

# 连续字符识别的级联 HMM 训练算法

赵 巍 刘家锋 唐降龙 吴 锐

(哈尔滨工业大学计算机学院 哈尔滨 150001)

**摘 要** 用较简洁的模型表达词典中成千上万的词条以及由这些词组成的类别无以计数的句子,是连续手写识别技术要解决的关键问题之一.利用 HMM 对时间序列的较强的建模能力这一特点,提出了可用于连续字符识别的 HMM 级联模型;给出了字符 HMM 模型连接概率和模型间状态转移概率的定义,并通过修正训练算法的重估公式,重估字符模型的连接参数,描述了用于手写体识别的字符 HMM 的设计方案;给出了级联训练算法重估公式和级联 Baum-Welch 训练算法描述.将级联方法应用于手写体英文单词的识别,结果表明,级联模型的描述能力与比较系统相比有明显的优势,获得了良好的识别性能.

**关键词** 连续手写字符识别;隐马尔可夫模型;级联模型;模型间状态转移;级联训练

**中图法分类号** TP391

## Cascaded HMM Training Algorithm for Continuous Character Recognition

ZHAO Wei LIU Jia-Feng TANG Xiang-Long WU Rui

(School of Computer, Harbin Institute of Technology, Harbin 150001)

**Abstract** Continuous handwritten recognition, including the recognition of sentences, words and character sequences is a new and important branch in character recognition. One of the key techniques in Continuous Handwritten Character Recognition is how to model every word entry in lexicon and every sentence that made up by the words. Due to the HMM's characteristic of time sequence modeling capability, a Cascaded Hidden Markov Models (Cascaded HMM), which defines model connection probability and state transition probability between HMMs, is proposed in this study. By modifying the Baum-Welch algorithm reestimation formula, Cascaded HMM reestimates connection parameter of character HMMs. The description of Cascaded Baum Welch Algorithm and Cascaded Viterbi Algorithm are given.

The cascaded idea of recognizing continuous character is with the strategy of free-segmentation and dynamic programming. Meanwhile, Cascaded HMMs do not model every item listed in lexicon, but combine character models as continuous text model. The cascaded HMM method could accurately describe the shape variability between adjacent characters in handwritten curve. In handwritten English word recognition task, test result shows that the cascaded model is prior to the baseline system. The method offers strong support for continuous recognition technology.

**Keywords** continuous handwritten character recognition; hidden Markov models; cascaded model; state transition between models; cascaded training

## 1 引言

连续手写体识别技术是将连续输入的不定长手写体模式从模拟形式(采样信号)转换成表达连续、正确且唯一的编码的技术. 然而, 对于词典中成千上万的词条以及由这些单词组成的类别无以计数的句子, 用尽可能少的简洁的模型来表达, 是连续手写识别技术需要解决的关键问题之一.

作为一种统计模型, HMM 被广泛应用在许多领域, 其理论基础是 1970 年由 Baum 等人建立起来的, 随后由 CMU 的 Baker 和 IBM 的 Jelinek 等人将其应用到语音识别之中. Bell 实验室 Rabiner 等人在 20 世纪 80 年代中期对 HMM 的介绍逐渐使 HMM 为世界各国从事模式识别处理的研究人员所了解和熟悉, 进而成为公认的有效的识别方法. HMM 的成功应用在于它具有较强的对时间序列的建模能力<sup>[1]</sup>. 它的缺点是自适应能力、鲁棒性不理想, 对信号底层信息(如手写信号的笔段、语音的音素等)的建模能力差; HMM 需对状态的分布作先验假设, 而这并不一定适合于待测模式; 参数空间大, 所需训练样本多, 训练时间长; 训练受初值影响较大, 易收敛于局部极值点等.

随着对 HMM 模型理论应用的不断深入, 研究者拓展了 HMM 模型, 或是将其与其它方法结合, 如结合小波域<sup>[2-3]</sup>、结合神经网络<sup>[4-5]</sup>等方法, 以改进算法的效率; 或是对 HMM 模型本身改进或修正, 如线性预测 HMM<sup>[6-7]</sup>、利用 Gibbs 分布取代 Markov 链的 HMM、在 Markov 链中考虑状态驻留时间的 HMM<sup>[8]</sup>、状态插值 HMM<sup>[9]</sup>等, 以改善模型的描述能力. 然而, 这些都无法为不定长时间序列建立独立的统计模型.

本文在建立单词的统计模型方面做了有效的尝试——提出并建立了单词级联模型, 给出了模型的数学描述.

考虑到字符上下文形变对识别准确率的影响, 可以为每个字符建立多个模型, 也可以只用一个模型. 后者在训练时一般要将字符间的形变信息记录在字符模型中, 通常是将单词样本在字符分界点位置做人工或自动标注, 再用切分后的字符训练模型参数, 从而将该字符的多种形变记录下来. 本文根据两个连接概率的定义和单词级联模型的数学描述, 通过级联训练算法得到单词级联模型  $\lambda^w = (\pi^w, A^w, B^w)$  后, 从  $\pi^w$  和  $A^w$  中还原出该单词每个字符初始状

态的概率  $\pi_{i_{r,p}}$  和转移至结束状态的概率  $a_{i_{r,p}}$ , 从而得到级联训练后的字符 HMM 模型. 这种还原同时修改了两个相邻字符的模型参数, 是字符形变的新的处理方法, 也是算法改进的关键之处.

字符 HMM 模型的连接方式有顺序连接<sup>[10-12]</sup>和循环连接<sup>[13]</sup>两种. 本文的级联方法将字符模型以顺序方式连接, 与文献[13]的 HMM 有限状态循环网(详见表 1)的循环连接方式有所不同, 又在处理字符间形变信息的方式上与其它采用顺序连接方式的识别方法有所区别.

