HMM及其在语音识别中的应用

2015 04 22

余学辉

Introduction

一、**HMM基础理论**

二、HMM应用于语音识别

三、HMM用于语音识别的改进

3

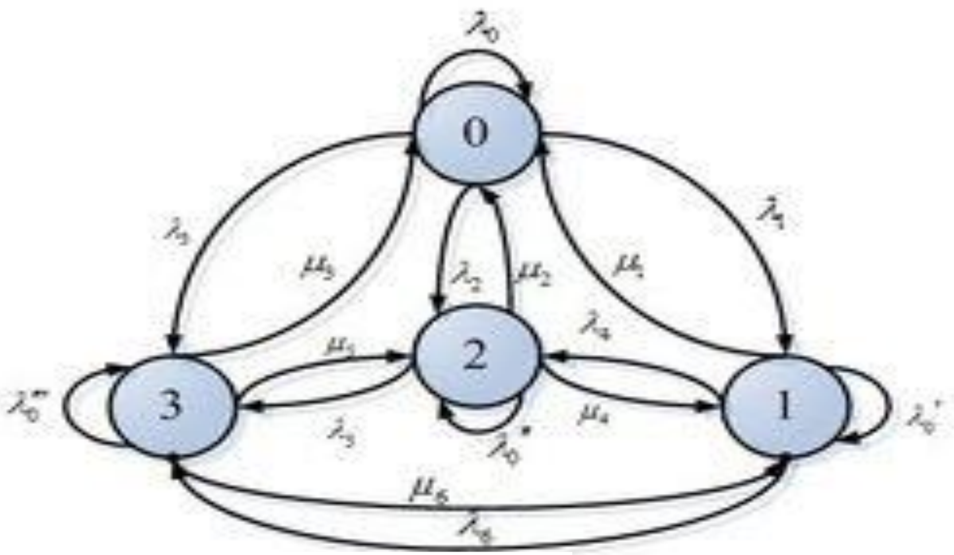
HMM基础理论

1. 离散马尔科夫链
2. HMM的定义（基本元素）
3. HMM的分类
4. HMM的三个问题及三个算法

离散马尔科夫链

1.一个系统有N个状态，$S_1, S_2, \ldots, S_N$ 。
随时间推移，系统连续地从一个状态跳转到另一个状态的随机过程即为马尔科夫链。

N = 4时的模型图



设q_t为时间t的状态，系统在时间t处于状态S_j的概率取决于其在时间1,2,...,t-1的状态，该概率为：

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots)$$

离散马尔科夫链

基本假设：

1.一阶马尔科夫链：

系统在t时间的状态只与其在时间t-1的状态相关

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \ldots) = P(q_t = S_j \mid q_{t-1} = S_i)$$

