

# 多观察序列连续隐含马尔柯夫模型的无溢出参数重估

何 强<sup>1</sup>, 毛士艺<sup>1</sup>, 张有为<sup>2</sup>

(1. 北京航空航天大学电子工程系, 北京 100083; 2. 五邑大学信息科学研究所, 广东江门 529020)

**摘 要:** 在语音识别系统的 HMM 模型训练阶段, 由于 Baum-Welch 算法中前向概率和后向概率包含大量连乘项, 计算结果数值会越来越小, 以致产生溢出. 在单观察序列情况下采用定标技术可以妥善地解决溢出问题. 在多观察序列情况下, 则会引入各序列对 HMM 的输出概率作为修正系数, 其数值很小, 溢出问题仍存在. 本文分析了溢出问题产生的原因, 针对多观察序列的情况, 将优化目标函数由输出概率的连乘改为对数累加和形式, 推导出一套改进的 Baum-Welch 算法. 该算法降低了 HMM 参数重估算法的计算复杂度, 提高了稳定性, 避免了溢出问题.

**关键词:** 隐含马尔柯夫过程; Baum-Welch 算法; 溢出; 定标

**中图分类号:** TN912.3 **文献标识码:** A **文章编号:** 0372-2112 (2000) 10-0098-04

Re-estimation of Continuous Hidden Markov Model with Multiple Observation without Overflow

HE Qiang<sup>1</sup>, MAO Shi-yi<sup>1</sup>, ZHANG You-wei<sup>2</sup>

(1. Dept. of EE, Beijing University of Aeronautics and Astronautics, Beijing 100083, China)

(2. Research Institute of Information Science, Wuyi University, Jiangmen 529020, China)

**Abstract:** In the training phase of HMM system, due to the cumulative production, the evaluation of the forward and backward probabilities in Baum-Welch algorithm needs a large dynamic range which will exceed the precision range of essentially any machine. This can be resolved by multiplying the forward and backward probability with scaling coefficients in single observation case. In multiple observation case, generally the output probability terms are introduced for each observation, which will cause the overflow problem again. In this paper, the cause of overflow problem is studied and a revised BW algorithm specially for multiple observation is deduced by redefining the optimization object function from cumulative production to sum of logarithms. The revised HMM parameter re-estimating algorithm is more stable and effective, and the overflow problem is eliminated.

**Key words:** HMM; Baum-Welch algorithm; overflow; scaling

## 1 引言

在语音识别系统的隐含马尔柯夫模型 (HMM) 训练阶段, 一般采用 Baum-Welch 算法对原始语音参数进行训练<sup>[1~3]</sup>. 在迭代计算中, 需要采用前向概率  $f_t(i)$  和后向概率  $b_t(j)$  来计算观察序列对当前 HMM 的输出概率  $P$ . 由于  $f_t(i)$ 、 $b_t(j)$  是概率连乘的形式, 当观察序列较长时, 得到的数值就会越来越小以至溢出, 因此必须进行定标 (Scaling). 在计算  $f_t(i)$  和  $b_t(j)$  的过程中, 动态地乘以定标系数从而将其归一化, 并利用参数重估公式的分子和分母包含相同的定标系数的特点, 通过对消来防止溢出<sup>[3]</sup>.

实际应用中必须采用多观察序列进行训练, 而多观察序列情况仍有溢出问题, 并且不能简单地套用单观察序列的定标算法. 本文分析了溢出问题产生的原因, 以及以往解决多观察序列参数重估溢出问题的方法, 将优化目标函数定义为对数累加和的形式, 推导了多观察序列无溢出问题的重估公式.

## 2 标准 Baum-Welch 算法及定标公式

本文以连续混和高斯概率密度函数 HMM 为例, 讨论 Baum-Welch 算法中的定标问题. HMM 为单链结构, 有  $L$  个状态, 每个状态有  $M$  个混和高斯概率密度函数 (pdf). 观察序列  $O$  长为  $T$ ,  $O = [o_1, o_2, \dots, o_T]$ ,  $o_t$  为观察向量, 维数是  $V$ .

