,主要研究方向:大词表连续语音识别。

出这一映射关系,来对模型参数做出有效调整,降低模型与自适应数据间的不匹配程度。比较常用的函数映射是线性变换。最大似然线性回归(Maximum Likelihood Linear Regression, MLLR)是基于线性变换的自适应方法中的典型代表,其采用最大似然准则估计线性变换的参数。在MLLR的基础上,出现了采用最大后验概率准则和鉴别性准则估计线性变换的最大后验概率线性回归(Maximum A Posteriori Linear Regression, MAPLR)[6-7]和鉴别性线性回归,如最小音素错误线性回归(Minimum Phone Error Linear Regression, MPELR)[8-9]、最小词分类错误线性回归(Minimum Word Classification Error, MWCELR)[10]、软分类边缘估计线性回归(Soft Margin Estimation Linear Regression, SMELR)[11],以及将最大后验概率和鉴别性相结合的鉴别性最大后验概率线性回归(Discriminative Maximum A Posteriori Linear Regression, DMAPLR)[12]。与MAP相比,基于线性变换的自适应方法的渐进性较差。基于说话人聚类的方法利用多个说话人相关(Speaker Dependent, SD)模型的线性组合来得到说话人自适应(Speaker Adaptation, SA)模型,该类方法需要估计的参数最少,适合于自适应数据极少的情况。基于本征音(Eigen Voice, EV)[13]的自适应方法、基于变换矩阵线性插值[14]的自适应方法和基于参考说话人加权(Reference Speaker Weighting, RSW)[15-16]的自适应方法是比较成功的例子。

针对以隐马尔可夫模型作为建模基础的声学模型,本文研究在较多自适应数据下的自适应方法。此方法是将先验分布和对自适应数据采用最小音素错误准则估计的统计量相结合,来得到新的模型参数;与MPE-MAP方法不同的是,先验分布中的超参数不是用MAP方法得到的,而是采用鉴别性自适应方法得到。根据超参数的估计方法,分别提出了基于最大互信息最大后验概率先验的最小音素错误最大后验概率(MMI-MAP prior based MPE-MAP, MPE-MMI-MAP)和基于H-criterion最大后验概率先验的最小音素错误最大后验概率(H-MAP prior based MPE-MAP, MPE-H-MAP)方法。在连续语音识别的任务自适应实验中,两种方法的识别性能都优于MPE-MAP、MMI-MAP和MAP方法。

1 鉴别性MAP的统一表示

在声学模型参数的最大似然(Maximum Likelihood, ML)估计与鉴别性估计中,模型参数被认为是固定值。而在MAP与鉴别性MAP算法中将声学模型参数看作是随机变量,结合参数的先验分布来求新的模型参数。鉴别性MAP的目标函数可表示如下:

$$M(\lambda) = F(\lambda) + \log P(\lambda) \tag{1}$$

其中:$F(\lambda)$为鉴别性目标函数,$P(\lambda)$是模型参数λ的先验分布。该目标函数的优化需要借助辅助函数来实现。函数在任意点是它本身的弱辅助函数和强辅助函数,因此式(1)的辅助函数如下所示:

$$m(\lambda,\hat{\lambda}) = f(\lambda,\hat{\lambda}) + \log P(\lambda) \tag{2}$$

其中:对数先验分布$\log P(\lambda)$采用如式(3)所示的形式,$f(\lambda,\hat{\lambda})$为鉴别性目标函数的辅助函数,$\hat{\lambda}$为旧的模型参数。当采用最大互信息(Maximum Mutual Information, MMI)和最小音素错误(Minimum Phone Error, MPE)鉴别性准则时,$f(\lambda,\hat{\lambda})$如式(4)所示。

