

隐马尔可夫模型 HMM

编写者：林海云 吴凤

1. 引言

隐马尔可夫模型(HMM)的理论基础是1970年前后由Baum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中。HMM是一个二重马尔可夫随机过程,包括具有转移概率的马尔可夫链和输出观测值的随机过程,其状态是不确定或不可观测的,只有通过贯彻序列的随机过程才能表现出来。1994年, Samaria F[5]将HMM应用到人脸识别中取得了一定的识别率。Nefian A V等人[6]将所提取出的2D-DCT特征向量引入HMM中进行人脸识别。随后,有不少的学者对HMM进行了改进,并将改进后的HMM应用到人脸识别中也取得了较好的实验效果。

2. HMM 简介

HMM是一个双内嵌随机过程,即HMM是由两个随机过程组成的,一个是隐含的转台转移序列,它对应一个单纯的Markov过程;另一个与隐状态有关的测试序列,,并且在这两个随机的过程中,有一个随机过程(转台转移序列)是不可观测的,只能通过另一个随机过程的输出观测序列进行推断,所以称为隐马尔可夫模型,基本要素包括:

- (1) 模型的状态数 N 。如果 S 是状态集合,则 $S=\{S_1, S_2, \dots, S_n\}$ 。模型在时间 t 的状态极为, q_t (q_t 属于集合 S , $1 \leq t \leq T$), T 是一个观察序列的长度。模型经历的状态序列记为 $\{q_1, q_2, \dots, q_t\}$;
- (2) 观察符号数 M , 设 V 是所有观察符号的集合,则 $V=\{v_1, v_2, \dots, v_n\}$;
- (3) 状态转移的概率分布 A 。状态转移的概率分布可表示为 $A\{a_{ij}\}$, 其中 $a_{ij}=P\{q_{i+1}=S_j | q_i=S_i\}$, $i \leq I, j \leq N$, 且满足 $a_{ij} \geq 0, \sum_{j=1}^N a_{ij}=1$ 表示时刻 t 从状态 S_i 转移到时刻 $t+1$ 状态 S_j 的转移概率。
- (4) 状态 S_i 条件下输出的观测变量概率分布 B 。假设观测变量的样本空间为 V , 在状态 S_i 时输出观测变量的概率分布可表示为: $B\{b_i(v), i \leq i \leq N, v \text{ 属于 } V\}$, 其中 $b_i(v)=f\{Q_t=v | q_t=S_i\}$, Q_t 为时刻 t 的观测随机变量, 可以是一个数值或向量, 观测序列记为 $O=\{O_1, O_2 \dots O_T\}$ 。值得注意观测变量的样本空间和概率分布可以为离散型, 也可以为连续型。

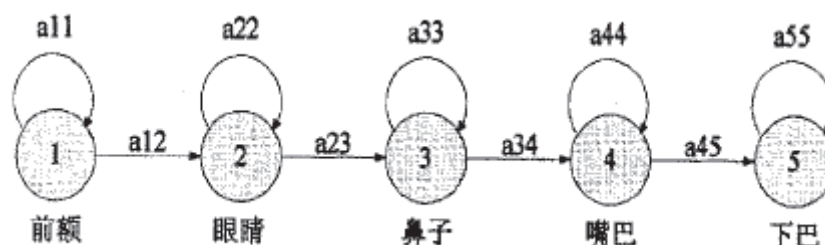
(5) 系统初始化状态分布 π 。系统初始状态概率分布可表示为 $\pi = \{\pi_i, i \leq i \leq N\}$, 其中 $\pi_i = P\{q_i = s_i\}$

综上可是, 要描述一个完整的 HMM, 需要模型参数 $\{N, M, A, B, \pi\}$ 。对于一个标准 HMM 模型, 需要解决模型训练, 隐状态估计和似然计算三个基本问题。