### 2.1 单观察序列 Baum-Welch 重估公式

首先回顾一下标准算法中前向概率、后向概率的迭代公式以及转移概率的重估公式<sup>[1~3]</sup>.

前向概率:

$$f_1(i) = \pi_i b_i(o_1), 1 \leq i \leq L \quad (1a)$$

$$f_t(i) = \sum_{j=1}^L f_{t-1}(j) a_{ji} b_i(o_t), 2 \leq t \leq T, 1 \leq i \leq L \quad (1b)$$

后向概率:

$$b_T(i) = 1, 1 \leq i \leq L \quad (2a)$$

$$b_t(i) = \sum_{j=1}^L a_{ij} b_j(o_{t+1}) b_{t+1}(j), t < T, 1 \leq i \leq L \quad (2b)$$

单观察序列情况下的转移概率重估公式为:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} {}_t(i) a_{ij} b_j(o_{t+1}) {}_{t+1}(j)}{\sum_{t=1}^{T-1} {}_t(i) {}_t(i)}, 1 \leq i, j \leq L \quad (3)$$

其中:

$$b_l(o_t) = \prod_{j=1}^M w_{lj} p_{lj}(o_t), 1 \leq l \leq L, 1 \leq t \leq T \quad (4)$$

是观察向量  $o_t$  对第  $l$  个 HMM 状态的输出概率, 而  $p_{lj}(o) \sim N(\mu_{lj}, U_{lj})$  是第  $l$  状态的第  $j$  个 pdf. 由于递推公式 (1b) 和 (2b) 含有大量的连乘项, 如果不加处理,  ${}_t(i)$ 、 ${}_t(j)$  在  $t$  较大时就会出现溢出.

## 2.2 单观察序列定标技术

采用定标技术处理前后, 前向概率和后向概率的变化如下<sup>[3]</sup>:

$$\hat{\alpha}_t(i) = \left[ \prod_{s=1}^t c_s \right] {}_t(i) = C_t {}_t(i), 1 \leq i \leq L, 1 \leq t \leq T \quad (5a)$$

$$\hat{\alpha}_t(j) = \left[ \prod_{s=t}^T c_s \right] {}_t(j) = D_t {}_t(j), 1 \leq j \leq L, 1 \leq t \leq T \quad (5b)$$

其中定标系数  $c_t$  由前向概率计算<sup>[3]</sup>. 后向概率与前向概率采用共同的定标系数. 定标公式 (5) 代入式 (3), 经过对消之后, 得到转移概率重估公式为:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) a_{ij} b_j(o_{t+1}) \hat{\alpha}_{t+1}(j)}{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) a_{il} b_l(o_{t+1}) \hat{\alpha}_{t+1}(l)} \quad (6)$$

## 2.3 传统多观察序列定标

多观察序列情况下, 总和输出概率定义为:

$$P = \prod_{k=1}^K P(O^{(k)} | ) = \prod_{k=1}^K P_k \quad (7)$$

由于式 (6) 是基于事件发生概率的, 将不同观察序列对应相同事件的概率相加, 即可得到多观察序列时的重估公式<sup>[3]</sup>. 这样由式 (3) 就可以得到多观察序列情况下转移概率的重估公式:

$$\bar{a}_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} {}_t^{(k)}(i) a_{ij} b_j(o_{t+1}^{(k)}) {}_{t+1}^{(k)}(j)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} {}_t^{(k)}(i) {}_t^{(k)}(i)}, 1 \leq i, j \leq L \quad (8)$$

其中  $K$  为序列总数,  $T_k$  为第  $k$  个序列的长度. 这里  ${}_t(i)$ 、 ${}_t(j)$  是没有经过定标处理的.