$$\log P(\lambda) = -\frac{1}{2}\left[\tau^l \log(2\pi\sigma_{jmd}^2) + \frac{\tau^l((\mu_{jmd}^{\text{prior}})^2 + (\sigma_{jmd}^{\text{prior}})^2) - 2\tau^l\mu_{jmd}^{\text{prior}}\mu_{jmd} + \tau^l\mu_{jmd}^2}{\sigma_{jmd}^2}\right] = Q(\tau^l,\tau^l\mu_{jmd}^{\text{prior}},\tau^l((\mu_{jmd}^{\text{prior}})^2 + (\sigma_{jmd}^{\text{prior}})^2) \mid \mu_{jmd},\sigma_{jmd}^2) \tag{3}$$

$$\begin{aligned} f(\lambda,\hat{\lambda}) = {} & \sum_j\sum_m Q(\gamma_{jm}^{\text{num}},\theta_{jmd}^{\text{num}}(O_d),\theta_{jmd}^{\text{num}}(O_d^2) \mid \mu_{jmd},\sigma_{jmd}^2) - \\ & \sum_j\sum_m Q(\gamma_{jm}^{\text{den}},\theta_{jmd}^{\text{den}}(O_d),\theta_{jmd}^{\text{den}}(O_d^2) \mid \mu_{jmd},\sigma_{jmd}^2) + \\ & \sum_j\sum_m Q(D_{jm},D_{jm}\hat{\mu}_{jmd},D_{jm}(\hat{\mu}_{jmd}^2 + \hat{\sigma}_{jmd}^2) \mid \mu_{jmd},\sigma_{jmd}^2) \end{aligned} \tag{4}$$

其中:μ_{jmd}^{prior}、$(\sigma_{jmd}^{\text{prior}})^2$是对数先验分布$\log P(\lambda)$中的超参数,$\tau^l$是控制自适应对先验信息依赖程度的参数,也可以说它是控制自适应速度的参数,D_{jm}是在每个状态的每个高斯上单独计算的平滑参数,$\theta_{jmd}(O_d)$、$\theta_{jmd}(O_d^2)$和γ_{jm}为自适应数据在第j个状态的第m个高斯上的一阶统计量、二阶统计量和占有概率。统计量与占有概率可分为两类,即num(numerator)和den(denominator)两类,它们在MMI和MPE准则中的含义不同。

将式(3)、(4)代入式(2),然后分别对均值和方差求导并让其等于零,即可得到均值和方差的更新公式:

$$\mu_{jmd} = \frac{\{\theta_{jmd}^{\text{num}}(O_d) - \theta_{jmd}^{\text{den}}(O_d)\} + D_{jm}\hat{\mu}_{jmd} + \tau^l\mu_{jmd}^{\text{prior}}}{\{\gamma_{jm}^{\text{num}} - \gamma_{jm}^{\text{den}}\} + D_{jm} + \tau^l} \tag{5}$$

$$\sigma_{jmd}^2 = \frac{\{\theta_{jmd}^{\text{num}}(O_d^2) - \theta_{jmd}^{\text{den}}(O_d^2)\} + D_{jm}(\hat{\mu}_{jmd}^2 + \hat{\sigma}_{jmd}^2)}{\{\gamma_{jm}^{\text{num}} - \gamma_{jm}^{\text{den}}\} + D_{jm} + \tau^l} + \frac{\tau^l((\mu_{jmd}^{\text{prior}})^2 + (\sigma_{jmd}^{\text{prior}})^2)}{\{\gamma_{jm}^{\text{num}} - \gamma_{jm}^{\text{den}}\} + D_{jm} + \tau^l} - \mu_{jmd}^2 \tag{6}$$

对于MPE-MAP算法而言,式(1)中的$F(\lambda)$为MPE的目标函数,如式(7)所示。

$$F_{\text{MPE}}(\lambda) = \sum_{r=1}^{R}\sum_{W_i\in W_r^{\text{lat}}} P_\lambda(W_i \mid O_r)A(W_i,W_r) = \sum_{r=1}^{R}\sum_{W_i\in W_r^{\text{lat}}} \frac{P_\lambda(O_r \mid W_i)P(W_i)A(W_i,W_r)}{\sum_{W_k\in W_r^{\text{lat}}} P_\lambda(O_r \mid W_k)P(W_k)} \tag{7}$$

