

The Application of Hidden Markov Models in Offline Handwriting Digital Recognition

Xiangxiang Zhang, Le Zhou, Yongling Qian, Xingling Zhang, Chenguang Cui, Xuerui Wang

School of Mathematical Science and Engineering, Anhui University of Technology, Ma'anshan Anhui
Email: 814081807@qq.com

Received: Apr. 30th, 2018; accepted: May 16th, 2018; published: May 23rd, 2018

Abstract

In this paper, we use hidden Markov models (HMM) to study the off-line handwritten numeral recognition. This paper can generally be divided into three parts: the first part introduces the basic theory of HMM; the second part introduces the digital image preprocessing and feature extraction, and this part belongs to image processing, and it is very important in the handwriting recognition, and especially extracting a stable and effective feature determines the successful recognition; the third part implements recognition process, and we use Matlab to obtain an off-line handwriting recognition system.

Keywords

Hidden Markov Models, Handwritten Numeral Recognition, Preprocessing, Feature Extraction

隐马尔可夫模型在脱机手写数字识别中的应用

张祥祥, 周 乐, 钱永林, 张星玲, 崔晨光, 王雪蕊

安徽工业大学数理科学与工程学院, 安徽 马鞍山
Email: 814081807@qq.com

收稿日期: 2018年4月30日; 录用日期: 2018年5月16日; 发布日期: 2018年5月23日

摘 要

本文主要用隐Markov模型(HMM)来研究脱机手写数字识别。本文的内容分为三个部分, 第一个部分是介绍HMM的基本理论; 第二部分是介绍数字图片的预处理和特征提取, 这部分属于图像处理方向, 并且在手写体数字识别中很重要, 特别是提取一个稳定而有效的特征决定着识别是否成功; 第三部分是具体地实现这一识别过程, 并且用Matlab实现了一个脱机手写数字识别系统。

关键词

Markov模型, 手写体数字识别, 预处理, 特征提取

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

隐 Markov 模型(Hidden Markov Models, 简称为 HMM)是一种统计模型, 今天正在手写体识别和语音识别等各个领域中获得广泛的应用。而有关它的理论基础, 却是在 1970 年前后由 Baum 等人建立起来的, 随后由 CMU(卡内基·梅隆大学)的 Jim and Janet Baker (贝克夫妇, 李开复的师兄师姐)和 IBM 的 Fred Jelinek 分别独立地提出用 HMM 来识别语音, 语音识别的错误率相比人工智能和模式匹配等方法降低了三倍(从 30%到 10%)。八十年代李开复博士坚持采用 HMM 的框架, 成功地开发了世界上第一个大词汇量连续语音识别系统 Sphinx。由于 Bell 实验室 Rabiner 等人在 80 年代中期对 HMM 的深入浅出的介绍, 才逐渐使 HMM 为世界各国从事语音处理的研究人员所了解和熟悉, 进而成为公认的一个研究热点。

HMM 在语音识别中的成功应用[1], 使得人们将 HMM 引入到手写识别中来[2] [3] [4] [5]。目前, 利用 HMM 进行脱机手写字符识别是脱机手写字符研究的一个热点的方向。此外, HMM 在人脸识别和图像分割等模式识别领域有成功的应用。

2. 隐 Markov 模型的基本理论

2.1. Markov 链

马尔可夫链, 因安德烈·马尔可得名, 是指数学中具有马尔可夫性质的离散事件随机过程。该过程中, 在给定当前知识或信息的情况下, 过去(即当前以前的历史状态)对于预测将来(即当前以后的未来状态)是无关的。

这一类随机过程, 它具有所谓的“无后效性(Markov 性)”, 即要确定过程的将来的状态, 知道它此刻的情况就足够了, 并不需要对它以往状况的认识, 这类过程称为 Markov 过程。而 Markov 链是 Markov 过程的特殊情况, 即 Markov 链是状态和时间参数都离散的 Markov 过程。从数学上, 可以给出如下定义:

随机过程 $\{X_n, n=1, 2, \dots\}$, 在任一时刻 n , 它可以处在状态 S_1, S_2, \dots, S_N , 且它在 $n+1$ 时刻所处的状态为 q_{n+1} 的概率, 只与它在 n 时刻的状态 q_n 有关, 而与 n 时刻以前它所处状态无关, 即有:

$$\begin{aligned} P(X_{n+1} = q_{n+1} | X_n = q_n, X_{n-1} = q_{n-1}, \dots, X_1 = q_1) \\ = P(X_{n+1} = q_{n+1} | X_n = q_n) \end{aligned}$$

其中, $q_1, q_2, \dots, q_n, q_{n+1} \in (S_1, S_2, \dots, S_N)$ 。

则称 $\{X_n, n=1, 2, \dots\}$ 为 Markov 链, 并且称条件概率 $P(X_{n+1} = q_{n+1} | X_n = q_n)$ 为一步转移概率, 当这个转移概率只与状态 q_n, q_{n+1} 的具体取值有关, 而与时刻 n 无关, 称这个 Markov 链为齐次 Markov 链, 在本文中, 讨论的 Markov 链都是齐次的。

当 q_n, q_{n+1} 分别取值 S_i, S_j 时, 记 $a_{ij} = P(X_{n+1} = S_j | X_n = S_i)$, 则 $a_{ij}, 1 \leq i, j \leq N$ 可以构成一个转移概率矩阵, 即

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

对于上面的矩阵显然有 $0 \leq a_{ij} \leq 1$, $\sum_{j=1}^N a_{ij} = 1$ 。

描述 Markov 链的最重要参数就是转移概率矩阵 A , 但 A 还决定不了初始分步, 即由 A 求不出 $q_1 = S_i$ 概率, 这样, 完全描述 Markov 链, 除了 A 矩阵之外, 还必须引进初始概率向量 $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, 其中

$$\pi_i = P(q_1 = S_i), 1 \leq i \leq N$$

显然有 $0 \leq \pi_i \leq 1$, $\sum_{i=1}^N \pi_i = 1$ 。

2.2. 隐 Markov 模型的定义

隐马尔可夫模型是关于时序的概率模型, 描述由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列, 再由各个状态生成一个观测而产生观测随机序列的过程。隐藏的马尔可夫链随机生成的状态的序列, 称为状态序列(state sequence); 每个状态生成一个观测, 而由此产生的观测的随机序列, 称为观测序列(observation sequence)。序列的每一个位置又可以看作是一个时刻。

隐 Markov 模型是在 Markov 链的基础上发展起来的。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, 这样的模型就称为隐 Markov 模型, 它是一个双重随机过程, 其中之一是齐次 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测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感知状态的存在及其特性。因而称之为“隐 Markov 模型”, 即 HMM。

2.3. 维特比算法与隐 Markov 模型

维特比算法(Viterbi)针对 HMM 中的解码或者预测问题, 寻找最可能的隐藏状态序列: 对于一个特殊的隐马尔可夫模型及一个相应的观察序列, 找到生成此序列最可能的隐藏状态序列。

在利用 Viterbi 算法进行识别的过程中, 其每步所得到的概率的具体值是没有意义的, 而且计算出的概率值总是很小很小的, 另外我们所需要的只是各模型对该观察序列的概率值之间的相对大小, 因此, 在计算过程中, 将每步得到的概率值取对数, 简化计算, 即:

定义

$$\delta_t(i) = \max_{i_1, i_2, \dots, i_{t-1}} \log P(q_1 = S_{i_1}, q_2 = S_{i_2}, \dots, q_{t-1} = S_{i_{t-1}}, q_t = S_i, O_1 O_2 \cdots O_t | \lambda)$$

那么, 初始化式变为

$$\delta_t(i) = \log \pi_i + \log b_i(O_t), \quad 1 \leq i \leq N$$

递归运算式变为

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) + \log a_{ij}] + \log [b_j(O_t)]$$

