HMM 转移概率的新的重估算法

李 健, 王作英
(清华大学电子工程系, 北京 100084)

摘 要: 将隐含马尔可夫模型(HMM:Hidden Markov Model)引入到语音识别中来是一个巨大的贡献. 但是在经典的HMM 中关于状态转移概率 $a_{ij}(i\neq j)$ 与自转移概率 a_{ii} 的独立性假设, 导致了这个模型的不协调性. 事实上, 段长分布概率与状态转移概率并非相互独立的, 由其中的一个就可以唯一的确定另外一个. 本文从段长分布概率出发说明了以上关于转移概率独立性假设的不合理性, 并得到了转移概率新的重估算法. 这个新算法比经典 HMM 的 Baum-Welch 迭代算法重估转移概率效果更好, 前者比后者相对误识率下降了大约 5%.

关键词: 语音识别; 隐含马尔可夫模型; 转移概率
中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2001) 12A-1833-03

A New Re-estimation Algorithm of HMM′s Transition Probability

LI Jian, WANG Zuo-ying
(Department of Electronic Engineering, Tsinghua University, BeiJing 100084, China)

Abstract: It is a great contribution that introduces hidden Markov model (HMM) in speech recognition. But the assumption of the traditional HMM that state transition probabilities $a_{ij}(i\neq j)$ are independent of self-trainsiton probabilty a_{ij} leads to disharmomy to the model. In fact, duration distribution probability and transition probability are dependent and one can be determined by another. From the viewpoint of duration distribution probabilty, this paper shows that the aussumption of independence of transition probability is unreasonable. And a new re-estimation algorithm of transition probability is deduced. The new algorithm performs better than the Baum-Welch algorithm. The average error reduction is about 5%.

Key words: speech recogonition; hidden Markov model (HMM); transition probability

1 引言

将隐含马尔可夫模型(HMM)引入到语音识别中来是一个巨大的贡献, 它将语音识别的研究向前推进了一大步, 并成为了当前最为流行的语音识别模型, 在语音识别的研究中占据着主导地位[1,2]. 很多成功的语音识别系统都是基于经典HMM, 象剑桥大学开发的 HTK[3]、卡内基-梅隆大学开发的 sphinx[4]等.

但是经典 HMM 在描述语音信号上仍然存在着一些缺陷. 其中一个缺陷就是状态转移概率 $a_{ij}(i\neq j)$ 与自转移概率 a_{ii} 的独立性假设. 事实上, 段长分布概率和状态转移概率并非相互独立, 由其中的一个就可以唯一决定另外一个, 由状态段长分布得出的转移概率并不是独立的. 转移概率独立性的假设导致了模型的不协调性, 使模型对语音信号的描述不够精确, 从而影响了系统的性能.

为克服经典 HMM 状态转移独立性假设的缺陷, 本文从更符合语音信号产生机理的角度出发, 根据状态段长分布得出状态转移概率, 并推导了转移概率的重估公式. 在对孤立字的试验中, 取得了比经典 HMM 的 Baum-Welch 算法进行重估更好的效果.

2 从段长分布得出转移概率

从状态驻留长度分布可以得到状态的转移概率[5]. 设系统在 i 状态驻留长度 τ 已达到 $(k-1)$, 在 k 时刻仍驻留在 i 状态的概率为 a_{ii}, 设事件 $A=\{\tau\geq k-1\}$, $B=\{\tau\geq k\}$, 则有:

$$\begin{aligned}a_{ii}&=P_i(B/A)=P_i(AB)/P_i(A)=P_i(B)/P_i(A)\\&=P_i(\tau\geq k)/P_i(\tau\geq k-1),\ k\geq 1\end{aligned}\tag{1}$$