其中:W_r^{lat}是识别的对应声学矢量O_r的词图,W_i、W_k表示在词图W_r^{lat}上的任意候选词序列,$A(W_i,W_r)$为给定参考文本W_r时候选词序列W_i上的音素正确率。$F_{\text{MPE}}(\lambda)$对应的辅助函数如式(4)所示。对数先验$\log P(\lambda)$中的超参数μ_{jmd}^{prior}、$(\sigma_{jmd}^{\text{prior}})^2$通过MAP的方式计算得到,其计算方法如下:

$$\mu_{jmd}^{\text{prior}} = \mu_{jmd}^{\text{MAP}} = \frac{\theta_{jmd}^{\text{ml}}(O_d) + \tau^{\text{MAP}}\mu_{jmd}^{\text{orig}}}{\gamma_{jm}^{\text{ml}} + \tau^{\text{MAP}}} \tag{8}$$

$$(\sigma_{jmd}^{\text{prior}})^2 = (\sigma_{jmd}^{\text{MAP}})^2 =$$

$$\frac{\theta_{jmd}^{\mathrm{ml}}(O_d^2)+\tau^{\mathrm{MAP}}((\mu_{jmd}^{\mathrm{orig}})^2+(\sigma_{jmd}^{\mathrm{orig}})^2)}{\gamma_{jm}^{\mathrm{ml}}+\tau^{\mathrm{MAP}}}-(\mu_{jmd}^{\mathrm{MAP}})^2 \quad (9)$$

其中:$\theta_{jmd}^{\mathrm{ml}}(O_d)$、$\theta_{jmd}^{\mathrm{ml}}(O_d^2)$ 和 $\gamma_{jm}^{\mathrm{ml}}$ 表示采用最大似然估计得到的一阶统计量、二阶统计量和占有概率,$\mu_{jmd}^{\mathrm{orig}}$、$(\sigma_{jmd}^{\mathrm{orig}})^2$ 为原始的均值和方差。将式(8)、(9)代入式(5)、(6)中,即可得到 MPE-MAP 算法中均值和方差的更新公式。辅助函数与更新公式中的 num 和 den 统计量是根据在旧的模型参数下目标函数对音素对数似然值的导数的正负来区分的:

$$\gamma_q^{\mathrm{MPE}}=\frac{1}{k}\times\frac{\partial F_{\mathrm{MPE}}(\lambda)}{\partial\log P(\boldsymbol{O}_r\mid q)}\Big|_{\lambda=\hat{\lambda}}$$

其中:num 代表 γ_q^{MPE} 为正时的统计量,den 代表 γ_q^{MPE} 为负时的统计量。各个统计量的计算可参考文献[17]。

MPE-MAP 算法进行了两级 MAP 自适应。在第一级 MAP 自适应中,先验分布中的超参数用原始的均值和方差,更新时用 ML 估计的统计量;在第二级 MAP 自适应中,用第一级 MAP 自适应中得到的均值和方差作先验分布中的超参数,更新时用 MPE 估计的统计量。

2　对 MPE-MAP 中先验分布参数估计的改进

在各种 MAP 自适应中,自适应后的均值向量实际上是先验均值与估算的样本均值向量的线性加权之和。MPE-MAP 算法用 MAP 自适应估计的均值和方差作为先验分布的中心,其最终得到的均值是 MPE 估计的均值与原始模型经 MAP 自适应后的均值的线性组合。本文提出 H-MAP 自适应,用 ML 准则与 MMI 准则的线性组合进行 MAP 自适应,实验结果表明该方法的性能优于 MAP 自适应。由于 MMI-MAP 和 H-MAP 自适应的性能要优于 MAP 自适应[2]4,所以采用 MMI-MAP 和 H-MAP 算法估计 MPE-MAP 自适应中的先验分布中心,可以得到更为准确的先验分布,从而使自适应后的模型参数更为准确。

