

一种基于 SCHMM 的手语识别方法

张岱¹, 柯珂²

(1. 中国网通(集团)有限公司北京市分公司, 北京 100075;

2. 华北科技学院, 北京 101601)

摘要:手语识别的研究具有重大的学术价值和广泛的应用前景。在近些年的手语识别工作中, 隐马尔科夫模型(Hidden Markov Models, 简称 HMMs)起到了重要的作用。基于 HMM 的统计框架是当前动态识别领域的主流方法, 同时也是该文的研究工作的理论基础。提出将半连续隐马尔科夫模型(SCHMM)用于手语识别, 在理论上证明了 SCHMM 优于离散隐马尔科夫模型(DHMM)和连续隐马尔科夫模型(CHMM), 可以避免 DHMM 中因矢量量化造成的信息损失, 在保证识别率的前提下降低模型的复杂性和运算量。

关键词:手语识别; HMM; Semi-Continuous Hidden Markov Model; 观察值

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2009)07-0149-03

SCHMM for Sign Language Recognition

ZHANG Dai¹, KE Ke²

(1. Beijing Branch, China Netcom (Group) Company Limited, Beijing 100075, China;

2. North China University of Science and Technology, Beijing 101601, China)

Abstract: The research of sign language recognition has great academic value and broad application prospect. In recent works on sign language recognition, Hidden Markov Models (HMMs) has played an important role. The statistical frame based on the HMM is the mainstream method in dynamic recognition domain recently; also is this article's basic theory. Presents a Semi-Continuous Hidden Markov Model for sign language recognition, shows that SCHMM is prior to the DHMM and the CHMM in theory. SCHMM avoid the information loss because of the vector estimate in DHMM; debase the complexity and the operations at the same recognition rate.

Key words: sign language recognition; HMM; semi-continuous hidden Markov model; observe value

0 引言

手语是聋人之间交流的主要方式, 手势是人们之间交流的重要辅助手段, 因此手语和手势识别的研究具有重要的实际意义; 同时, 手语识别的研究还有助于促进和谐人机交互的发展, 开发出自然和谐的人机交互方式, 进而完全取代传统的基于鼠标和键盘的人机交互模式, 因此手语识别的研究具有重大的理论意义; 除此之外手语识别的研究也有助于机器人的示范学习和虚拟现实等研究的发展和进步^[1]。

手语是一种规范化手势, 手语识别将手语信号转换为文字或语音信息, 以方便手语的传播和理解^[2]。最早的手语自动识别的尝试开始于 20 世纪 80 年代

末。根据手语输入设备不同, 分为采用视觉和数据手套两种方法^[3]。基于视觉的手语识别如 Charaphayan 和 Marble 研究了一种用图像处理的方法来识别 31 个美国手语词, 可以正确识别出其中的 27 个。Starner 等对 40 个词构成的具有很强语法约束的短小语句进行识别, 识别率为 99.2%^[4,5]。基于数据手套的手语识别有: Takahashi 和 Kishino 用一只 VPL 数据手套识别 46 个日本手指字母, 可正确识别出其中的 30 个。Liang 等对连续的我国台湾手语识别进行了研究, 他们设定时变得手形参数变化的值来确定手势流中手语词的开始和结束, 在 250 个手语词组成的词汇集上, 他们取得了 80.4% 的识别率^[6]。

近年来, 基于统计隐马尔科夫模型(HMM)的方法一直是手语识别研究的主流方法^[7]。目前的手语识别系统也通常采用 HMM, 包括离散隐马尔科夫模型(DHMM)和连续隐马尔科夫模型(CHMM)作为训练模型。SCHMM 是作为 DHMM 和 CHMM 的折中, 本质上讲是二者的一般形式, 对于大词汇量手语识别系

收稿日期: 2008-10-16; 修回日期: 2009-01-02

基金项目: 国家自然科学基金资助项目(60533030); 北京市自然科学基金资助项目(4061001)

作者简介: 张岱(1982-), 男, 北京人, 助理工程师, 研究方向为模式识别。

统,可以期待采用 SCHMM,既可以取得与基于 CHMM 相当的高识别率,还极大地降低了模型复杂性与运算量,保证系统的实时性的识别效果。

1 HMM 模型

1.1 HMM 模型简介

隐马尔科夫模型(Hidden Markov Model, HMM)是一种具有学习能力的统计模型。HMM 的基本理论在 20 世纪 60 年代末提出并加以研究^[8]。HMM 是在 Markov 链的基础之上发展起来的。由于实际问题比 Markov 链模型所描述的更为复杂,观察到的事件并不是与状态一一对应,而是通过一组概率分布相联系,这样的模型就成为 HMM。它是一个双重随机过程,其中之一是 Markov 链,这是基本随机过程,它描述状态的转移。另一个随机过程描述状态和观察值之间的统计对应关系。这样,站在观察者的角度,只能看到观察值,不像 Markov 链模型中的观察值和状态一一对应,因此,不能直接看到状态,而是通过一个随机过程去感知状态的存在及其特性,因而称之为“隐”Markov 模型,即 HMM^[9]。按状态的观察概率密度的连续性, HMM 可分为连续隐马尔科夫模型(CHMM, Continuous Hidden Markov Model)、半连续隐马尔科夫模型(SCHMM, Semi-Continuous Hidden Markov Model)和离散隐马尔科夫模型(DHMM, Discrete Hidden Markov Model)。