本文针对为不定长手写体时间序列建模的问题, 根据 HMM 模型的基本理论<sup>[1,14-15]</sup>, 提出了可用于连续字符识别的 HMM 级联模型. 本文第 2 节简单介绍用于连续字符识别的字符 HMM 的建立和初始化训练算法, 由于级联 HMM 是在传统 HMM 的基础上扩展得到的, 在训练算法重估公式的描述上与传统的 BW 算法在形式上有相似之处. 为了方便对比, 这部分适当给出了 BW 算法的关键公式; 第 3 节介绍用于连续字符识别的 HMM 级联模型及两个描述级联模型的概率的定义; 第 4 节给出级联训练算法及描述; 第 5 节给出基于级联 BW 算法和级联 Viterbi 算法的连续英文单词识别测试实验及结果分析; 最后给出结论.

## 2 HMM 字符模型设计及训练

### 2.1 字符 HMM 的建立

HMM 模型的结构主要由状态以及状态之间的连接方式来描述<sup>[16-18]</sup>. 设计用于连续字符识别的 HMM 级联模型时, 充分考虑了手写体随意书写的特点——大量存在的笔段冗余和丢失的情况, 因此本文采用了带有跨越的左-右形式的 Markov 链作为字符 HMM 的基本结构, 如图 1 所示. 模型中每个状态上有一个自循环, 用以吸收该笔段上的冗余笔段, 同时每个状态不仅能够转移到下一个状态, 而且能跨越转移到该状态之后的任一状态, 利用状态的跨越来描述笔段丢失的情况.

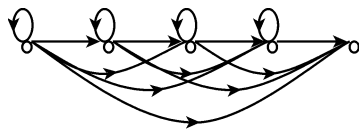


图 1 带有跨越的左右模型结构

考虑到书写过程中笔段丢失的情况非常普遍, 模型可以从任何一个状态开始发生转移, 因此设初始概率  $\pi$  的每个元素均不为 0; 同时, 为了保证字符

最后一个笔段对应的状态也可以被跨越,在模型的最后增加一个虚状态作为结束状态,模型状态转移均终止于结束状态.这样,模型的状态数表示为 $N=T_{st}+1$ ,其中, $T_{st}$ 为相应字符的标准笔段数,通过对训练样本的统计得到.

本文采用连续型的 HMM. 假设状态的输出满足高斯分布,用高斯混合模型(Gauss Mixture Model, GMM)来描述观察数据,设状态输出观察值的概率密度函数为  $B$ ,则状态  $j$  输出观察值  $o_t$  的概率为

$$b_j(o_t)=\sum_{h=1}^H C_{jh} \frac{1}{\sqrt{2\pi}|\boldsymbol{\Sigma}_{jh}|^{1/2}} \cdot \exp\left\{-\frac{1}{2}(o_t-\boldsymbol{M}_{jh})^T \boldsymbol{\Sigma}_{jh}^{-1}(o_t-\boldsymbol{M}_{jh})\right\} \tag{1}$$

其中,  $H$  为 GMM 中的高斯数;  $C_{jh}$ ,  $\boldsymbol{M}_{jh}$ ,  $\boldsymbol{\Sigma}_{jh}$  分别为第  $j$  个状态第  $h$  个高斯分量的加权值、均值矢量和协方差矩阵.

2.2 字符 HMM 的训练重估公式

本文采用 Baum-Welch 算法解决 HMM 训练. 对于给定的一个观察值序列  $O=o_1, o_2, \dots, o_T$ , 用  $\sum_{i=1}^{T-1} \xi_i(i)$  表示从状态  $i$  转移出去的次数的期望值, 而用  $\sum_{i=1}^{T-1} \xi_i(i, j)$  表示从状态  $i$  转移到状态  $j$  次数的期望值, Baum-Welch 算法的重估公式如下:

$$\bar{\pi}_i = \xi_1(i) \tag{2}$$

$$\bar{a}_{ij} = \frac{\sum_{i=1}^{T-1} \xi_i(i, j)}{\sum_{i=1}^{T-1} \xi_i(i)} \tag{3}$$

$$\bar{b}_{jh} = \frac{\sum_{i=1}^T \xi_i(j)}{\sum_{i=1}^T \xi_i(j)} \tag{4}$$

$o_t \sim N(\boldsymbol{M}_{jh}, \boldsymbol{\Sigma}_{jh})$

设  $L$  个观察序列为  $O^{(l)}, l=1, 2, \dots, L$ , 其中  $O^{(l)}=o_1^{(l)}, o_2^{(l)}, \dots, o_{T_l}^{(l)}$ , 假定各个观察序列独立, 使用多个观测序列的 Baum-Welch 算法的重估公式修正为

$$\bar{\pi}_i = \frac{\sum_{l=1}^L \alpha_1^{(l)}(i) \beta_1^{(l)}(i) / P(O^{(l)} | \lambda)}{\sum_{l=1}^L \sum_{i=1}^{T_l-1} \alpha_i^{(l)}(i) a_{ij} b_j(o_{i+1}^{(l)}) \beta_{i+1}^{(l)}(j) / P(O^{(l)} | \lambda)} \tag{5}$$

$$\bar{a}_{ij} = \frac{\sum_{l=1}^L \sum_{i=1}^{T_l-1} \alpha_i^{(l)}(i) a_{ij} b_j(o_{i+1}^{(l)}) \beta_{i+1}^{(l)}(j) / P(O^{(l)} | \lambda)}{\sum_{l=1}^L \sum_{i=1}^{T_l-1} \alpha_i^{(l)}(i) \beta_i^{(l)}(j) / P(O^{(l)} | \lambda)}, \tag{6}$$

$1 \leq i \leq N, 1 \leq j \leq N$

$$\bar{C}_{jh} = \frac{\sum_{t=1}^T \gamma_t(j, h)}{\sum_{t=1}^T \xi_t(j)} \tag{7}$$

$$\bar{\boldsymbol{M}}_{jh} = \frac{\sum_{t=1}^T o_t \gamma_t(j, h)}{\sum_{t=1}^T \gamma_t(j, h)} \tag{8}$$

$$\bar{\boldsymbol{\Sigma}}_{jh} = \frac{\sum_{t=1}^T \gamma_t(j, h) (o_t - \bar{\boldsymbol{M}}_{jh}) (o_t - \bar{\boldsymbol{M}}_{jh})^T}{\sum_{t=1}^T \gamma_t(j, h)} \tag{9}$$

其中,

$$\gamma_t(j, h) = \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} C_{jh} b_{jh}(o_t) \beta_t(j) / P(O | \lambda).$$

3 HMM 字符模型级联

级联是状态转移在字符之间的扩展. 级联过程将字符模型组合在一起, 实现对不定长序列(如连续手写体)的建模. 这种组合不是模型的简单连接, 而是在字符模型的基础上为表征笔迹在字符间的变化所设定的特殊连接. 级联示意图如图 2 所示.