2.1　MPE-MMI-MAP

MPE-MMI-MAP 算法是指用数据的 MMI-MAP 统计量作为先验分布的中心,来平滑 MPE 估计的模型参数。在 MPE-MMI-MAP 算法中,式(1)中的 $F(\lambda)$ 仍为 MPE 的目标函数,如式(7)所示,而对数先验 $\log P(\lambda)$ 中的超参数 $\mu_{jmd}^{\mathrm{prior}}$、$(\sigma_{jmd}^{\mathrm{prior}})^2$ 是通过 MMI-MAP 的方式计算得到。

在 MMI-MAP 算法中,式(1)中的 $F(\lambda)$ 为 MMI 准则的目标函数,如式(10)所示。对数先验 $\log P(\lambda)$ 中的超参数 $\mu_{jmd}^{\mathrm{prior}}$、$(\sigma_{jmd}^{\mathrm{prior}})^2$ 仍通过 MAP 的方式计算得到。式(2)中的 $f(\lambda,\hat{\lambda})$ 为 MMI 准则的辅助函数。MMI 准则和 MPE 准则的辅助函数具有相同的形式,但各个统计量的含义和计算方式不同。在 MMI 准则中,标示为 num 与 den 的占有概率和统计量分别是在正确路径与整个搜索空间上计算的。各个统计量的计算可参考文献[17]。因此在 MPE-MMI-MAP 中,对数先验 $\log P(\lambda)$ 中的超参数 $\mu_{jmd}^{\mathrm{prior}}$、$(\sigma_{jmd}^{\mathrm{prior}})^2$ 的计算方法如式(11)~(12)所示。

$$F_{\mathrm{MMI}}(\lambda)=\sum_{r=1}^{R}\log\frac{P(\boldsymbol{O}_r\mid W_r)P(W_r)}{\sum_i P(\boldsymbol{O}_r\mid W_i)P(W_i)} \quad (10)$$

$$\mu_{jmd}^{\mathrm{prior}}=\mu_{jmd}^{\mathrm{MMI\text{-}MAP}}=\frac{\{\theta_{jmd}^{\mathrm{num}}(O_d)-\theta_{jmd}^{\mathrm{den}}(O_d)\}+D_{jm}\hat{\mu}_{jmd}+\tau^I\mu_{jmd}^{\mathrm{MAP}}}{\{\gamma_{jm}^{\mathrm{num}}-\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I} \quad (11)$$

$$(\sigma_{jmd}^{\mathrm{prior}})^2=(\sigma_{jmd}^{\mathrm{MMI\text{-}MAP}})^2=\frac{\{\theta_{jmd}^{\mathrm{num}}(O_d^2)-\theta_{jmd}^{\mathrm{den}}(O_d^2)\}+D_{jm}(\hat{\mu}_{jmd}^2+\hat{\sigma}_{jmd}^2)}{\{\gamma_{jm}^{\mathrm{num}}-\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I}+\frac{\tau^I((\mu_{jmd}^{\mathrm{MAP}})^2+(\sigma_{jmd}^{\mathrm{MAP}})^2)}{\{\gamma_{jm}^{\mathrm{num}}-\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I}-(\mu_{jmd}^{\mathrm{MMI\text{-}MAP}})^2 \quad (12)$$

将式(11)、(12)代入式(5)、(6)中,得到 MPE-MMI-MAP 算法的均值和方差的更新公式。需要注意的是,在式(11)、(12)与式(5)、(6)中标记为 num、den 的各个量的含义和计算方法以及 τ^I 的取值不同。

2.2　MPE-H-MAP

有研究者采用 H-criterion 目标函数估计声学模型参数,以解决 MMI 估计的过训练问题。本文采用 H-criterion 目标函数进行 MAP 自适应,提出 H-MAP 算法。在 H-MAP 算法中,式(1)中的目标函数 $F(\lambda)$ 为 ML 准则与 MMI 准则的线性组合,其具体形式如式(13),对数先验分布 $\log P(\lambda)$ 中的超参数 $\mu_{jmd}^{\mathrm{prior}}$、$(\sigma_{jmd}^{\mathrm{prior}})^2$ 通过 MAP 方式计算得到。式(2)中的 $f(\lambda,\hat{\lambda})$ 如式(14)所示。将式(14)、(3)代入式(2),然后分别对均值和方差求导并让其等于零,得到 H-MAP 的参数计算方法,如式(15)、(16)所示。