终止式变为

$$\log P^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

3. 数字图片的预处理与特征提取

3.1. 数字图片的预处理

所有的模式识别包含三个步骤：预处理、特征提取、分类识别。其中第一步预处理的好坏将直接影响到后续步骤。一般情况下，预处理有如下过程：二值化、平滑去噪、字符切割、归一化、倾斜校正、细化。

1) 二值化：对于字符识别，图片的彩色信息作用不大，相反还会增加算法复杂度，因此必须首先把图片转化成灰度图。本文采用的是 Otsu 提出的最大类间方差方法进行二值化，该方法的优点就是计算简单，稳定有效。

2) 平滑去噪：噪声去除已经成为图像处理及其重要的步骤，然而现在还没有一个通用的滤波去噪方法对所有的图片适用，一般的图像预处理必须根据实际情况选择不同的滤波去噪方法进行比较，最后得出最佳的滤波去噪方法[6]。

3) 字符切割：通常获取的图片除字符外，有较大的空白区，这对我们的研究没有任何价值，而且造成存储和时间资源浪费。为此，需要通过字符提取，去除空白区，提取有用的字符区。

4) 归一化：归一化的目的是将字符图片转化成一个标准尺寸，以便获取固定维度的特征向量，同时减少同类字符的变化对识别的影响，改善分类精度。

5) 倾斜校正：倾斜校正首先要确定倾斜角，然后再进行旋转。确定倾斜角有 Hough 变换法、Radon 变换法，矩方法等。本文采用的是 Radon 变换法。Radon 变换法是基于投影的矫正方法，利用图像水平方向和垂直方向的投影特征来进行倾斜判断。主要用到投影的方差和均方差、投影特征矢量、梯度方向场等统计特性。

6) 细化：不同的书写工具会造成笔画宽度不一样，所以这必将影响有些的特征提取。字符细化是通过一定的处理算法将字符重要的像素点保留下来，去除无关紧要的点，得到字符笔化骨架的技术[7]。

3.2. 特征提取

对数字识别特征提取有多种方法，本论文采用的方法是先提取数字的轮廓特征，然后对图像进行裁剪之后按照像素占总像素的比率提取出特征值。把每个数字图形定义成一个 $N * N$ ($5 * 5$) 的模板，将每个样品的长度和宽度 N 等分，平均有 $N * N$ 个等分，对每一等分进行像素个数统计，除以每一份的面积总数即得特征值。具体步骤如下：

1) 调入样本图片，找出图片中数字的上、下、左、右边界。对图像进行裁剪处理。

2) 将数字区域平均分成 $5 * 5$ 的小区域。

3) 计算 $5 * 5$ 的每个小区域中黑像素所占比例，第一行的 5 个比例值保存到特征的前 5 个，第二行对应着特征的 6~10，如此保存样本的特征值。本论文提取每个数字 25 个样本的特征值之后，把它们保存在结构体 Mytemplet 里，构成了特征库。

下面以 0 为例看一下求取特征值的方法，首先把图像按边缘裁剪放缩后，然后分成 $5 * 5$ 的小区域(见图 1)，然后计算出每个小区域中黑像素所占的比例，存储之后即为 0 的一个样本特征值(见表 1)。

4. 识别过程

4.1. HMM 的脱机手写数字识别的一般过程

本文主要是应用 HMM 来对 0~9 这 10 个阿拉伯数字识别，首先我们要利用一组已知的手写样本来分

别训练数字 0~9 的 HMM，即 $\lambda_0, \lambda_1, \dots, \lambda_9$ 。然后对于一个待识别的数字样本，先要对其预处理，然后提取一组特征 $O_1 O_2 \dots O_T$ ，把它作为 HMM 的观察值序列，再用 Viterbi 算法分别计算它与求出来的 10 个 HMM 的匹配率，匹配率最高对应的数字即为该样本识别结果。