2. 时齐马尔科夫链：

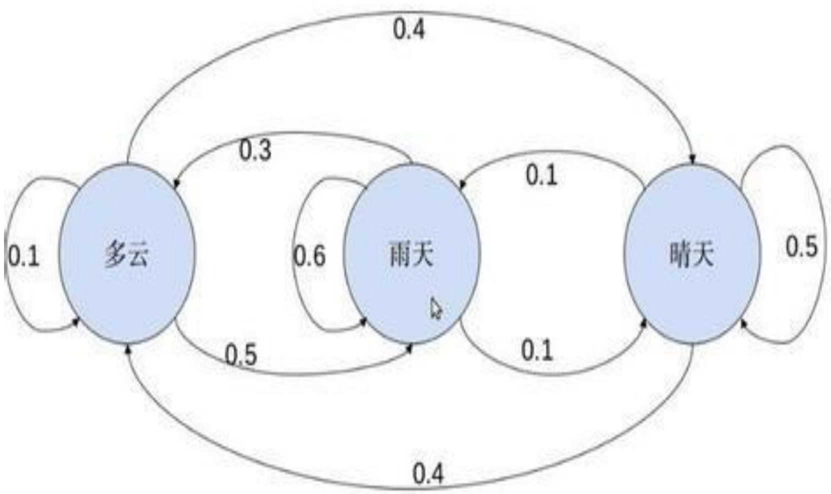
更进一步，当等式右边独立于时间t时，可以表示为

$$P(q_t = S_j \mid q_{t-1} = S_i) = a_{i,j}, 1 \le i, j \le N$$

其中，$a_{ij} \geq 0$, $\sum_{j=1}^{N} a_{ij} = 1$

离散马尔科夫链

- 很多时候这是不符实际的假设；
- 该假设很大程度上简化了模型，基于这两个假设模型图中所有的参数都是固定的常数



对该图进行建模：

M(S,A):

S = {S1：多云，S2：雨天，S3：晴天}

$A = [a_{ij}] =$

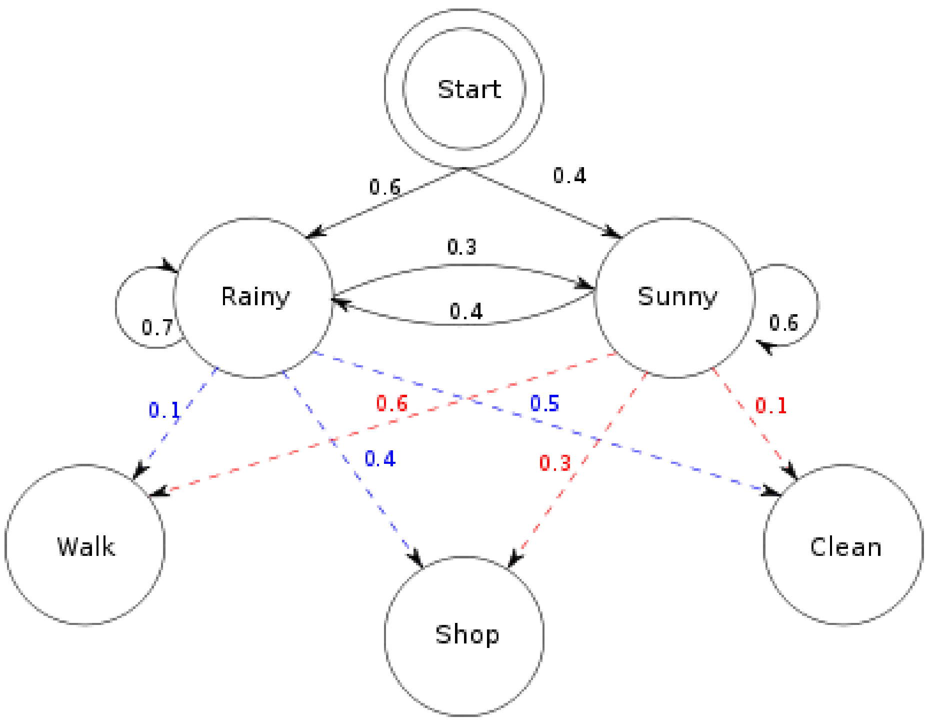
	S1	S2	S3
S1	0.1	0.5	0.4
S2	0.3	0.6	0.1
S3	0.4	0.1	0.5

注：每一行和为 1

HMM的定义

1. HMM(Hidden Markov Model, **隐马尔科夫模型**)

基于一阶时齐马尔科夫链模型



建模： $\lambda = (N, M, A, B, \pi)$

N是隐状态的数量 N=2

M是可观测状态的数量 M=3

π = [0.6,0.4]