$$F_{\mathrm{H}}(\lambda)=(1-\alpha)F_{\mathrm{ML}}(\lambda)+\alpha F_{\mathrm{MMI}}(\lambda) \quad (13)$$

$$f_{\mathrm{H}}(\lambda,\hat{\lambda})=\sum_j\sum_m Q(\gamma_{jmd}^{\mathrm{num}},\theta_{jmd}^{\mathrm{num}}(O_d),\theta_{jmd}^{\mathrm{num}}(O_d^2)\mid\mu_{jmd},\sigma_{jmd}^2)-\alpha\sum_j\sum_m Q(\gamma_{jmd}^{\mathrm{den}},\theta_{jmd}^{\mathrm{den}}(O_d),\theta_{jmd}^{\mathrm{den}}(O_d^2)\mid\mu_{jmd},\sigma_{jmd}^2)+\sum_j\sum_m Q(D_{jm},D_{jm}\hat{\mu}_{jmd},D_{jm}(\hat{\mu}_{jmd}^2+\hat{\sigma}_{jmd}^2)\mid\mu_{jmd},\sigma_{jmd}^2) \quad (14)$$

$$\mu_{jmd}^{\mathrm{H\text{-}MAP}}=\frac{\{\theta_{jmd}^{\mathrm{num}}(O_d)-\alpha\theta_{jmd}^{\mathrm{den}}(O_d)\}+D_{jm}\hat{\mu}_{jmd}}{\{\gamma_{jm}^{\mathrm{num}}-\alpha\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I}+\frac{\tau^I\mu_{jmd}^{\mathrm{MAP}}}{\{\gamma_{jm}^{\mathrm{num}}-\alpha\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I} \quad (15)$$

$$(\sigma_{jmd}^{\mathrm{H\text{-}MAP}})^2=\frac{\{\theta_{jmd}^{\mathrm{num}}(O_d^2)-\alpha\theta_{jmd}^{\mathrm{den}}(O_d^2)\}+D_{jm}(\hat{\mu}_{jmd}^2+\hat{\sigma}_{jmd}^2)}{\{\gamma_{jm}^{\mathrm{num}}-\alpha\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I}+\frac{\tau^I((\mu_{jmd}^{\mathrm{MAP}})^2+(\sigma_{jmd}^{\mathrm{MAP}})^2)}{\{\gamma_{jm}^{\mathrm{num}}-\alpha\gamma_{jm}^{\mathrm{den}}\}+D_{jm}+\tau^I}-(\mu_{jmd}^{\mathrm{H\text{-}MAP}})^2 \quad (16)$$

其中:num 和 den 的含义与 MMI-MAP 中相同,α 的取值是 0.8[18]。

MPE-H-MAP 算法是指用数据的 H-MAP 统计量作为先验分布的中心,来平滑 MPE 估计的模型参数。在 MPE-H-MAP 算法中,式(1)中的 $F(\lambda)$ 仍为 MPE 的目标函数,如式(7)所示。而对数先验 $\log P(\lambda)$ 中的超参数 $\mu_{jmd}^{\mathrm{prior}}$、$(\sigma_{jmd}^{\mathrm{prior}})^2$ 是通过 H-MAP的方式计算得到,其计算方法如上面所述。将式(15)、(16)代入式(5)、(6)中,得到 MPE-H-MAP 算法的均值和方差的更新公式。

在MPE-MMI-MAP算法和MPE-H-MAP算法中进行了三级MAP自适应。在第一级MAP自适应中用原始的均值和方差作为先验分布中的超参数,更新时用ML估计的统计量;在第二级和第三级MAP自适应中,先验分布中的超参数用上一级的均值和方差,更新时分别用MMI和MPE估计的统计量。