更形象地说, HMM 可以分为两个部分,一个是 Markov 链,由 π, A 描述,产生的输出为状态序列,另一个是一个随机过程,由 B 描述,产生的输出为观察值序列,如图 1 所示, T 为观察值时间长度。

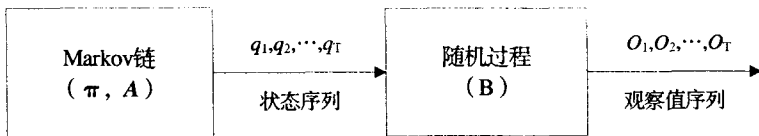


图 1 · HMM 组成示意图

可以记一个 HMM 为 $\lambda = (N, M, \pi, A, B)$, 或简写为 $\lambda = (\pi, A, B)$, 式中 N 为模型中 Markov 链状态的数目, M 为每个状态对应的可能的观察值数目^[10]。 $\pi = \{\pi_i\}$ 表示初始状态概率集合 $\pi_i = P[q_1 = \theta_i]$, $1 \leq i \leq N$ 并满足约束条件 $0 \leq \pi_i \leq 1, \sum_i \pi_i = 1$ 。 $A = \{a_{ij}\}$ 为状态转移概率矩阵,

$$a_{ij} = P(q_{t+1} = \theta_j / q_t = \theta_i) \quad 1 \leq i, j \leq N \quad (1)$$

并满足约束条件

$$a_{ij} \geq 0, 1 \leq i, j \leq N, 1 \leq i \leq N \quad (2)$$

$B = \{b_j(k)\}$ 为观测信号概率密度, $b_j(k)$ 为 j 状

态下观察符号 k 的概率密度。当观察符号 k 是离散符号时, $b_j(k)$ 是离散的概率值, 当观察符号 k 取为连续的矢量时, $b_j(k)$ 是连续概率密度函数, 通常取为 Gaussian 混合密度函数。

HMM 假定在同一个状态之内的观察值都是独立、同分布的, 因此, HMM 能够很好地描述具有若干稳定区域的高维时序信号。在使用 HMM 进行手语识别时, 将每个手语词的训练数据训练出一个 HMM, 在对一个未知的手语数据进行测试时, 使得观察值的后验概率最大的那个 HMM 对应的手语词即为识别结果。

1.2 SCHMM 模型及其优势描述

半连续隐马尔可夫模型(SCHMM, Semi-Continuous Hidden Markov Model)是在对离散 HMM 和连续 HMM 思想进行综合后提出来的。

离散 HMM 处理的观察值必须在 M 个离散值组成的集合之内。因此, 对手语信号段分帧产生的特征参数序列, 必须经过矢量量化(VQ)之后, 才能用于离散 HMM 的训练和识别, 此时, VQ 的码本(code book)中 M 个码字(code word), 就是 HMM 相应的 M 个观察值集合。这样的处理过程就产生了离散 HMM 的两个缺陷: 首先, 将手语数据, 确切地说, 是各帧的特征参数, 经过 VQ, 造成了一些信息丢失。此外, VQ 的码本训练和离散 HMM 训练是两个分开的最优化过程, 而不是一起进行优化训练。另一方面, 连续 HMM 虽然能直接处理手语数据, 但为了取得手语识别的良好性能, 必须要较多的概率函数进行混合, 以得到某个状态对应的观察值概率函数, 对 $b_j(X) = \sum_{v_j \in \eta(X)} f(X/v_j) b_i(j)$ 高斯型概率函数来说, K 值就要比较大。这样, 不仅模型复杂, 运算量大, 而且, 需要更多的训练数据才能得到可靠的模型参数。

实际上半连续 HMM 也是为了克服离散 HMM 和连续 HMM 的不足之处而提出来的。SCHMM 的基本思想为: 离散

HMM 使用 VQ 产生 M 个码字组成的码本, 这 M 个码字实际上就是将训练矢量空间划分为 M 个部分, 这种划分是丢失手语特征信息的原因, 因此, 用 M 个高斯型概率函数取代 M 个码字, 这样, 训练矢量空间的划分就不会损失信息, 而且, VQ 的训练和 HMM 的训练可在一个优化过程中完成。此时, 状态对应的观察值概率函数为

$$b_j(X) = \sum_{j=1}^M f(X/v_j) b_i(j) \quad (3)$$

其中, $f(X/v_j)$ 就是第 j 个码字 v_j 对应的高斯概率函数, $b_i(j)$ 就是 $P(v_j | q_i = \theta_i)$ 。实际上, 由于码本中码