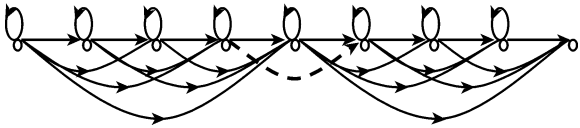


图 2 模型级联示意图

由于模型的最后一个状态为虚状态, 且为该字符模型的结束标志, 在做模型连接时这一状态将被跨越. 图 2 中示意的字符模型之间的跨越用一条虚线弧表示. 它实际上代表的是一族弧线, 表征上一模型的最后一个或几个状态可能转移至下一模型的任意状态.

由于级联过程中将用到每个字符的所有状态, 为了方便说明和避免混淆, 作如下约定:

(1) 在单词  $w=(r_1, r_2, \dots, r_p, \dots, r_R)$  中,  $r_p$  为单词的第  $p$  个字符,  $p=1, 2, \dots, R$ ,  $R$  为单词所含字符个数;

(2) 每个字符模型 Markov 链的状态数(包括结束状态)为  $N_{r_p}$ ;

(3) 设  $r_p$  和  $r_{\bar{p}}$  为单词  $w$  的两个相邻字符, 用  $i_{r_p}$  和  $j_{r_{\bar{p}}}$  分别作为字符  $r_p$  和  $r_{\bar{p}}$  模型状态的下标, 则  $i_{r_p}=1, 2, \dots, N_{r_p}$ ;  $j_{r_{\bar{p}}}=1, 2, \dots, N_{r_{\bar{p}}}$ ;  $p, \bar{p}=1, 2, \dots, R$ .

(4) 每个字符的结束状态的下标均用  $e$  表示. 给出两个描述级联模型的概率的定义.

定义 1. 字符  $r_p$  和  $r_{\bar{p}}$  相邻的概率为字符模型连接概率, 记为  $p_{r_p r_{\bar{p}}}$ .

字符  $r_p$  的后面既可以连接任意字符, 此时  $p_{r_p r_{\bar{p}}}$  值为常量, 即  $r_p$  与任意字符相邻的概率相等; 也可以连接有统计语言模型限制的字符, 如同现字符等, 此时  $p_{r_p r_{\bar{p}}}$  为字符  $r_p$  和  $r_{\bar{p}}$  同现概率, 该概率表达字符  $r_p$  和  $r_{\bar{p}}$  的合法连接, 不可能同时出现的字符其同现概率为零. 同现概率通过统计字典得到.

**定义 2.** 对相邻的字符  $r_p$  和  $r_{\bar{p}}$ , 由字符  $r_p$  的  $i$  状态转移至字符  $r_{\bar{p}}$  的  $j$  状态的转移概率称为模型间状态转移概率, 记为  $a_{i_{r_p} j_{r_{\bar{p}}}}$ .

图 2 中终止于字符  $r_p$  的模型虚状态的所有转移弧线, 可能起始于该字符模型的任意一个状态, 当级联发生时, 该弧线可能终止于字符  $r_q$  的模型的任意状态. 因此, 有三个概率值与模型间状态转移概率相关, 分别是字符  $r_p$  模型的当前状态  $i$  转移到当前字符的结束状态  $e$  的概率  $a_{i_{r_p} e}$ , 字符  $r_{\bar{p}}$  的初始状态  $j$  的概率  $\pi_{j_{r_{\bar{p}}}}$  以及模型连接概率  $p_{r_p r_{\bar{p}}}$ . 模型间状态转移概率可以由下面公式计算得到

$$a_{i_{r_p} j_{r_{\bar{p}}}} = a_{i_{r_p} e} \times \pi_{j_{r_{\bar{p}}}} \times p_{r_p r_{\bar{p}}} \quad (10)$$

这时, 标记字符结束的虚状态因字符级联而不再使用, 如图 3 所示.

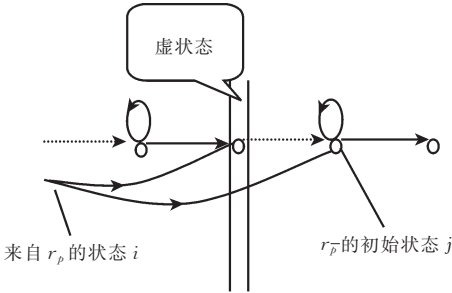


图 3 模型间状态转移示意图

级联模型建立在字符 HMM 模型的基础上, 使用 Baum-Welch 算法的重估式 (5)~(9) 和字符训练样本得到每个字符的模型  $\lambda_{r_p} = (\pi_{r_p}, \mathbf{A}_{r_p}, \mathbf{B}_{r_p})$ . 将级联后得到的模型称为级联 HMM 模型, 简称级联模型. 级联模型可表示为

$$\lambda^w = (\pi^w, \mathbf{A}^w, \mathbf{B}^w) \quad (11)$$

设某词  $w$  的级联模型有  $N^w$  个状态, 由于构成该词的字符模型的结束状态被跨越, 而级联模型需要一个全局结束状态. 因此, 除了末尾字符外, 其它字符在级联模型中的状态均比原来少了一个结束状态. 区别起见, 将级联模型中的字符模型表示为  $\lambda'_{r_p} = (\pi'_{r_p}, \mathbf{A}'_{r_p}, \mathbf{B}_{r_p})$ ,  $p=1, 2, \dots, R$ . 而级联模型状态数为  $N^w = \sum_{p=1}^R (N_{r_p} - 1) + 1$ .