3 实验结果和评价

3.1 实验配置

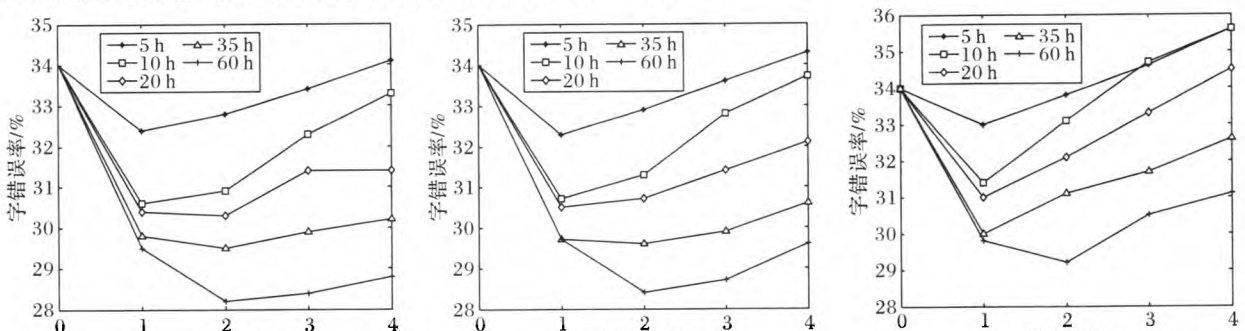
首先用实验室收集的8 kHz采样,男女平衡,大约2 000 h的语料,采用MPE方式训练了基线声学模型,然后用自适应语音分别采用MAP、MPE-MAP、MMI-MAP、H-MAP、MPE-MMI-MAP和MPE-H-MAP自适应方法对基线声学模型进行了自适应。用作自适应的语音大约有60 h,该自适应数据是实验室收集的来自网络的实际语音检索数据。为了研究各种自适应方法在不同数量的自适应数据下的性能,从60 h的自适应语音中随机抽取生成了5,10,20,35 h的自适应数据集,然后用各个自适应数据集分别采用上述6种自适应方法对基线声学模型做自适应,得到了6种自适应方法在各个自适应数据集下的声学模型。

搭建了一个大词汇量连续语音中英文混合识别系统对前文提出的鉴别性自适应方法进行了实验,其声学模型分别为上文提到的基线声学模型及各个自适应后的声学模型,然后在完全相同的解码环境下用大约2 h的测试集进行了测试。测试集是与自适应集互不重合的来自网络的实际语音检索数据。实验中所使用的特征为感知线性预测系数(Perceptual Linear Predictive, PLP),包括13维静态特征及对应的一阶、二阶、三阶差分,经过异方差线性鉴别分析(Heteroscedastic Linear Discriminant Analysis, HLDA)之后特征维数从52维降为39维。音素集中音素的个数是122,音素集的生成方法可参考文献[19]。声学模型的结构为自左向右每个音素3状态的三音子隐马尔可夫模型(Hidden Markov Model, HMM),模型经过基于决策树的状态聚类之后最终的状态数为8 655,每个状态的高斯数为40。

本文分别采用MAP、MPE-MAP、MMI-MAP、H-MAP、MPE-MMI-MAP和MPE-H-MAP算法对MPE方式训练的声学模型进行了自适应。各种自适应方法中的权重的取值是根据经验设定的。在MAP中先验权重系数为10;在MPE-MAP中,两个级别的MAP自适应中的先验权重系数分别为10,50;在MMI-MAP和H-MAP中,两个级别的MAP自适应中的先验权重系数分别为10,100;在MPE-MMI-MAP和MPE-H-MAP中,三个级别的MAP自适应中的先验权重系数分别为10,100和50。

3.2 实验结果和评价

图1是在不同数量的自适应数据下,MPE-MMI-MAP、MPE-H-MAP和MPE-MAP自适应方法的识别性能随着迭代次数的变化情况。由实验结果可见,当自适应数据比较少时,迭代1次性能最好,当自适应数据增多时,迭代2次性能最好。