A是隐状态的转移矩阵

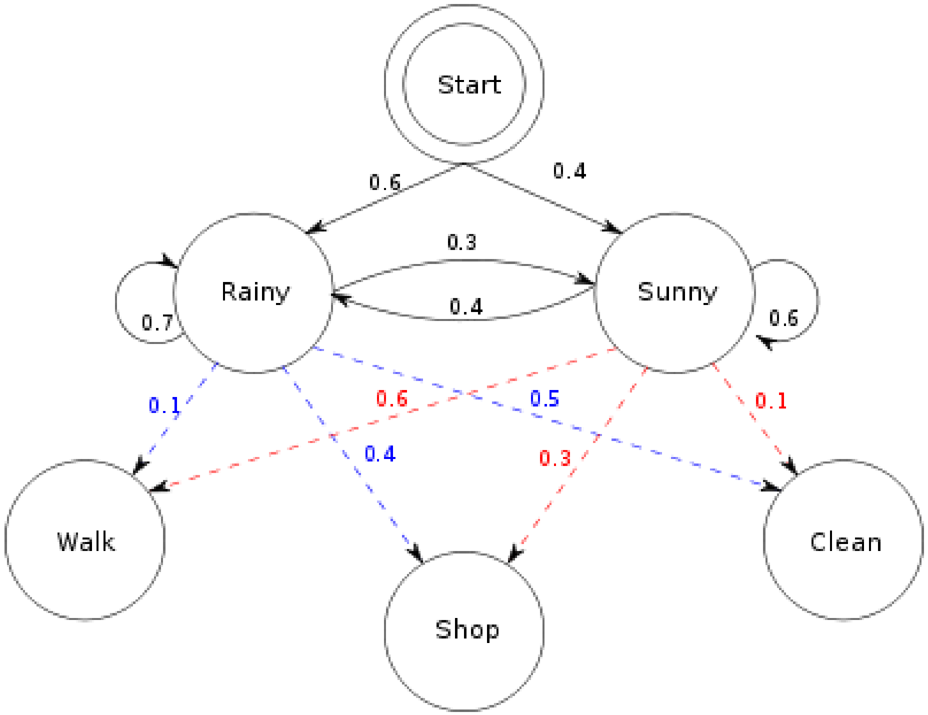
$A = [a_{ij}]$ =

	Rainy	Sunny
Rainy	0.7	0.3
Sunny	0.4	0.6

HMM的定义

1. HMM(Hidden Markov Model, **隐马尔科夫模型**)

基于一阶时齐马尔科夫链模型



建模： $\lambda = (N, M, A, B, \pi)$

B是混淆矩阵

$B = b_j(k) =$

	walk	shop	clean
rainy	0.1	0.4	0.5
sunny	0.6	0.3	0.1

一般用S表示隐状态组

S = [S1=Rainy，S2=Sunny]

一般用O表示可观测状态组

O = [O1=Walk，O2=Shop，O3=Clean]

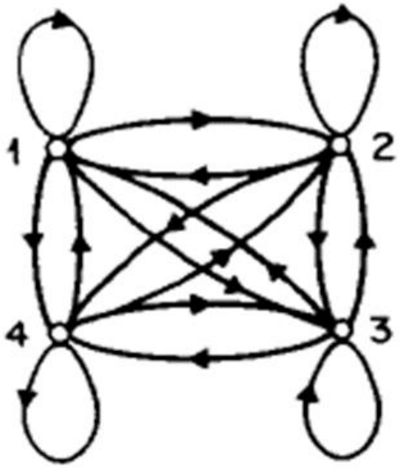
HMM定义

- HMM是一个双随机过程

HMM的分类

- 按照HMM的状态转移矩阵(A)分类
  - 遍历型模型(ergodict model)

  特点：从任何一个状态出发可以到达另外的任何一个状态



(a)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}.$$

a_{ij} != 0 for every i, j

HMM的分类

- 左右型模型(left-right model)