式 (5) 代入式 (8) 得:

$$\begin{aligned} \bar{a}_{ij} &= \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \frac{1}{C_t^{(k)}} \hat{\alpha}_t^{(k)}(i) a_{ij} b_j(o_{t+1}^{(k)}) \frac{1}{D_{t+1}^{(k)}} \hat{\alpha}_{t+1}^{(k)}(j)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \frac{1}{C_t^{(k)}} \hat{\alpha}_t^{(k)}(i) \frac{1}{D_t^{(k)}} \hat{\alpha}_t^{(k)}(i)} \\ &= \frac{\sum_{k=1}^K \frac{1}{C_t^{(k)} D_{t+1}^{(k)}} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^{(k)}(i) a_{ij} b_j(o_{t+1}^{(k)}) \hat{\alpha}_{t+1}^{(k)}(j)}{\sum_{k=1}^K \frac{1}{C_t^{(k)} D_t^{(k)}} \sum_{t=1}^{T_k-1} \hat{\alpha}_t^{(k)}(i) \hat{\alpha}_t^{(k)}(i) / c_t^{(k)}} \quad (9) \end{aligned}$$

其中  $C_t^{(k)} D_{t+1}^{(k)} = P_k^{-1}$ , 即  $O^{(k)}$  对 的输出概率  $P(O^{(k)} | )$  的倒数, 对应于不同序列其数值不同, 因而不能对消.

采用式 (9) 计算转移概率可以保证得到的数值与未定标前保持一致, 同时考虑了  ${}_t(i)$ 、 ${}_t(j)$  的定标问题. 但是在实

际计算中, 由于  $P_k$  含有概率的连乘, 其动态范围仍会很大而导致溢出, 因此必须进行处理. 比如忽略数值太小的因式, 或者设法将上式采用对数方法表示. 这些措施增加了计算的复杂度, 并给计算造成不稳定因素. 实际上, 这种方法是消除了  ${}_t(i)$ 、 ${}_t(j)$  的溢出问题, 却又引入了  $P_k$  的溢出问题, 因此溢出的可能仍然存在.

值得说明的是, 在识别阶段的 Viterbi 算法中, 由于是采用概率的对数累加和代替概率的连乘来计算  $\log(P_k)$ , 动态范围小, 因而没有训练阶段中的溢出问题.

## 3 改进的 Baum-Welch 算法重估公式

从式 (9) 可知, 分子和分母都含有  $P_k$ , 动态范围大, 如果能够设法消除, 也就解决了溢出问题. 由下面的分析可知, 恰当地定义优化目标函数所推导得到的重估公式中, 可以避免  $P_k$  显示地出现, 从而避免溢出.

### 3.1 转移概率的重估

这里用 Lagrange 数乘法推导重估公式<sup>[1,5]</sup>. 定义新的优化目标函数为:

$$\begin{aligned} \log(P) &= \log \left[ \prod_{k=1}^K P(O^{(k)} | ) \right] = \sum_{k=1}^K \log [P(O^{(k)} | )] \\ &= \sum_{k=1}^K \log(P_k) \end{aligned} \quad (10)$$

可以得到:

$$\begin{aligned} a_{ij} &= \frac{a_{ij} \frac{\partial}{\partial a_{ij}} \log(P)}{\sum_{l=1}^L a_{il} \frac{\partial}{\partial a_{il}} \log(P)} = \frac{a_{ij} \frac{\partial}{\partial a_{ij}} \left[ \sum_{k=1}^K \log(P_k) \right]}{\sum_{l=1}^L a_{il} \frac{\partial}{\partial a_{il}} \left[ \sum_{k=1}^K \log(P_k) \right]} \\ &= \frac{a_{ij} \frac{\partial}{\partial a_{ij}} \left[ \sum_{k=1}^K \frac{1}{P_k} \frac{\partial P_k}{\partial a_{ij}} \right]}{\sum_{l=1}^L a_{il} \frac{\partial}{\partial a_{il}} \left[ \sum_{k=1}^K \frac{1}{P_k} \frac{\partial P_k}{\partial a_{il}} \right]} \quad (11) \end{aligned}$$

可以证明<sup>[5]</sup>:

$$\frac{\partial P_k}{\partial a_{ij}} = \sum_{t=1}^{T_k-1} {}_t^{(k)}(i) b_l(o_{t+1}^{(k)}) {}_{t+1}^{(k)}(j) \quad (12)$$