(a) MPE-MMI-MAP (b) MPE-H-MAP (c) MPE-MAP

图1 3种自适应方法的识别性能随迭代次数的变化情况

对用MPE方式训练的声学模型,分别采用MAP、MPE-MAP、MMI-MAP、H-MAP、MPE-MMI-MAP和MPE-H-MAP算法进行自适应。表1列出了在不同数量的自适应数据集上6种算法在测试集上的字错误率。

表1 自适应方法字错误率对比 %

自适应方法	自适应数据集小时数					
	0	5	10	20	35	60
MAP	34.0	33.2	32.4	32.4	32.3	31.9
MPE-MAP	34.0	33.0	31.4	31.0	30.0	29.2
MMI-MAP	34.0	32.8	31.4	31.4	30.1	29.9
H-MAP	34.0	32.9	31.5	31.5	30.4	30.1
MPE-MMI-MAP	34.0	32.4	30.6	30.3	29.5	28.2
MPE-H-MAP	34.0	32.3	30.7	30.5	29.6	28.4

基线MPE模型的字错误率是34.0%。采用MAP进行自适应时只迭代了一次,因为实验发现增加迭代次数,性能并没有提高。采用其他方法进行自适应时,在不同数量的自适应数据集上,出现最好性能的迭代次数不同,表1中给出的是性能最好的识别结果。结果表明:

1)在各种数量的自适应数据集的情况下,MPE-MMI-MAP和MPE-H-MAP算法的识别性能都高于MPE-MAP算法。在60 h的自适应数据下,MPE-MMI-MAP和MPE-H-MAP分别比MPE-MAP相对提高3.4%和2.7%,比MAP相对提高11.6%和11.0%。其原因是用MMI-MAP和H-MAP比用MAP估计的先验分布中心更为准确,从而使自适应后的模型参数更为准确。

2)在各种数量的自适应数据集的情况下,MPE-MMI-MAP与MPE-H-MAP算法的识别性能相差不大。MMI-MAP与H-

MAP 算法的识别性能基本相当,因此用 MMI-MAP 与 H-MAP 估计的先验分布中心相差不大,这应该是 MPE-MMI-MAP 与 MPE-H-MAP 两种算法识别性能相当的原因。

3) MPE-MAP 与 MMI-MAP 相比,当自适应数据少时,两者识别性能相当,随着自适应数据量的增大,MPE-MAP 的性能优于 MMI-MAP。

4 结语

本文对 MPE-MAP 自适应中的先验分布中心的估计进行了研究,提出了 MPE-MMI-MAP 和 MPE-H-MAP 算法,分别采用 MMI-MAP 和基于 H-criterion 准则的 H-MAP 估计 MPE-MAP 中的先验分布。构建了大词汇量连续语音识别系统进行声学模型自适应实验,识别结果表明,在不同数量的自适应数据的情况下,MPE-MMI-MAP 与 MPE-H-MAP 均能提高系统的识别性能,两种估计先验分布参数的方法性能相差不大。

参考文献:

[1] SHINODA K. Speaker adaptation techniques for automatic speech recognition [EB/OL]. [2012-10-10]. http://www.apsipa.org/proceedings_2011/pdf/APSIPA305.pdf.

[2] POVEY D, WOODLAND P C. Discriminative MAP for acoustic model adaptation [C]// Proceedings of the 2003 IEEE International Conference on Acoustics, Speech and Signal Processing. Washington, DC: IEEE Press, 2003: 312 - 315.

[3] JIANG D N, KANEVSKY D, GOEL V, *et al.* Investigating performance of the discriminative methods for long-term speaker adaptation [C]// Proceedings of the 13th Annual Conference of the International Speech Communication Association. Lakeville: Curran Associates Inc, 2012: 1766 - 1769.

[4] POVEY D, GALES M J F, KIM D Y, *et al.* MMI-MAP and MPE-MAP for acoustic model adaptation [C]// Proceedings of the 8th European Conference on Speech Communication and Technology. Bonn: International Speech Communication Association, 2003: 1981 - 1984.