字较多,即 M 值较大,因此为了减少计算量,(3)式一般简化为

$$b_j(X) = \sum_{v_j \in \eta(X)} f(X/v_j) b_i(j) \quad (4)$$

其中, $\eta(X)$ 表示那些使 $f(X/v_j)$ 足够大的码字的集合,这样 $\eta(X)$ 就比 M 小了很多。

从半连续 HMM 的观察值概密表达式(4)式可知,它事实上是离散 HMM 和连续 HMM 的一种一般形式。当 $\eta(X)$ 只选一个使 $f(X/v_j)$ 最大的码字,即与 X 最接近的码字,此时,半连续 HMM 就变为离散 HMM。另一方面,半连续 HMM 也是连续 HMM,只不过 N 个状态的概密函数共同拥有 M 个概密函数,而且 $b_i(j)$ 就是混合若干概密函数时的加权系数。

剩下的关键问题就是在一次优化过程中训练 VQ 和 HMM,这可由与 Baum - Welch 算法类似的重估公式思想来解决,设训练序列为 $O = O_1 O_2 \cdots O_T$,可以推导出:

$$\bar{b}_i(j) = \frac{\sum_{t=1}^T \gamma_t(i, j)}{\sum_{i=1}^M \xi_t(i)}, 1 \leq j \leq M \quad (5)$$

对高斯型概密函数 $f(X/v_j)$ 的均值矢量 $\bar{\mu}_j$ 和方差矩阵 $\bar{\Sigma}_j$,有

$$\bar{\mu}_j = \frac{\sum_{t=1}^T \gamma_t(j) O_t}{\sum_{t=1}^T \gamma_t(j)}, 1 \leq j \leq M \quad (6)$$

$$\bar{\Sigma}_j = \frac{\sum_{t=1}^T \gamma_t(j) (O_t - \bar{\mu}_j)(O_t - \bar{\mu}_j)^T}{\sum_{t=1}^T \gamma_t(j)} \quad (7)$$

$$1 \leq j \leq M$$

其中

$$\gamma_t(j, k) = \begin{cases} \pi_i b_i(j) f(O_1/v_j) \beta_1(i)/P(P|\lambda), & t = 1 \\ \sum_i \chi_{t-1}(k, i, j), & 1 < t \leq T \end{cases} \quad (8)$$

而

$$\chi_{t-1}(k, i, j) = a_i(i) a_{ij} b_j(k) f(O_{t+1}/V_k) \beta_{t+1}(j)/P(O|\lambda) \quad (9)$$

在手语识别过程中,半连续 HMM 比离散和连续

HMM 有更好的性能,应用半连续的 HMM 可以在保证识别速度的前提下大大降低计算量。它是经典离散 HMM 的一种较为成功的修正形式。

2 结束语

手语识别的正确率和识别速度对于基于 HMM 模型的手语识别系统是相当重要的,笔者提出以 SCHMM 作为手语识别模型,它是离散隐马尔科夫模型和连续隐马尔科夫模型的一种折中方案,克服了 DHMM 识别精度低的缺点,可以取得与基于 CHMM 相当高的识别率,又降低了 CHMM 计算的复杂性,从而保证手语识别的识别效果。在以后的工作中,将进一步研究 SCHMM 在手语识别中的应用。

参考文献:

- [1] 范保玲,王 民,董颖娣. 基于肤色监测技术的手势分割[J]. 计算机技术与发展,2008,18(3):105-106.
- [2] 王 红,刘建辉. 人工智能在决策支持系统中的应用与研究[J]. 微计算机信息,2005,21(3):177-178.
- [3] 吴江琴,高 文,陈熙霖,等. 基于 ANN/HMM 的中国手语识别系统[J]. 计算机工程与应用,1999,35:1-4.
- [4] Marten A, Nico K, Harri C. Assumptions for long-term stochastic population forecasts in 18 European countries[J]. European Journal of Population,2007,23:33-69.
- [5] Baum L E, Petrie. Statistical inference for probabilistic function of finite state Markov chains[J]. Annals of Math. Statistic, 1996,6:1554-1563.
- [6] 仁海宾,祝远新,徐光祐,等. 连续动态手势的时空表观建模及识别[J]. 计算机学报,2000,23(8):824-828.
- [7] 宋益波,高 文,尹宝才,等. 文本驱动的聋哑人手语合成系统[J]. 计算机学报,1999,22(7):733-739.
- [8] Baum L E, Geon J A. An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology [J]. Bull. Amer. Meteorology, 1967(2):360-363.
- [9] 赵 巍,刘家峰,唐降龙,等. 连续字符识别的级联 HMM 训练算法[J]. 计算机学报,2007,30(12):2144-2145.
- [10] Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proc. IEEE, 1989(2): 257-286.

中国计算机学会会刊、中国科技核心期刊
《计算机技术与发展》欢迎订阅,邮发代号:52-127