式 (12) 代入式 (11) 得:

$$a_{ij} = \frac{\sum_{k=1}^K \frac{1}{P_k} \sum_{t=1}^{T_k-1} {}_t^{(k)}(i) a_{ij} b_l(o_{t+1}^{(k)}) {}_{t+1}^{(k)}(j)}{\sum_{k=1}^K \frac{1}{P_k} \sum_{l=1}^L \sum_{t=1}^{T_k-1} {}_t^{(k)}(i) a_{il} b_l(o_{t+1}^{(k)}) {}_{t+1}^{(k)}(j)} \quad (13)$$

式中  $1/P_k$  由对数操作产生, 正是系数对消所需要的. 这里的  ${}_t(i)$ 、 ${}_t(j)$  仍是未定标的. 此式形式比较复杂, 下面进行变形和整理. 由于:

$$P_k = \prod_{i=1}^L {}_t^{(k)}(i) {}_t^{(k)}(i) \quad (14)$$

定义过渡概率  ${}_t(i, j)$  为观察序列在时刻  $t$  处于状态  $i$ , 在时刻  $t+1$  处于状态  $j$  的概率, 即:

$$\begin{aligned} {}_t(i, j) &= P(q_t = i, q_{t+1} = j | O, ) = \frac{P(q_t = i, q_{t+1} = j, O | )}{P(O | )} \\ &= \frac{{}_t(i) a_{ij} b_j(o_{t+1}) {}_{t+1}(j)}{\sum_{i=1}^L {}_t(i) {}_t(i)} \quad (15) \end{aligned}$$

这是一个条件概率, 尽管  $\iota(i)$ 、 $\iota(j)$  有溢出可能, 但是分子分母处于相同数量级, 其动态范围是有限的. 可得转移概率的重估公式为:

$$\bar{a}_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{l=1}^L \iota^{(k)}(i, j)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{l=1}^L \iota^{(k)}(i, l)} \quad (16)$$

这样只需对  $\iota(i, j)$  进行定标即可. 将式(5)代入式(16)可得:

$$\iota^{(k)}(i, j) = \frac{\sum_{l=1}^L \Lambda_t^{(k)}(i) a_{il} b_l(o_{t+1}^{(k)}) \Lambda_{t+1}^{(k)}(j) c_t^{(k)}}{\sum_{l=1}^L \Lambda_t^{(k)}(l) \Lambda_{t+1}^{(k)}(l)} \quad (17)$$

该式分子和分母处于相同的数量级, 而且都是经过定标处理的, 不存在溢出问题. 式(16)和式(17)就是多观察序列情况下转移概率的无溢出重估公式.

### 3.2 混合高斯元参数的重估

对于连续混和高斯密度 HMM, 还必须对各 pdf 的均值向量、方差矩阵, 以及每个 HMM 状态对各 pdf 的权系数进行重估. 同样, 用 Lagrange 数乘法可以得到相应的公式.

首先给出观察序列  $o$  在时刻  $t$  相对于状态  $l$  中第  $j$  个 pdf 的条件概率<sup>[3]</sup>:

$$\iota(l, j) = \frac{\sum_{i=1}^L \Lambda_i(l) \Lambda_i(l)}{\sum_{i=1}^L \Lambda_i(l) \Lambda_i(i)} \cdot \frac{w_{lj} N(o_t, \mu_{lj}, U_{lj})}{\sum_{i=1}^M w_{li} N(o_t, \mu_{li}, U_{li})} \quad (18)$$

该式是已经定标之后的形式, 据此很容易得到下面的重估公式.