在自左向右的 Markov 模型中, 不可逆转的顺序性使我们有理由认为当系统由状态 i 跳到状态 $i+j\,(j>1)$ 时, 等效于离开状态 i 后, 滑过了状态 $i+1$、状态 $i+2$、…和状态 $i+j-1$, 最后达到了状态 $i+j$ 才停留, 在滑过上述状态时没有停留(驻留长度为零). 因此, 在段长分布独立的条件下, 在 $k+1$ 时刻从状态 i 跨越到状态 $i+j$ 的概率为:

$$\begin{aligned}a_{i,i+j}&=(1-a_{ii})P_{i+1}(\tau=0)\cdots P_{i+j-1}(\tau=0)P_{i+j}(\tau\geq 1)\\&=(1-a_{ii})\Big[\prod_{s=i+1}^{i+j-1}(1-a_{ss})\Big]a_{i+j,i+j}\\&=\prod_{s=i}^{i+j-1}(1-a_{ss})a_{i+j,i+j},\ 1\leq j\leq N-i\end{aligned}\tag{2}$$

从上面两个个公式可以看出, HMM 的转移概率完全可以由状态驻留长度分布确定, 段长分布概率与状态转移概率并

收稿日期: 2001-07-18; 修回日期: 2001-09-20

非相互独立的. 从式(2)还可以看出, 转移概率 $a_{i,i+j}(1\le j\le N-1)$ 不是独立变量, 完全可以由自转移概率 a_{ii}, $a_{i+1,i+1}\cdots\cdots a_{i+j,i+j}$ 表示. 所以经典 HMM 的 Baum-Welch 算法[2], 关于状态转移概率与自转移概率的独立性假设是不成立的. 为此, 我们必须抛掉现有的对 a_{ij} 进行独立地训练的 Baum-Welch 算法. 新的算法只需对自转移概率 a_{ii}, $i=1,2,\dots,N$ 进行估计即可, 然后从式(2)得出向其它状态转移的转移概率. 在对孤立字的试验中, 取得了比经典 HMM 的 Baum-Welch 算法进行重估更好的效果.

3 吸收态的引入

根据式(1)和(2), 很容易得到:

$$\sum_{j=i}^{N} a_{ij}=1-(1-a_{ii})\prod_{j=i+1}^{N} p_j(\tau=0) \tag{3}$$

如果系统所有的状态都是可跨越的($P_j(\tau=0)>0, j=1,2,\dots,N$), 那么 $\sum_{j=i}^{N} a_{ij}=1$ 的约束不再满足, 所以必须引入一个吸收状态 $N+1$. 同时可以定义

$$a_{i,N+1}=1-\sum_{j=i}^{N} a_{ij}=(1-a_{ii})\prod_{j=i+1}^{N} p_j(\tau=0)=\prod_{j=i}^{N}(1-a_{ii}) \tag{4}$$

吸收态是一个实际上不输出观测矢量的状态. 在推导中为了进行统一的表达, 假定吸收态的一个输出矢量 O_e(或 O_{T+1}), 定义:

$$b_i(O_e)=\delta_{i,N+1}=\begin{cases}1, & i=N+1\\ 0, & i=1,2,\dots,N\end{cases} \tag{5}$$

$$b_{N+1}(O_t)=0,\ t=1,2,\dots,T$$

这里吸收态的引入跟一般的基于经典 HMM 的系统(比如 HTK)[3]中吸收态的引入是不一样的. HTK 引入吸收态只是为了进行模型的连接, 而没有给出理论上的依据.

4 自转移概率的重估公式的推导

由于 a_{ii} 是一个概率值, 所以对每一个 a_{ii} 都有限制: $a_{ii}\le 1$ 和 $a_{ii}\ge 0$. 这样一共有 $2N$ 组约束, 将约束条件简记为:

$$g_k(a_{11}\cdots a_{NN})\le 0,\ 1\le k\le 2N \tag{6}$$

估计的目标是使 $Prob(O/M)$ 最大, 这个概率值可以用前向后向公式表示如下

$$P=Prob(O_1O_2O_3\cdots O_TO_e/M)=\sum_{k=1}^{N+1}\alpha_t(k)\beta_t(k),\ 1\le t\le T \tag{7}$$

这是一个不等式约束的极值问题. 由不等式约束极值问题的 Kuhn-Tucker 必要条件可知, 极值点满足 $K-T$ 条件, 所以必须找到其 $K-T$ 点.