3. Hmm 在人脸识别的应用

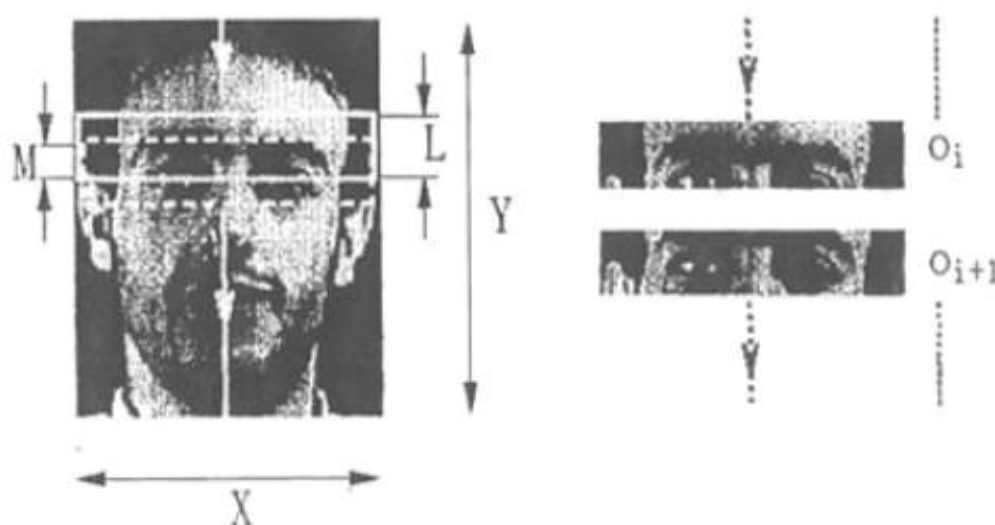
每个人的脸谱都有其个人特征, 同一个人的脸在不同拍摄条件下拍摄的照片中还存在种种随机性变异, 但其空间结构却具有稳定的相似共性, 即自上而下循序可分为前额、眼睛、鼻子、嘴和下巴 5 个组成部分。人脸的个人特征首先表现为上述组成部分形状及其相互联接关系的不同, 即可用一个一维隐马尔可夫模型表示。上述人脸 5 个部分的几何特征对应 HMM 的状态序列, 状态的出现及其转移可用模型中概率矩阵描述。对于一个既定的人脸, 所对应的 HMM 应当是惟一的, 人脸识别模型研究工作的任务就在于通过分析已收集的人脸图像去分析和建立其隐马尔可夫模型。

HMM 由一组状态序列组成。每个对象的人脸用一个 HMM 的参数来表示, 人脸的差异就是通过模型参数的不同表现出来的。人脸图像的 HMM 由以下参数表示: N 为垂直方向超状态数目; A 为状态转移概率矩阵, $A\{a_{ij}: 1 \leq i, j \leq N\}$; B 为状态的观察值概率密度函数, $B\{b_i(v), i \leq i \leq N, v \text{ 属于 } V\}$, 系统初始状态概率分布可表示为 $\pi = \{\pi_i, i \leq i \leq N\}$, 其中 $\pi_i = P\{q_i = s_i\}$



5 状态的人脸 HMM

本文取 5 个超状态分别对应人脸的前额、眼睛、鼻子、嘴和下巴, 分别描述和代表人脸的宏观特征。



3.1 特征提取

在人脸系统中，人脸图像的特征抽取是该系统的关键。人脸图像特征抽取的目的是获取一组“少而精”的分类特征，即获取特征数目少而分类错误概率小的特征向量。

对人脸检测划定的入脸区域，采用条状窗口从上到下的遍历扫描，如上图所示，如果将每个窗口所含像素作为观察向量，维数太高，且对影响识别光线角度等因素敏感。所以选取每个窗口的 KLT 稀疏为观察稀疏，也可以提取 AD-DCT 系数作为观察向量。在基于二阶 HMM 和嵌入式 HMM 人脸识别中提取 KLT 系数提取 2D-KCT 系数作为特征向量，主要是由于与 JPEG 图像压缩标准兼容。却又快速算法。

对于 $N \cdot N$ 的图像块，二维 DCT 公式如下：

$$C[I, k] = 4 \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m, n] \cos\left(\frac{(2m+1)i\pi}{2 \cdot N}\right) \cos\left(\frac{(2n+1)k\pi}{2 \cdot N}\right)$$

离散余弦变换的变换系数是一种较好的变换系数特征。离散余弦变换 DC' 的变换核为实数的余弦函数，不涉及复数运算，因而 DCT 的计算速度比变换核为复指数的 DFT 要快的多。而且 DCT 是

K-L 变换的最好近似。数据压缩是离散余弦变换的一个主要应用。另外，因为离散余弦变换与离散傅立叶变换很相近，所以可以对它进行有效地计算。基于这两个特性，可以用它作为人脸的特征。DCT 也具有好的模式识别性能，而且

由于 DCT 的计算速度，使用 DCT 进行人脸识别比 KLT 更有价值。这里我们选用离散余弦变换系数作为 HMM 的观察向量。HMM 观察序列的获得采用了遍历方法。假设人脸图像的长和宽尺寸分别 x 和 y ，在人脸图像平面上用一个像素点数为 $X \cdot L$ 采样窗，从上向下滑动，相邻窗口在垂直方向有 M 行，即步长为 $(L-M)$ 来获取观察图像块。如果直接取采样窗内图像块的像素点灰度

值构造观察向量，则存在下述问题：第一，观察向量维数太大，为 $x \times L$ ，计算分析量太大，第二，像素灰度值对光照条件变化及其他噪声干扰影响太敏感。而二维 DCT 的结果是能量分布向低频成分集中，因此取每个图像块 2D-D 交换后左上角的低频系数组成观察向量。