式 (11) 中,  $\pi^w$  用来描述该级联模型初始状态的概率分布. 当可能的起始字符  $\hat{r}$  被确定后, 其它字符  $r (r \neq \hat{r})$  的模型状态的初始概率将全部为 0, 即这些非起始字符的各状态没有可能成为级联模型的初始状态. 这样,

$$\pi^w = \begin{pmatrix} \pi'_{r_1} \\ \pi'_{r_2} \\ \vdots \\ \pi'_{r_R} \end{pmatrix} \xrightarrow{r_1 = \hat{r}} \begin{pmatrix} \pi'_{\hat{r}} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_{N_{\hat{r}}-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (12)$$

其中,  $\pi^w$  的前  $N_{\hat{r}} - 1$  个非零元素构成的向量是字符  $\hat{r}$  的模型初始状态概率向量;  $\pi^w$  为  $N^w$  维列向量,  $N^w = \sum_{p=1}^R (N_{r_p} - 1) + 1$ ; 由于  $\pi^w$  中除了前  $N_{\hat{r}} - 1$  个元素外其余元素均为零, 因此有  $\sum_{m=1}^{N^w} \pi_m^w = \sum_{i=1}^{N_{\hat{r}}-1} \pi'_i = 1$ . 其中  $\pi_m^w$  表示  $\pi^w$  的第  $m$  个分量,  $m=1, 2, \dots, N^w$ .

$\mathbf{A}^w$  为级联后的状态转移概率矩阵,

$$\mathbf{A}^w = (a_{i_{r_p} j_{r_{\bar{p}}}}), i_{r_p} = 1, 2, \dots, N_{r_p}; j_{r_{\bar{p}}} = 1, 2, \dots, N_{r_{\bar{p}}} \quad (13)$$

一个有  $R$  个字符的单词  $w = (r_1, r_2, \dots, r_R)$ , 其状态转移概率矩阵  $\mathbf{A}^w$  可以用式 (14) 表示.

$$\mathbf{A}^w = \begin{pmatrix} \mathbf{A}'_{r_1} & \mathbf{A}_{r_1 \rightarrow r_2} & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{A}'_{r_2} & \mathbf{A}_{r_2 \rightarrow r_3} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{A}'_{r_{R-1}} & \mathbf{A}_{r_{R-1} \rightarrow r_R} \\ 0 & 0 & \dots & 0 & \mathbf{A}'_{r_R} \end{pmatrix} \quad (14)$$

当式 (13) 中字符  $r_p = r_{\bar{p}}$  时, 有  $a_{i_{r_p} j_{r_{\bar{p}}}} = a_{ij}$ ,  $i, j=1, 2, \dots, N_{r_p}$ , 表示是在一个字符  $r_p$  内部发生的模型状态转移, 其状态转移概率矩阵为  $\mathbf{A}_{r_p}$ . 由于级联时屏蔽了字符结束状态  $e$ , 将字符  $r_p$  的模型的所有终止于结束状态的弧改为跨越结束状态并转移至相邻字符  $r_{\bar{p}}$  的模型的某些状态的弧. 因此, 形式上有

$$\mathbf{A}_{r_p} = \begin{pmatrix} \mathbf{A}'_{r_p} & \mathbf{a}_e \\ \mathbf{0} & 0 \end{pmatrix}_{N_{r_p} \times N_{r_p}}, p=1, 2, \dots, R-1. \text{ 其中, } \mathbf{a}_e$$

是  $r_p$  模型中所有终止于结束状态  $e$  的转移概率构成的  $N_{r_p} - 1$  维列向量. 这样,  $\mathbf{A}^w$  中相应于字符  $r_p$  的状态转移矩阵由  $\mathbf{A}_{r_p}$  修正为维数为  $(N_{r_p} - 1) \times (N_{r_p} - 1)$  的  $\mathbf{A}'_{r_p}$ . 同时, 利用式 (10) 将  $\mathbf{a}_e$  与  $\pi_{j_{r_{\bar{p}}}}$ ,  $p_{r_p r_{\bar{p}}}$  合并得到

了  $\mathbf{A}_{r_p \rightarrow r_{\bar{p}}}, \mathbf{A}_{r_{\bar{p}} \rightarrow r_p}$  中记录了模型间的状态转移概率。

当字符  $r_p$  是单词最后一个字符  $r_R$  时, 为整个单词保留一个结束状态。因此,  $\mathbf{A}_{r_R} = \mathbf{A}'_{r_R}$ , 维数为  $N_{r_R} \times N_{r_R}$ 。

显然, 矩阵  $\mathbf{A}^w$  的维数为  $\left[ \sum_{p=1}^R (N_{r_p} - 1) + 1 \right] \times \left[ \sum_{p=1}^R (N_{r_p} - 1) + 1 \right]$ 。

$\mathbf{A}^w$  中记录了单词  $w$  每个字符每个状态上的自循环及状态的自左至右的转移、跨越以及相邻字符的模型间状态转移。因此级联模型的状态转移概率矩阵  $\mathbf{A}^w$  是字符模型内部状态转移概率矩阵和模型之间的状态转移概率矩阵块的组合, 如式(14)。

$\mathbf{B}^w$  为级联模型状态输出观察值的概率矩阵。尽管模型的扩张使相应的观察序列的长度增加, 其状态输出观察值的概率描述没有变化。并且, 由于结束状态是一个虚拟状态, 不产生观察值。因此,  $\lambda$  和  $\lambda'$  中的  $\mathbf{B}^w$  没有区别。

4 级联训练算法

级联 HMM 的训练算法在 Baum-Welch 算法的重估公式(5)~(9)的基础上修正公式, 用  $L_w$  个手写训练样本, 重估字符模型的连接参数。以手写英文单词为例说明。