特点：

$$a_{ij} = 0, \quad j < i \tag{45}$$

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{46}$$

$$a_{ij} = 0, \quad j > i + \Delta \tag{47}$$

$$a_{NN} = 1 \tag{48a}$$

$$a_{Ni} = 0, \quad i < N. \tag{48b}$$

(b)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}.$$

HMM的分类

- 按B矩阵分类：

  1. 离散HMM(DHMM)：

     可观测序列的个数是离散的（有限或可列）

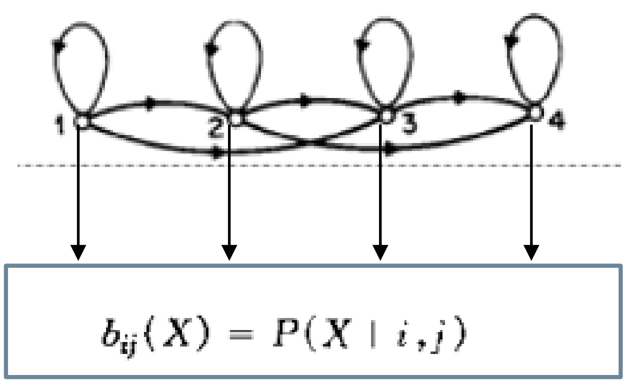
  2. 连续HMM(CHMM)：

  3. 半连续型：

HMM的分类

2. 连续HMM(CHMM):

$$b_{ij}(X) = P(X \mid i,j) = \frac{1}{(2\pi)^{p/2} \left| \Sigma_{ij} \right|^{1/2}} \exp\left\{ -\frac{1}{2}(X - \mu_{ij}) \Sigma_{ij}^{-1} (X - \mu_{ij})^t \right\}$$



- 连续型的变量往往服从某种函数分布，这里采用高斯函数
- 采用高斯函数的概率密度分布函数
- X表示可观测状态的数学表示，一般X是多维的，所以使用多维高斯概率密度函数

HMM的分类

高斯M元混合密度

$$b_{ij}(X) = \sum_{m=1}^{M} w_{ijm} b_{ijm}(X) = \sum_{m=1}^{M} w_{ijm} \frac{1}{(2\pi)^{p/2} \left| \sum_{ijm} \right|^{1/2}} \exp\left\{ -\frac{1}{2}(X - \mu_{ijm}) \sum_{ijm}^{-1} (X - \mu_{ijm})^t \right\}$$

这里 w_{ijm} 是混合系数，又叫分歧概率(Branch Probability)；$b_{ijm}(X)$ 叫分歧密度

HMM的分类

3. 半连续型：

$$b_{ij}(X)=\sum_{k=1}^{J}P(k\mid i,j)N(X,\mu_k,\sum_k)=\sum_{k=1}^{J}w_{ijk}N(X,\mu_k,\sum_k)$$

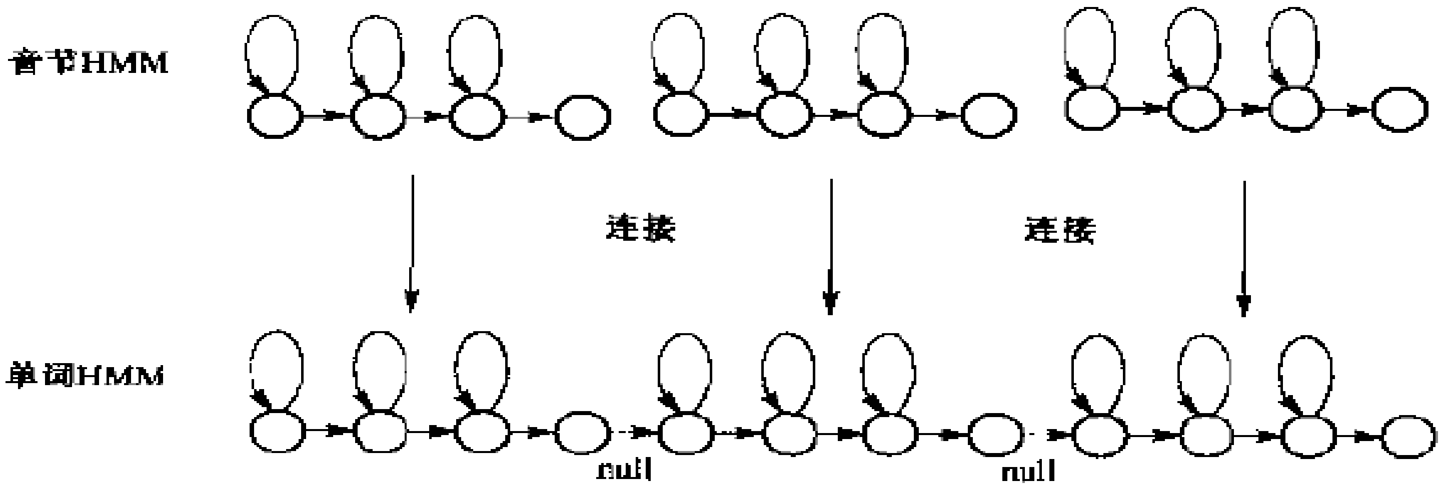
从式(5-42)可以看出，半连续型 HMM 的每个状态的输出概率分布是由几个正态分布函数叠加而成的，但是这些正态分布函数与状态无关(实际上与模型也无关)，即每个状态都使用共同的正态分布函数；而权值 w_{ijk} 与状态有关；k 实际上是离散 HMM 中码本的码矢，共有 J 个。