对于我们的问题, $K-T$ 条件表述为:

$$\nabla P-\sum_k w_k\nabla g_k=0 \tag{8}$$

$$w_kg_k=0,\ k=1,2,\dots,2N \tag{9}$$

$$w_k\ge 0,\ k=1,2,\dots,2N \tag{10}$$

从式(9)可知要么 $g_k=0$, 要么 $w_k=0$. 如果 $g_k=0$, 对应 $a_{ii}=0$ 或 $a_{ii}=1$. 显然为零的转移概率在自左向右的模型中是不可能的. 转移概率为 1 表明该状态只能进行自转移, 系统将永远停留在该状态, 这是不合理的. 所以 $g_k\ne 0$. 这样我们便得出 $w_k=0$. 于是 $K-T$ 条件可以表示为:

$$\nabla p=0 \tag{11}$$

那么 P 对 a_{ii}, $i=1,2,\dots,N$ 求偏导可以得到 N 个等式, 分别求解即得到 a_{ii} 的表达式.

$$\frac{\partial P}{\partial a_{ii}}=\sum_{k=1}^{N}\frac{\partial\alpha_i(k)}{\partial a_{ii}}\beta_t(k)+\sum_{k=1}^{N}\alpha_t(k)\frac{\partial\beta_i(k)}{\partial a_{ii}} \tag{12}$$

式(12)等号右边的两部分可以分别化简.

® 第一项: $\sum_{k=1}^{N}\frac{\partial\alpha_i(k)}{\partial a_{ii}}\beta_i(k)$

因为

$$\frac{\partial\alpha_i(k)}{\partial a_{ii}}=\frac{\partial\left[\sum_{l=1}^{N}\alpha_{t-1}(l)a_{lk}b_k(O_t)\right]}{\partial a_{ii}}=\sum_{l=1}^{N}\frac{\partial\alpha_{i-1}(l)}{\partial a_{ii}}a_{lk}b_k(O_t)+\sum_{l=1}^{N}\frac{\partial a_{lk}}{\partial a_{ii}}\alpha_{i-1}(l)b_k(O_t) \tag{13}$$

所以

$$\begin{aligned}\sum_{k=1}^{N}\frac{\partial\alpha_t(k)}{\partial a_{ii}}\beta_t(k)&=\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\partial\alpha_{t-1}(l)}{\partial a_{ii}}a_{lk}b_k(O_t)\beta_t(k)\\&\quad+\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\partial a_{lk}}{\partial a_{ii}}\alpha_{t-1}(l)b_k(O_t)\beta_t(k)\\&=\sum_{l=1}^{N}\frac{\partial\alpha_{t-1}(l)}{\partial a_{ii}}\beta_{t-1}(l)+\sum_{l=1}^{N}\frac{a_{li}}{a_{ii}}\alpha_{t-1}(l)b_i(O_t)\beta_t(i)\\&\quad-\sum_{k=i+1}^{N}\sum_{l=1}^{i}\frac{a_{lk}}{1-a_{ii}}\alpha_{t-1}(l)b_k(O_t)\beta_t(k)\\&=\sum_{l=1}^{N}\frac{\partial\alpha_{t-1}(l)}{\partial a_{ii}}\beta_{t-1}(l)+\frac{1}{a_{ii}}\alpha_t(i)\beta_t(i)\\&\quad-\frac{1}{1-a_{ii}}\sum_{k=i+1}^{N}\sum_{l=1}^{i}a_{lk}\alpha_{t-1}(l)b_k(O_t)\beta_t(k)\end{aligned} \tag{14}$$

同理有:

$$\begin{aligned}\sum_{l=1}^{N}\frac{\partial\alpha_{t-1}(l)}{\partial a_{ii}}\beta_{t-1}(l)&=\sum_{m=1}^{N}\frac{\partial\alpha_{t-2}(m)}{\partial a_{ii}}\beta_{t-2}(m)+\frac{1}{a_{ii}}\alpha_{t-1}(i)\\&\quad\cdot\beta_{t-1}(i)-\frac{1}{1-a_{ii}}\sum_{k=i+1}^{N}\sum_{l=1}^{N}a_{lk}\alpha_{t-2}(l)\\&\quad\cdot b_k(O_{t-1})\beta_{t-1}(k)\end{aligned} \tag{15}$$