状态  $l$  第  $j$  个 pdf 的权值:

$$\bar{w}_{lj} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{i=1}^L \iota^{(k)}(l, j)}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{i=1}^L \iota^{(k)}(l, i)} \quad (19)$$

状态  $l$  第  $j$  个 pdf 的均值向量:

$$\bar{\mu}_{lj} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{i=1}^L \iota^{(k)}(l, j) o_t^{(k)}}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{i=1}^L \iota^{(k)}(l, j)} \quad (20)$$

状态  $l$  第  $j$  个 pdf 的方差矩阵:

$$\bar{U}_{lj} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{i=1}^L \iota^{(k)}(l, j) (o_t^{(k)} - \mu_{lj})(o_t^{(k)} - \mu_{lj})}{\sum_{k=1}^K \sum_{t=1}^{T_k-1} \sum_{i=1}^L \iota^{(k)}(l, j)} \quad (21)$$

在单观察序列的情况下,  $\iota(l, j)$  分母的数值是常数, 在以上各式的计算中可以约去, 从而可以简化计算. 但是在多观察序列的情况下, 每个  $\iota(l, j)$  的分母是不同的, 不能简单地略去, 必须保留在计算公式中.

## 4 实验结果

实验表明, 该方法可以稳定地收敛. 以汉语拼音“shuang”的 40 个语音样本进行测试, 采样频率为 16kHz, 帧移 10ms, 每帧 256 点, 计算 12 阶 MFCC 参数. 试验中语音信号的初始分段是平均分为 4 段, 对应 4 个状态, 每个状态对应 3 个 pdf. 初始

分段后得到的数据用 K 均值聚类算法分类计算出各 pdf 的均值和方差. 之后采用改进的 Baum-Welch 算法进行迭代计算, 每次迭代之后, 计算所有观察序列总的 Viterbi 输出概率(对数)之和  $P_{out}$ , 当  $P_{out}$  的变化小于  $10^{-6}$  时停止迭代. 试验结果是  $P_{out}$  稳定增大至收敛. 其变化曲线如图 1. 可以看到, 两种算法收敛速度基本相当, 但是改进算法得到的总和输出概率略有提高, 表明改进算法得到的模型参数有更高的精度. 不过后续的测试表明, 两者在识别性能上没有明显区别.

表 1 和表 2 列出了采用传统重估算法和改进的无溢出重估算法完成一次迭代所需计算量的对比. 其中第一列乘数项为该行计算部分的循环次数. 新算法在重估过程的各个步骤的计算量都有减少, 而且表 1 中还未包括中间变量的对数化处理所需要的计算. 可见新的参数重估算法不仅克服了多观察序列情况下的溢出问题, 而且在算法复杂度上得到了降低.

表 1 传统 HMM 参数重估算法计算复杂度

项目	加法	乘法	高斯概率函数
中间变量 $\times K$	$T + L(T-1) + 2L^2(T-1) + 2nL(T-1) + nL^2(T-1)$	$2L + T + 3L(T-1) + 3L^2(T-1) + 6nL + nL(T-1) + 3nL^2(T-1)$	$3nL + nL(T-1) + 3nL^2(T-1)$
转移概率 $\times (L-1)$ $(L-2)/2$	$2K + 2K(T-1)$	$1 + K + 5K(T-1)$	$nK(T-1)$
pdf 参数 $\times nL$	$K(2V+3) + KT(2V+2)$	$2K+1 + 2V+4KV + KT(3V+2)$	

表 2 无溢出 HMM 参数重估算法计算复杂度

项目	加法	乘法	高斯概率参数
中间变量 $\times K$	$T + L + nL + L(T-1) + 2L^2(T-1) + 2nL(T-1)$	$2L + T + 3L(T-1) + 3L^2(T-1) + 2nL + 3nL$	$3nL + nL(T-1) + 3nL^2(T-1)$
转移概率 $\times (L-1)$ $(L-2)/2$	$2K + 2K(T-1)$	$1 + 5K(T-1)$	$nK(T-1)$
pdf 的参数 $\times nL$	$K(2V+3) + KT(2V+2)$	$2K+1 + KT(3V+2)$	