HMM的分类

其他形式：

1. 空转移：

允许转移到某一状态不产生输出。



图模型中用虚线连接，大量用于连续语音识别中连接若干音节模型

HMM的分类

2. 参数捆绑：

参数捆绑的基本思想是在 HMM 的不同状态转移弧的参数之间建立一定的关系，使得不同状态转移弧使用相同的参数，其目的就是使模型中的独立的状态参数减少，从而使得参数估计变得更简单。

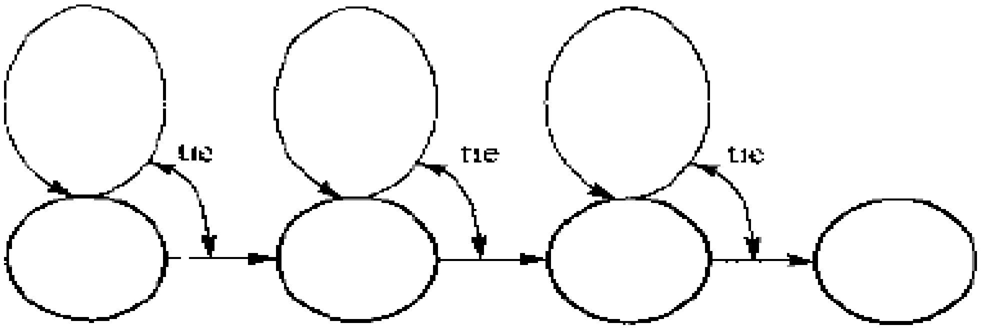


图 5-7 具有参数捆绑的连续型 HMM

应用领域

- 统计学
- 生物信息学
- 金融分析
- 语音识别
- 自然语言处理
- 网络信息安全
- 行为分析

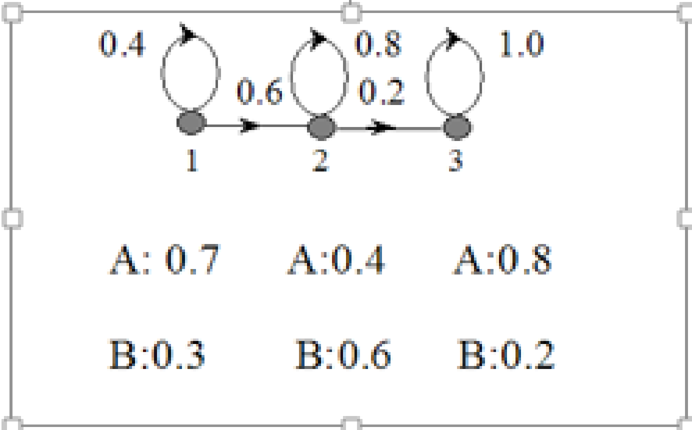
三个问题

- **评估问题**：给定可观测序列$O = O_1O_2 \ldots O_T$以及模型$\lambda = (A, B, \pi)$，如何计算该可观测序列出现的概率$P(O|\lambda)$？

- **解码问题**：给定可观测序列$O = O_1O_2 \ldots O_T$，以及模型$\lambda = (A, B, \pi)$，如何选取一个状态序列$S = s_1s_2 \ldots s_2$，使得该可观测序列出现的可能性最大?