取不同大小的采样窗和采样图像块二维 DCT 的不同系数进行模型训练及识别精度比较。取较小的采样窗组成观察向量组，模型训练时间较长，识别精度并没有显著提高；取较多的采样图像块二维 DCT 系数，训练后模型的识别精度反而降低。对于人脸识别问题，为了更好地提取出反映人脸的根本特征，提取过于细微或仅靠宏观的特征信息，识别效果并不好。因此，取 L 为 16， M 为 12，取每个采样图像块的 6 个二维 DCT 系数组成观察向量。

3. 2HMM建模训练和人脸识别工作流程

每个人脸对象通常选取多个训练样本来训练 HMM。例如每个对象取 5 幅人脸图像组成训练样本，分别由正面脸像，小旋转角度(小于 10 度，下同)的左侧面脸像、小角度旋转右侧面脸像，小角度的仰视脸像和小角度的俯视脸像组成。人脸数据库中每个人的训练过程包括以下步骤：

- 1、对图像进行采样，进行图像预处理。
- 2、划定在人脸区域，得到观察向量序列。
- 3、初始化 HMM，确定 HMM 的状态数 N 和观察向量的维数。
- 4、迭代计算初始 HMM 参数。首先，图像被统一分割，每部分对应 HMM 的一个状态。然后，上述的分割数据被 viterbi 分割代替。这一过程的输出是一个初始的 HMM 模型，用作下一步重估 IMM 参数的输入。
- 5、使用 Baum. wblch 方法对 HMM 模型参数进行重估。根据训练图像的观察向量，HMM 参数将会被调整到一个局部极大值。这个过程的输出即为人脸数据库中人脸图像最终的 HMM 模型。训练结束后，人脸数据库中每个人都建立了一个特定的 HMM 模型。

识别过程是通过一个简单的 Viterbi 分类器进行的。先提取待识别的人脸图片的观察向量，然后使用 ViteIbi 算法计算该图像属于人脸数据库中每个人的概率，待识别的人即被识别为概率最大的那个人。识别过程如下：

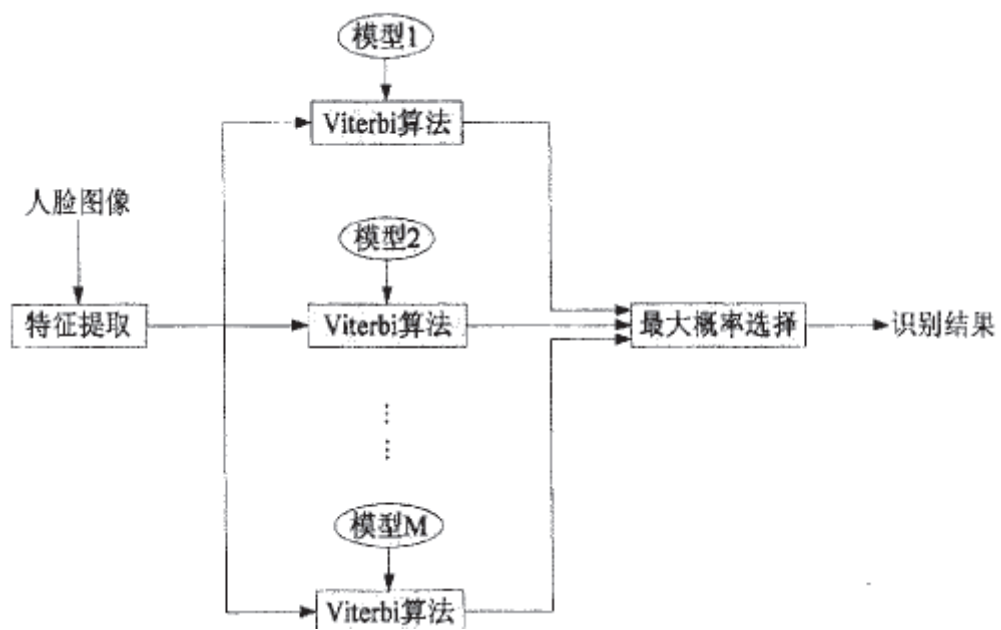
- 1、对待识别的人脸图片进行采样，得到观察向量序列 Q_{test}
- 2、上一步得到的观察向量序列与人脸库中的每个人的 mIM 模型进行匹配，计算下面的概率：

$$\Pr(O_{test}/r(K) | 1 \leq K \leq S)$$

实际应用中，为了计算的方便，常使用 Viterbi 算法计算出最大相似概率作

为上式的替代。

- 3、待识别的人即被识别为第 2 步算出概率最大的那个人。



流程图