## 5 结论

由于 HMM 训练阶段采用的 Baum-Welch 算法中, 计算前向概率和后向概率需要进行大量的连乘, 容易产生溢出. 在单观察序列情况下, 通过定标技术可以解决此问题. 但是实际应用中必须采用多观察序列, 否则无法提供足够的训练样本对 HMM 模型进行充分训练. 这时如果简单地套用单观察序列的定标技术, 就会在重估公式中引入参数  $P_k$ , 溢出问题仍然存在. 本文以各观察序列输出概率对数和的形式作为新的优化

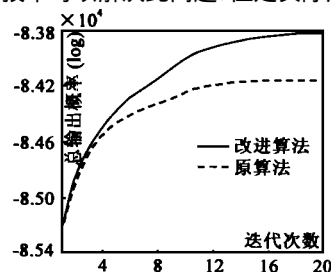


图 1 两种算法收敛性能的比较

目标函数,用 Lagrange 数乘法推导得到了无溢出多观察序列参数重估公式,由于避免了  $P_k$  的出现,也就防止了溢出的产生.实验结果表明,改进的算法是收敛的,计算公式更为简洁,计算复杂度低于原算法.

#### 参考文献:

- [ 1 ] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi. An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition [J]. B. S. T. J., 1983, 62(4): 1035 - 1074.
- [ 2 ] B. H. Juang, et al. Maximum likelihood estimation for multivariate mixture observations of markov chains [J]. IEEE Trans. IT, 1986, 32(2): 307 - 309.
- [ 3 ] L. Rabiner, B. H. Juang. Fundamentals of Speech Recognition [M]. PTR Prentice Hall, 1993.
- [ 4 ] 杨行峻,迟惠生等.语音信号数字处理 [M]. 电子工业出版社, 1995. 8.
- [ 5 ] 何强.基于 DSP 的实时语音识别理论方法和应用研究 [D]. 博士学位论文,北京航空航天大学,2000. 4

#### 作者简介:



何 强 1972 年出生,1993 和 1996 年分别获得解放军军械工程学院计算机应用专业学士学位和电路与系统专业硕士学位,1996 年起在北京航空航天大学电子工程系攻读博士学位.主要研究兴趣:信号与信息处理、语音信号处理和语音识别.



毛士艺 1935 年出生,北京航空航天大学教授,博士生导师,中国电子学会会员,主要研究领域为信号与信息处理、雷达信号处理、多传感器数据融合、数据互联与多目标跟踪.

张有为 1937 年出生,五邑大学教授,北京航空航天大学博士生导师,中国电子学会高级会员,主要研究领域为信号与信息处理、语音信号处理、电子系统分析.

(上接第 94 页)

#### 参考文献:

- [ 1 ] ITU-T Rec. M. 3010. Principles for a Telecommunication Management Network [S], 1995.
- [ 2 ] FETS 300 653. Generic Managed Object Class Library for the Network Level View [S]. 1996.
- [ 3 ] TINA Consortium. Connection Management Architecture, March [S], 1995.
- [ 4 ] D. P. Griffin and P. Georgatsos. A TMN system for VPC and routing management in ATM networks [A]. Proc. of ISINM '95 [C], London: Chapman and Hall, 1995: 356 ~ 369.
- [ 5 ] T. Saydam, J. P. Gaspoz, and P. A. Bique. Object-oriented design of a VPN bandwidth management system [A]. Proc. of ISINM '95 [C], London: Chapman and Hall, 1995: 344 - 355.
- [ 6 ] Gails A, Brianza C, and Leone C et al. Towards integrated network management for ATM and SDH networks supporting a global broadband connectivity management service [A]. Proc. of the IS&N 97 [C], 1997: 263 - 274.
- [ 7 ] W. J. Barr, T. Boyd and Y. Inoue. The TINA Initiative, v31 n3 [J]. IEEE Commun. Mag. March 1993: 70 - 76.
- [ 8 ] OMG. The Common Object Request Broker Architecture and Specification [S], 1996.
- [ 9 ] OMG. Mobile Agent Facility Specification [S], 1996.
- [ 10 ] Tianning Zhang et al. Intelligent Agents in Network and Service Management [A], INFOCOM 96 [C], 1996: 1855 - 1861.