[5] MACHLICA L, ZAJIC Z, MULLER L. Discriminative adaptation based on fast combination of DMAP and DfMLLR [C]// Proceedings of the 11th Annual Conference of the International Speech Communication Association. Bonn: International Speech Communication Association, 2010: 534 - 537.

[6] HU T Y, TSAO Y, LEE L S. Discriminative fuzzy clustering maximum a posteriori linear regression for speaker adaptation [C]// Proceedings of the 13th Annual Conference of the International Speech Communication Association. Lakeville: Curran Associates Inc, 2012.

[7] TSAO Y, ISOTANI R, KAWAI H, *et al.* An environment structuring framework to facilitating suitable prior density estimation for MAPLR on robust speech recognition [C]// ISCSLP 2010: Proceedings of the 7th International Symposium on Chinese Spoken Language Processing. Piscataway, NJ: IEEE Press, 2010: 29 - 32.

[8] WANG L, WOODLAND P C. MPE - based discriminative linear transforms for speaker adaptation [J]. Computer Speech and Language, 2008, 22(3): 256 - 272.

[9] PIRHOSSEINLOO S, JAVADI S. A combination of maximum likelihood Bayesian framework and discriminative linear transforms for speaker adaptation [J]. International Journal of Information and Electronics Engineering, 2012, 2(4): 552 - 555.

[10] ZHU B, YAN Z J, HU Y, *et al.* Investigation on adaptation using different discriminative training criteria based linear regression and MAP [C]// ISCSLP 2008: Proceedings of the 6th International Symposium on Chinese Spoken Language Processing. Piscataway, NJ: IEEE Press, 2008: 93 - 96.

[11] MATSUDA S, TSAO Y, LI J, *et al.* A study on soft margin estimation of linear regression parameters for speaker adaptation [C]// Proceedings of the 10th Annual Conference of the International Speech Communication Association. Lakeville: Curran Associates Inc, 2010: 1603 - 1606.

[12] TSAO Y, ISOTANI Y, KAWAI H, *et al.* Increasing discriminative capability on MAP-based mapping function estimation for acoustic model adaptation [C]// Proceedings of International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2011: 5320 - 5323.

[13] ZHANG W L, NIU T, ZHANG L H, *et al.* Rapid speaker adaptation based on maximum-likelihood variable subspace [J]. Journal of Electronics & Information Technology, 2012, 34(3): 571 - 575. (张文林, 牛铜, 张连海, 等. 基于最大似然可变子空间的快速说话人自适应方法[J]. 电子与信息学报, 2012, 34(3): 571 - 575.)

[14] XU X H, ZHU J. Speaker adaptation with transformation matrix linear interpolation [J]. Wuhan University Journal of Natural Sciences, 2004, 9(6): 927 - 930.

[15] TENG W X, GRAVIER G, BIMBOT F, *et al.* Rapid speaker adaptation by reference model interpolation [C]// Proceedings of the 8th Annual Conference of the International Speech Communication Association. Lakeville: Curran Associates Inc, 2008: 258 - 261.

[16] TENG W X, GRAVIER G, BIMBOT F, *et al.* Speaker adaptation by variable reference model subspace and application to large vocabulary speech recognition [C]// Proceedings of International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2009: 4381 - 4384.

[17] XU R. Discriminative training of acoustic models and its application in automatic speech recognition [D]. Beijing: Chinese Academy of Sciences, Institute of Acoustics, 2009. (徐燃. 自动语音识别中声学模型鉴别性训练的研究与应用 [D]. 北京: 中国科学院声学研究所, 2009.)

[18] WOODLAND P C, POVEY D. Large scale discriminative training for speech recognition [C]// Proceedings of International Workshop on Automatic Speech Recognition. Piscataway, NJ: IEEE Press, 2000: 7 - 16.

[19] ZHANG Q Q. Mandarin-English bilingual acoustic modeling for automatic speech recognition [D]. Beijing: Chinese Academy of Science, Institute of Acoustics, 2010. (张晴晴. 中英文混合双语声学建模 [D]. 北京: 中国科学院声学研究所, 2010.)