4.1 级联训练算法重估公式

设当前字符为  $r_p, r_p \in X, X = \{ 'a', 'b', \dots, 'z', 'A', 'B', \dots, 'Z' \}$ 。用  $r_{\bar{p}}$  表示  $r_p$  的相邻字符。  $r_{\bar{p}}$  代表了如下 3 种情况:

- (1)  $r_{\bar{p}} \neq r_p$ : 与  $r_p$  相邻的字符, 字符不同, 显然此时  $\bar{p} \neq p$ ;
- (2)  $r_{\bar{p}} = r_p$ , 且  $\bar{p} \neq p$ : 与  $r_p$  相邻字符, 字符相同;
- (3)  $r_{\bar{p}} = r_p$ , 且  $\bar{p} = p$ : 特殊指定此时  $r_{\bar{p}}$  与  $r_p$  是同一字符。

手写单词观察序列以  $O^{(l)}$  表示,  $l = 1, 2, \dots, L_w$ , 其中  $O^{(l)} = o_1^{(l)}, o_2^{(l)}, \dots, o_{T_l}^{(l)}, O^{(l)} \in \Omega$ ,  $\Omega$  为观察序列的所有单词的集合。统计  $L_w$  个观察序列, 记

$\Omega_p = \{ w_p \mid w_p \text{ 是 } L_w \text{ 个观察序列中含字符 } r_p \text{ 的单词} \}$ ;

$\Omega_{p\bar{p}} = \{ w_{p\bar{p}} \mid w_{p\bar{p}} \text{ 是 } L_w \text{ 个观察序列中字符 } r_p \text{ 和 } r_{\bar{p}} \text{ 相邻的单词} \}$ 。

显然有  $\Omega_{p\bar{p}} \subseteq \Omega_p$ 。设这两个集合包含元素个数分别为  $L_{\Omega_p}$  和  $L_{\Omega_{p\bar{p}}}$ , 则  $L_{\Omega_{p\bar{p}}} \leq L_{\Omega_p}$ 。

由于  $\mathbf{A}^w$  的转移概率计算复杂, 这里使用  $I$  和  $J$

作为状态的广义下标。设  $t$  时刻 Markov 链所处状态为  $q_t$ , 则  $q_t \in \{ \theta_1, \dots, \theta_I, \dots, \theta_J, \dots, \theta_{N^w} \}$ 。其中  $\theta_I, \theta_J$  表示级联模型 Markov 链的第  $I$  个、第  $J$  个状态,  $1 \leq I \leq N^w, 1 \leq J \leq N^w$ 。

设  $\theta_I$  是字符  $r_p$  的模型的第  $i$  个状态,  $\theta_I$  状态转移的目标状态为  $\theta_J$ , 则有  $\theta_J = \begin{cases} \theta_j^{r_p}, & p = \bar{p} \\ \theta_j^{r_{\bar{p}}}, & p \neq \bar{p} \end{cases}$ 。即由  $\theta_I$  起始的状态转移发生在字符  $r_p$  模型内部或发生在  $r_p, r_{\bar{p}}$  模型之间。

定义  $\xi_t(I, J)$  为给定训练序列  $O$  和模型  $\lambda$  时, 级联 HMM 模型在  $t$  时刻处于状态  $I, t+1$  时刻处于状态  $J$  的概率,  $\xi_t^{(l)}(I, J) = P(q_t = \theta_I, q_{t+1} = \theta_J \mid O^{(l)}, \lambda)$ , 有

$$\xi_t^{(l)}(I, J) = \alpha_t^{(l)}(I) a_{IJ} b_J(o_{t+1}^{(l)}) \beta_{t+1}^{(l)}(J) / P(O^{(l)} \mid \lambda) \tag{15}$$

定义  $\xi_t(I)$  为级联 HMM 模型在时刻  $t$  处于  $I$  状态的概率,  $\xi_t^{(l)}(I) = P(q_t = I \mid O^{(l)}, \lambda) = \sum_J \xi_t(I, J)$ , 有

$$\xi_t^{(l)}(I) = \alpha_t^{(l)}(I) \beta_t^{(l)}(J) / P(O^{(l)} \mid \lambda) \tag{16}$$

$\sum_{t=1}^{T-1} \xi_t(I)$  表示从状态  $I$  转移出去的次数的期望值,  $\sum_{t=1}^{T-1} \xi_t(I, J)$  表示从状态  $I$  转移到状态  $J$  次数的期望值。

用式(17)~(21) 计算和更新每个字符  $r_p$  模型的参数。

$$\bar{\pi}_I = \sum_{l=1}^{L_{\Omega_p}} \xi_1^{(l)}(I) \tag{17}$$

$$\bar{a}_{IJ} = \frac{\sum_{l=1}^{L_{\Omega}} \sum_{t=1}^{T_l-1} \xi_t^{(l)}(I, J)}{\sum_{l=1}^{L_{\Omega}} \sum_{t=1}^{T_l-1} \xi_t^{(l)}(I)} \tag{18}$$

其中,

$$\bar{a}_{IJ} = \begin{cases} \bar{a}_{ij}, & 1 \leq i, j \leq N_{r_p}; & p = \bar{p} \\ \bar{a}_{i_{r_p} j_{r_{\bar{p}}}}, & 1 \leq i_{r_p} \leq N_{r_p}, 1 \leq j_{r_{\bar{p}}} \leq N_{r_{\bar{p}}}; & p \neq \bar{p} \end{cases};$$

$$L_{\Omega} = \begin{cases} L_{\Omega_p}, & p = \bar{p} \\ L_{\Omega_{p\bar{p}}}, & p \neq \bar{p} \end{cases}.$$