$$\begin{aligned}\sum_{l=1}^{N}\frac{\partial\alpha_2(l)}{\partial a_{ii}}\beta_2(l)&=\sum_{m=1}^{N}\frac{\partial\alpha_1(m)}{\partial a_{ii}}\beta_1(m)+\frac{1}{a_{ii}}\alpha_2(i)\beta_2(i)\\&\quad-\frac{1}{1-a_{ii}}\sum_{k=i+1}^{N}\sum_{l=1}^{N}a_{lk}\alpha_1(l)b_k(O_2)\beta_2(k)\\&=\frac{1}{a_{ii}}\delta_{1i}\alpha_1(1)\beta_1(1)+\frac{1}{a_{ii}}\alpha_2(i)\beta_2(i)\\&\quad-\frac{1}{1-a_{ii}}\sum_{k=i+1}^{N}\sum_{l=1}^{i}a_{lk}\alpha_1(l)b_k(O_2)\beta_2(k)\end{aligned} \tag{16}$$

于是

$$\begin{aligned}\sum_{k=1}^{N}\frac{\partial\alpha_t(k)}{\partial a_{ii}}\beta_t(k)&=\frac{1}{a_{ii}}\delta_{1i}\alpha_1(1)\beta_1(1)+\frac{1}{a_{ii}}\sum_{s=2}^{t}\alpha_s(i)\beta_s(i)\\&\quad-\frac{1}{1-a_{ii}}\sum_{s=2}^{t}\sum_{k=i+1}^{N}\sum_{l=1}^{i}a_{lk}\alpha_{s-1}(l)b_k(O_s)\beta_s(k)\end{aligned} \tag{17}$$

® 第二项 $\sum_{k=1}^{N}\alpha_t(k)\frac{\partial\beta_t(k)}{\partial a_{ii}}$. 同上的推导, 可以有:

$$\sum_{k=1}^{N}\alpha_t(k)\frac{\partial\beta_t(k)}{\partial a_{ii}}=\frac{1}{a_{ii}}\sum_{s=t+1}^{T}\alpha_s(i)\beta_s(i)-\frac{1}{1-a_{ii}}\alpha_{T+1}(N+1)$$
$$-\frac{1}{1-a_{ii}}\sum_{s=t+1}^{T}\sum_{k=i+1}^{N}\sum_{l=1}^{i}a_{lk}\alpha_{s-1}(l)b_k(O_s)\beta_s(k)\quad(18)$$

将上面两部分的结果加起来就有:

$$\frac{\partial P}{\partial a_{ii}}=\frac{1}{a_{ii}}\delta_{1i}\alpha_1(1)\beta_1(1)+\frac{1}{a_{ii}}\sum_{t=2}^{T}\alpha_t(i)\beta_t(i)$$
$$-\frac{1}{1-a_{ii}}\sum_{t=2}^{T}\sum_{k=i+1}^{N}\sum_{l=1}^{i}a_{lk}\alpha_{t-1}(l)b_k(O_t)\beta_t(k)$$
$$-\frac{1}{1-a_{ii}}\alpha_{T+1}(N+1)\quad(19)$$

对于自左向右的模型来说,当 $l>k$ 时,有 $a_{lk}=0$,所以上式等号右边第三项的和部分可以化简为

$$\sum_{t=2}^{T}\sum_{k=i+1}^{N}\sum_{l=1}^{i}a_{lk}\alpha_{t-1}(l)b_k(O_t)\beta_t(k)=\sum_{l=1}^{i}\alpha_1(l)\beta_1(l)$$
$$-\sum_{k=1}^{i}\alpha_T(k)\beta_T(k)\quad(20)$$

因为$\frac{\partial P}{\partial a_{ii}}=0$,所以将(19)、(20)两式代入有

$$a_{ii}=\frac{\delta_{1i}\alpha_1(1)\beta_1(1)+\sum_{t=2}^{T}\alpha_t(i)\beta_t(i)}{\delta_{1i}\alpha_1(1)\beta_1(1)+\sum_{t=2}^{T}\alpha_t(i)\beta_t(i)+\sum_{l=1}^{i}\alpha_1(l)\beta_1(l)-\sum_{k=1}^{i}\alpha_T(k)\beta_T(k)+\alpha_{T+1}(N+1)}\quad(21)$$

在得到了自转移概率的基础上,可以由式(2)得到向其它状态转移的概率.