4.2. 脱机手写数字识别系统设计

我们将 HMM 的相关算法如 Viterbi 算法等用 Matlab 语言写出来，参考了 Kevin Murphy 在网站[8]上的 HMM 工具箱。在 Matlab R2016a 中设计这一脱机手写数字识别系统，最后用 Matlab 设计出了图形用户界面 GUI(见图 2)。

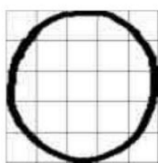


Figure 1. Number 0
图 1. 数字 0

Table 1. Eigenvalues of the number 0
表 1. 数字 0 的特征值

0.1322	0.2356	0.1689	0.2232	0.1522
0.3333	0	0	0	0.2900
0.2700	0	0	0	0.2689
0.2967	0	0	0	0.2911
0.0989	0.2356	0.1744	0.2300	0.1133



Figure 2. GUI
图 2. 图形用户界面

Table 2. System resulting data of standard experiment
表 2. 标准试验系统结果数据

导入的手写数字				识别得到的手写数字						正确率
0	0	0	0	0	6	0	6	0	0	80%
1	1	1	1	1	1	1	1	1	1	100%
2	2	8	2	2	2	8	2	8	2	60%
3	3	3	8	3	3	3	8	3	3	80%
4	6	4	4	4	4	4	4	4	1	80%
5	5	5	5	8	5	5	8	5	5	80%
6	6	6	6	6	6	6	6	6	6	100%
7	7	7	7	7	7	7	7	7	7	100%
8	8	8	8	8	6	8	8	8	8	90%
9	9	9	9	1	9	9	9	9	9	90%

系统具体功能：每导入一张外部图片，点击开始识别按钮即可识别图片中的手写数字，系统会输出识别结果。

4.3. 识别系统试验结果

多次试验结果表明，该识别系统识别率较高(见表 2)。

5. 结束语

本文基于 HMM 理论设计了一个脱机手写数字识别系统。设计过程中采用二值化、平滑去噪、字符切割、归一化、倾斜校正、细化等进行图片预处理。对数字识别特征提取有多种方法，本论文采用的方法是先提取数字的轮廓特征，然后对图像进行裁剪之后按照像素占总像素的比率提取出特征值，最后用 Viterbi 算法计算匹配率实现数字识别。经过多次试验表明，对于书写较为工整的孤立数字，系统识别率较高。本文的结论初步说明了，类似于 HMM 在语音识别中的成功运用，把 HMM 应用到脱机手写体识别中是可以的。

基金项目

安徽省教育厅重点项目(KJ2017A851)。

参考文献

- [1] 谢锦辉. 隐 Markov 模型(HMM)及其在语音处理中的应用[M]. 武汉: 华中理工大学出版社, 1995.
- [2] Mohamed, M.A. and Gader, P. (2000) Generalized Hidden Markov Models—Part II: Application to Handwritten Word Recognition. *IEEE Transactions on Fuzzy Systems*, **8**, No. 1.
- [3] 梁佳玉. 基于 HMM 的脱机自由手写英文单词识别系统[D]: [硕士学位论文]. 北京: 中科院自动化研究所, 2004.
- [4] 梁佳玉, 刘昌平, 黄磊. 脱机自由手写英文单词的识别[J]. 计算机应用, 2004, 24(9): 41-43.
- [5] 李辉熠, 李峰, 黄道昌. 基于 MNMM 的脱机手写体字符识别[J]. 长沙理工大学学报: 自然科学版, 2007, 4(2): 63-67.
- [6] 葛照君, 葛世龙, 盛磊. 基于 HMM 的脱机手写字符识别方法研究[J]. 江苏科技大学学报: 自然科学版, 2008, 22(6): 57-61.

- [7] 陈超等. 图像处理与 GUI 设计篇[M]. 北京: 电子工业出版社, 2011.
- [8] Murphy, K. (2005) Hidden Markov Model (HMM) Toolbox. <http://www.matlabsky.com/thread-2109-1-1.html>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: csa@hanspub.org