训练  $\mathbf{B}^w$  时, 由于  $b_J(X) = \sum_{h=1}^H C_{Jh} b_{Jh}(X) =$

$\sum_{h=1}^H C_{Jh} N(X, M_{Jh}, \Sigma_{Jh})$ , 有如下公式

$$\bar{C}_{Jh} = \frac{\sum_{l=1}^{L_{a_p}} \sum_{t=1}^{T_l} \gamma_t^{(l)}(J, h)}{\sum_{l=1}^{L_{a_p}} \sum_{t=1}^{T_l} \xi_t^{(l)}(J)} \quad (19)$$

$$\bar{M}_{Jh} = \frac{\sum_{l=1}^{L_{a_p}} \sum_{t=1}^{T_l} o_t^{(l)} \gamma_t^{(l)}(J, h)}{\sum_{l=1}^{L_{a_p}} \sum_{t=1}^{T_l} \gamma_t^{(l)}(J, h)} \quad (20)$$

$$\bar{\Sigma}_{Jh} = \frac{\sum_{l=1}^{L_{a_p}} \sum_{t=1}^{T_l} \gamma_t^{(l)}(J, h) (o_t^{(l)} - \bar{M}_{Jh}) (o_t^{(l)} - \bar{M}_{Jh})^T}{\sum_{l=1}^{L_{a_p}} \sum_{t=1}^{T_l} \gamma_t^{(l)}(J, h)} \quad (21)$$

其中,

$$\gamma_t^{(l)}(J, h) = \sum_{I=1}^{N_{r_p}-1} \alpha_{t-1}^{(l)}(I) a_{IJ} C_{Jh} b_{Jh} (o_t^{(l)}) \beta_t^{(l)}(J) / P(O^{(l)} | \lambda) \quad (22)$$

$\xi_t(J)$  表示从状态  $\theta_J$  转移出去的次数的期望值;  $1 \leq I \leq N_{r_p} - 1, 1 \leq J \leq N_{r_p} - 1; \Omega_p = \{w_p | w_p \text{ 是 } L_w \text{ 个观察序列中含字符 } r_p \text{ 的单词}\}, L_{a_p}$  是集合  $\Omega_p$  包含元素的个数.

得到  $\lambda^w = (\pi^w, A^w, B^w)$  后, 还要从  $\pi^w$  和  $A^w$  中还原出每个字符初始状态的概率  $\pi_{i_{r_p}}$  和转移至结束状态的概率  $a_{i_{r_p}e}$ , 从而得到级联训练后的字符 HMM 模型.

对级联模型中处于字符  $r_p$  的 Markov 链上的状态  $\theta_I, \theta_J$  有还原公式如下

$$\pi_{i_{r_p}} = \frac{\sum_{\Omega_p} \pi_I + \sum_{\Omega_{\bar{p}p}} a_{\cdot I}}{N_{r_p} - 1} \quad (23)$$

$$a_{ij} = \frac{\sum_{\Omega_p} a_{IJ}}{N_{r_p} - 1} \quad (24)$$

$$a_{i_{r_p}e} = \frac{\sum_{\Omega_{p\bar{p}}} a_{I\cdot}}{N_{r_p} - 1} \quad (25)$$

其中,  $\sum_{\Omega} a_{\cdot I}$  表示  $\Omega$  中起始状态任意, 终止状态为  $\theta_I$  的所有状态转移概率之和;  $\sum_{\Omega} a_{I\cdot}$  表示  $\Omega$  中起始状态为  $\theta_I$ , 终止状态任意的所有状态转移概率之和.

用式(19)~(21), 式(23)~(25)得到了  $r_p$  的经过级联训练的字符模型  $\lambda_p = (\pi_p, A_p, B_p)$ .

## 4.2 级联 HMM 训练算法描述

与 Baum-Welch 算法相同, 采用级联 Baum-Welch 算法训练存在的问题仍是算法容易收敛到一个局部极值点. 特别是当模型参数的训练空间比较大时, 参数的训练结果对初始化数据的依赖性很强, 如果设置的初始参数距最优点比较远, 训练结果很难达到最优. 由于采用的是混合高斯分布的连续 HMM 模型, 需要训练的模型参数相对比较多, 上述问题更加严重.

因此, 级联训练之前先采用 Baum-Welch 算法训练字符的 HMM 模型参数, 以此作为级联训练的初始参数, 使初始参数尽可能靠近全局最优值.

用  $L_w$  个手写单词观察序列, 对字符集  $R$  中每个  $r_p$  做有监督的级联训练.

级联 Baum-Welch 算法步骤描述如下:

1. 导入 Baum-Welch 算法训练的字符 HMM 模型.
2. 迭代结束控制参数初始化:  $Iteration = 0$ ; 迭代阈值  $DELTA$  初始化.
3. 开始新一轮迭代,  $l = 1$ .
4.
  - (1) 输入单词观察序列  $O^{(l)}, O^{(l)} \in \Omega$ ;
  - (2) 用更新的字符模型建立当前单词的级联模型  $\lambda^w$ ;
  - (3) 用前向算法或后向算法计算  $P(O^{(l)} | \lambda^w)$  的值;
  - (4) 根据式(15)、式(16)和式(22), 计算每个字符的期望  $\xi_t^{(l)}(I, J)$ ,  $\xi_t^{(l)}(I)$  和  $\gamma_t^{(l)}(J, h)$ ;
  - (5)  $l = l + 1$ ;
  - (6) 如果  $l > L_w$ , 转去执行步 5; 否则, 重复执行步(1)~步(5).
5. 用所得到的期望值, 根据式(17)~(21), 重新估计每个字符的  $\pi_I, \bar{a}_{IJ}, \bar{b}_J$ , 用式(23)~(25)更新每个字符的模型参数.
6. 计算  $delta = \left| \sum_{l=1}^{L_w} P(O^{(l)} | \lambda^w) - Iteration \right|$ .
7. 如果  $delta \geq DELTA$ , 则  $Iteration = \sum_{l=1}^{L_w} P(O^{(l)} | \lambda^w)$ , 重复执行步 3~6; 否则执行步 8.
8. 保存级联训练模型参数, 算法结束.

级联 Baum-Welch 训练得到的字符模型的集合, 称之为级联模型库. 对手写英文单词, 其模型库中包括 52 个英文字符的 HMM, 这 52 个字符的 HMM 模型可以组合出长度任意的字符串级联模型.