本文对两种不同拓扑结构的 HMM 作了试验. 试验一是每个状态可以向其后所有状态转移. 这样的模型识别效果是有限的,因为实际的汉语语音信号中"吃音"的现象不会太明显,特别是孤立字发音情况下,每个字发音都比较充分,因而一个状态一般都指向其后邻近的状态转移. 试验二是只允许一个状态向其后两个状态转移,这是国际上通用的 HMM 的拓扑结构,象 HTK、sphinx 等都采用了这种结构. 在这个试验中,a_{ii}仍由式(21)计算,$a_{i,i+1}$仍由式(2)计算,但为保证 $\sum_{j=i}^{i+2}a_{ij}=1$,转移概率 $a_{i,j+2}$不是由式(2)计算,而是由公式 $a_{i,i+2}=1-a_{ii}-a_{i,i+1}$得出.

5 实验结果

本文采用的数据集是由清华大学电子工程系语音识别实验室自行录制的,该数据集包括 50 个男性每人的 1254 个汉语孤立音节的录音数据,这些数据均为 16K 采样,16 比特量化. 提取的特征是 45 维的 MFCC 特征,其中 14 维的 MFCC 倒谱系数和它们的一阶及二阶差分特征,1 维能量特征和能量的一阶二阶差分特征.

表 **1**　两种算法的结果比较(%)

	试验一			试验二		
	Baum 算法	改进算法	误识率下降	Baum 算法	改进算法	误识率下降
$m1$	89.23	89.79	5.20	90.75	92.42	18.05
$m2$	81.10	81.18	0.42	82.22	83.49	7.14
$m3$	90.83	91.55	7.85	91.71	91.71	0.00
$m4$	83.01	84.53	8.95	83.81	84.13	1.98
$m5$	90.43	91.71	13.38	90.35	92.03	17.41
$m6$	92.11	92.42	3.93	93.14	93.62	7.00
$m7$	86.92	87.32	3.06	86.92	87.64	5.50
$m8$	87.32	87.88	4.42	87.64	88.36	5.83
$m9$	90.51	90.49	-0.21	91.71	92.50	9.53
$m10$	88.28	89.23	8.11	90.03	90.03	0.00
平均	87.97	88.61	5.29	88.83	89.59	6.85

本文进行的是与人无关的试验,试验策略是用 49 个说话人的文件作训练,另一个说话人的文件作识别. 如此轮换,作其它的文件. 表 1 中列出了对 10 个文件的试验结果. 表中的所有数据都是百分比.

两组试验的结果表明,改进算法的效果要比 Baum-welch 算法要好. 充分说明了由状态的驻留长度分布,推导出的转移概率比经典模型的转移概率更符合语音信号产生的机理. 这个结果更重要的意义还在于改进的训练算法指出了经典 HMM 所能达到的最优性能.

6 结论

本文首先从段长驻留长度推导出了状态的转移概率,论证了二者并不是相互独立的,而且一个状态向其它状态转移的概率可以完全由状态的自转移概率确定,并且在理论上说明了引入了吸收态的必要性. 从求解不等式约束极值问题的角度出发,推导出了状态的转移概率. 在对孤立字的非特定人的两组试验中,改进算法相对于 Baum 算法分别有 5.29% 和 6.85% 的误识率下降.

参考文献:

[1] L R Rabiner. A tutorial on Hidden Markov Models and selected applications in speech recognition [J]. Proc. IEEE, Feb. 1989, 77(2): 257-286.

[2] L R Rabiner, B H Juang. Fundamentals of Speech Recognition [M]. Prentice Hall, 1993.

[3] Steve Young, Dan Kershaw, et al. The HTK Book (for HTK Version 3.0). Downloaded from: < http://htk.eng.cam.ac.uk/ > .

[4] K F Lee. Automatic Speech Recognition: The Development of SPHINX System. Chapter 2 [A]. Kluwer Academic Publishers [C]. 1989

[5] 王作英. 基于段长分布的 HMM 语音识别模型 [A]. 第二届全国汉字、汉语识别会议(庐山) [C], 1989, 9.

作者简介:

李　健　男. 1974 年生于四川省遂宁市. 1997 年毕业于清华大学电子工程系,获学士学位. 1997 年于清华大学攻读博士学位.