通过扩展状态转移的概念, 算法清晰地描述、记录了模型之间的状态转移情况, 将原本互相孤立的

字符模型紧密地结合, 算法解决了无法为每个不定长时间序列建模的问题.

5 级联 HMM 训练算法实验

选用基于字符 HMM 的联机手写英文单词识

别系统 HMM 有限状态模型网循环搜索法(Finite State Network,FSN)<sup>[13]</sup> 作为比较系统. FSN 在字符连接方式上与 HMM 级联模型方法(Cascaded HMM)的系统有区别,但二者在字符的连字笔段的处理上都有所考虑. 表 1 给出 FSN 与 Cascaded HMM 的比较.

表 1 测试系统比较

	字符 HMM 模型		其它模型	模型总数	单词模型			识别前处理	解码	切分识别策略
	状态数	结构			连字笔段	延迟笔段的处理	模型建立方式			
FSN	10 状态	从左至右无跨越拓扑结构	1 个 3 状态连字模型, 4 个 10 状态的不完整字符模型	57	建立连字笔段 HMM	4 个不完整字符 HMM	字符模型、连字模型和不完整字符模型的简单连接	根据每个落笔一提笔轨迹的尺寸和形状进行分割并标注字符主体、延迟笔段和连字笔段	词典驱动 Viterbi 解码	先分割再识别
Cascaded HMM	与观察序列长度相关	从左至右有跨越拓扑结构	—	52	—	—	字符模型的级联连接	—	词典驱动级联 Viterbi 解码	自由切分—动态寻优

使用的数据是实验室建立的非特定人书写联机手写单词样本库 EW-1,词表有 11275 词,包含了英文单词的基本词汇及实用词汇. 样本库容量 15300 个单词样本;测试样本约为 15000 个单词,级联字符 HMM 训练样本 300 个. 字符训练样本总量约为 1450 个. 每个字符训练样本约 30 个. GMM 的高斯数为 10、特征 8 维(分别是笔段的端点坐标、长度、笔段弧长五等分后得到的子笔段的倾角),即字符模型的每个状态由一个混合高斯模型 GMM 描述,一个 GMM 由 10 个高斯分量构成,每个分量是 8 维高斯分布.

表 2 给出了两种方法的性能比较. 对训练样本的识别,HMM 级联模型方法效果令人满意;HMM 模型网循环搜索法只有训练字符模型,无单词训练过程,测试时加入了延迟笔段和连字笔段的人工标注,因此效果较好.

表 2 两种方法的性能比较

前 $n$ 候选 识别率/%	训练样本		测试样本	
	FSN 人工标注	Cascaded HMM	FSN	Cascaded HMM
1	90.87	93.52	82.34	89.26
2	91.29	94.16	84.07	90.42
3	91.71	94.87	85.92	90.94
4	93.14	95.24	87.13	91.61
5	93.96	95.70	89.60	92.35
6	94.50	96.49	91.27	92.89
7	95.08	97.32	92.15	93.62
8	95.27	98.07	92.99	94.11
9	95.93	99.16	93.84	94.79
10	96.22	99.97	94.17	95.34

对测试样本的识别,HMM 级联模型方法前 10 候选识别率为 95.34%;HMM 模型网循环搜索法用测试样本测试,无人工标注,级联模型法可降低模型网循环搜索法识别错误率在 29%~40%之间,第一候选错误降低率为 39.18%.

模型网循环搜索法引入了连字模型,但其字符模型状态数目固定,不足以覆盖手写体的多变性;为解决延迟笔段设定的 4 个不完整字符 HMM 有一定效果,但不显著. 比较系统的原实验条件要求测试样本被规范或自由书写在固定区域内,在分割、标注后识别<sup>[13]</sup>,系统实用性较差.

基于字符 HMM 模型级联的单词识别方法中,每个字符模型状态数是不固定的;字符内的,尤其是字符间的状态跨越可以较准确地表达手写单词的变化性. 对联机手写体的延迟笔段,由于用来判断字符迹象的信息不连贯,使用一个字符模型来表示一个可能跨越了几个其它字符的字符,容易引起模型表述的混乱. 方法可以在一定程度上由模型各状态高斯混合模型来表征和处理.

6 结 论

本文在总结各种基于 HMM 方法的英文单词识别技术的基础上,提出了级联的 HMM 方法. 对方法的构建思想、定义和算法给出了全面的阐述,并将之用于联机手写英文单词识别中. 通过级联训练

得到的字符 HMM 准确地描述了手写单词中相邻字符在上下文环境中的形变,即在字符模型的基础上为表征笔迹在字符间的变化设定一种特殊连接,通过修正训练算法的重估公式,重估字符模型的连接参数。

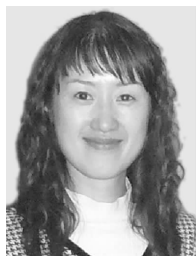
将字符模型的训练分两个阶段,首先使用单字符样本对字符 HMM 模型做初始训练,以防止算法收敛在局部极值点,然后用单词样本对相应字符 HMM 模型做级联训练,将相邻字符的形变参数分别记录在模型中,这是训练算法的核心部分,也是算法改进的关键。

已知  $N$  为字符模型状态数,  $N^w$  为单词级联模型状态数,  $T$  为样本序列的长度,级联 Baum-Welch 算法与 Baum-Welch 算法的时间复杂度在形式上是一致的,但前者的计算量比后者的计算量大,原因是单词级联模型状态数  $N^w$  及单词样本的时序长度均大于字符的模型状态数  $N$  及其时长. 识别时使用经过级联训练的字符模型,级联 Viterbi 解码算法的时间复杂度为  $O(N^2 T)$ . 测试结果表明,级联模型的描述能力与比较系统相比具有明显的优势,在实验中获得较好的识别性能,为连续识别技术提供了新的理论和方法支持。

## 参 考 文 献

- [1] Bengio Y. Markovian models for sequential data. *Neural Computing Surveys*, 1999, 2: 129-162
- [2] Choi H, Baraniuk R G. Multiscale image segmentation using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 2001, 10: 1309-1321
- [3] Guo Xiao-Wei, Tian Zheng, Lin Wei. Wavelet-domain hidden Markov overlapping tree model and artifact evaluating. *Computer Science*, 2005, 32(1): 208-211, 223(in Chinese)  
(郭小卫, 田铮, 林伟. 小波域隐 Markov 交叠树模型及块效应评价. *计算机科学*, 2005, 32(1): 208-211, 223)
- [4] Marukat S, Artières T, Dorizzi B, Gallinari P. Sentence recognition through hybrid Neuro-Markovian modeling// *Proceedings of the International Conference on Document Analysis and Recognition*. Seattle, Washington, US, 2001
- [5] Artières T, Gallinari P, Li H, Marukat S, Dorizzi B. From character to sentences: A hybrid Neuro-Markovian system for on-line handwriting recognition// Bunke H, Kandel A eds. *Hybrid Methods in Pattern Recognition*. Parution Courant; World Scientific Publ. Co., 2001: 1-27
- [6] Xie Jin-Hui. Linear predictive HMM in speech recognition, compression and enhancement[Ph. D. dissertation]. Wuhan: Huazhong University of Science and Technology, 1990(in Chinese)  
(谢锦辉. 线性预测 HMM 在语音识别、压缩和增强中的应用[博士学位论文]. 武汉: 华中理工大学, 1990)
- [7] Kenny P, Lennig M, Mermelstein P. A linear predictive HMM for vector-valued observations with applications to speech recognition. *IEEE Transactions on Acoustics, Speech, Signal Processing*, 1990, 38(2): 220-225
- [8] Chen M Y, Kundu A, Srihari S N. Variable duration hidden Markov model and morphological segmentation for handwritten word recognition. *IEEE Transactions on Image Processing*, 1995, 4(12): 1675-1687
- [9] Deng L, Kenny P, Lennig M. Modeling acoustic transitions in speech by state-interpolation hidden Markov models. *IEEE Transactions on Signal Processing*, 1992, 40(2): 265-271
- [10] Marti U-V, Bunke H. Handwritten sentence recognition// *Proceedings of the 15th International Conference on Pattern Recognition*. Barcelona, Spain, 2000, 3: 463-468
- [11] Artières T, Gallinari P. Stroke level HMMs for on-line handwriting recognition// *Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition*, 2002: 227-232
- [12] Bunke H, Roth M, Schukat-Talamazin E G. Off-line cursive handwriting recognition using hidden Markov models. *Pattern Recognition*, 1995, 28(9): 1399-1413
- [13] Sin B-K, Ha J Y, Oh S C, Kim J H. Network-based approach to online cursive script recognition. *IEEE Transactions on System, Man, and Cybernetics—Part B: Cybernetics*, 1999, 29(2): 321-328
- [14] Mohamed Magdi A, Gader Paul. Generalized hidden Markov models: Part I, Theoretical frameworks; Part II, Application to handwritten word recognition. *IEEE Transactions on Fuzzy System*, 2000, 8(1): 67-94
- [15] Li Xiao-Lin, Parizeau Marc, Palmonon Réjean. Training hidden Markov models with multiple observations—A combinatorial method. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2000, 22(4): 371-377
- [16] Jaeger S, Manke S, Waibel A. Npen++: An on-line handwriting recognition system// *Proceedings of the 7th International Workshop on Frontiers in Handwriting Recognition*. Amsterdam, Nijmegen: International Unipen Foundation, 2000: 249-260
- [17] Jung K, Yoon S, Kim H. Continuous HMM applied to quantization of on-line Korean character spaces. *Pattern Recognition Letters*, 2000, 21(3): 303-310
- [18] Liu Jia-Feng, Huang Jian-Hua, Tang Xiang-Long. A HMM based on-line Chinese character recognition system and improved training algorithm. *Journal of Chinese Information Processing*, 2001, 15(4): 47-52(in Chinese)  
(刘家锋, 黄剑华, 唐降龙. 基于 HMM 的联机汉字识别系统及其改进的训练方法. *中文信息学报*, 2001, 15(4): 47-52)





**ZHAO Wei**, born in 1972, Ph. D. , lecturer. Her current research interests include pattern recognition and Chinese information processing.

**LIU Jia-Feng**, born in 1968, Ph. D. , associate professor. His research interests include artificial intelligence, im-

age processing, pattern recognition etc.

**TANG Xiang-Long**, born in 1960, Ph. D. , professor, Ph. D. supervisor. His research interests include pattern recognition, information processing, and motion intelligence.

**WU Rui**, born in 1976, M. S. , lecturer. His current research interests include character recognition and Image processing.

## Background

Since the 1990s, interests of Hidden Markov Model (HMM) for handwritten character recognition have been growing significantly. As an effective combination of Markov chain and stochastic process, HMM could depict random signal using state transition and statistic mapping relation between state and observation. While in continue character recognition, it is difficult to establish an independent statistical model for each time sequence, or we say handwritten word, which sampled from tablet, because of the variability and illegibility of free handwritten text. How to represent a word model by using as less character HMM as possible is one of the hotspot in this field. Researchers have been making sustained attempt at novel approaches.

In this paper, a continuous stochastic model, named as cascaded HMM, is proposed. It could generalize an integrated word model for certain a free handwritten word by using single character HMMs. The key of the method is the proba-

bility of shifting between character models and the cascaded Baum-Welch Algorithm that are explained elaborately in the paper. By using Baum-Welch and cascaded Baum-Welch training, only 52 character HMMs are obtained. These HMMs will be connected as word level cascaded HMM and used in decoding free handwritten words. Without label any of the word, the experiment results show effectiveness of the cascaded algorithm and higher performance in online English word recognition.

This group has been working in off-line or on-line character recognition research for many years. Their interests have been in both pen-based device design and recognition strategy and algorithm design. Till now, they had gained many research fruit and reward in this area. By the year of 2006, they had published more than 100 papers in the international proceedings and journals on this topic.