- **学习问题**：给定观察序列，如何调整模型$\lambda = (A, B, \pi)$的参数，使得可观测序列出现的概率$P(O|\lambda)$最大?

解决方法：问题一



- HMM模型如上图所示
- 观察序列$O = ABAB$
- 计算$P(O|\lambda)$？

注：题中A,B表示可观测状态

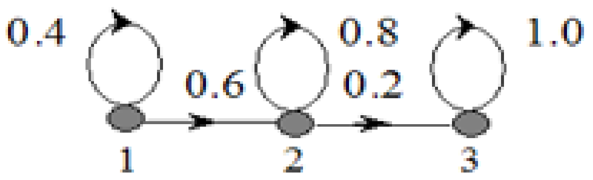
解决方法：问题一

- 求解方法：前向-后向算法
- 前向算法
  - 思想：高效率地计算向前变量，以求得最终结果
  - 前向变量：$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = s_i | \lambda)$
  - 过程：
    - 初始化：$\alpha_1(i) = \pi_i b_i(O_1)$，$1 \le i \le N$
    - 递归：$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1})$，$1 \le t \le T-1$，$1 \le j \le N$
    - 终止：$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

解决方法：问题一

- 后向算法
  - 思想：与前向算法本质是一样的，只不过递归的方向不同
  - 后向变量：$\beta_t(i) = P(O_{t+1}O_{t+2} \dots O_T | q_t = s_i, \lambda)$
  - 过程：
    - 初始化：$\beta_T(i) = 1$，$1 \le i \le N$
    - 递归：$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$，$t = T-1$，$T-2$，$\dots, 1$，$1 \le i \le N$
    - 终止：$P(O|\lambda) = \sum_{i=1}^{N} \beta_1(i)$

解决方法：问题二



A: 0.7　A:0.4　A:0.8

B:0.3　B:0.6　B:0.2

- HMM模型如上图所示
- 观察序列$O = ABAB$
- 求状态序列$S = s_1 s_2 \dots s_2$，使得观察序列出现的可能性最大

解决方法：问题二

-

- 求解方法：维特比算法
- 维特比算法:
  - 思想：利用动态规划求解
  - Viterbi变量:

$$\delta_t(i) = \max_{q_1,q_2,\ldots,q_{t-1}} P(q_1, q_2, \ldots, q_t = s_i, O_1, O_2, \ldots, O_t|\lambda)$$

解决方法：问题二

- 过程：
  - 初始化：$\delta_1(i) = \pi_i b_i(O_1)$，$\varphi_1(i) = 0$，$1 \le i \le N$
  - 递归：

$$\delta_t(j) = \left[\max_{1\le i\le N} \delta_{t-1}(i) a_{ij}\right] b_j(O_t)，2 \le t \le T，1 \le j \le N$$

$$\varphi_t(j) = arg\ \max_{1\le i\le N} [\delta_{t-1}(i) a_{ij}]$$

  - 终止：

$$p^* = \max_{1\le i\le N} [\delta_T(i)]$$

$$q_T^* = arg\ \max_{1\le i\le N} [\delta_T(i)]$$

  - 路径回溯：

$$q_t^* = \varphi_{t+1}(q_{t+1}^*)，t = T\text{-}1，T\text{-}2，\ldots，1$$

解决方法：问题三

- Baum-Welch 算法：
- 说明：该方法并不保证能获得最优解，只能拿获得一个局部最优解。
- 步骤：（假设只有一组可观测序列数据）
- 1. 按一定要求，适当的选择模型的参数

$$\lambda = (N, M, A, B, \pi)$$

- 2. 进行参数调整：

解决方法：问题三

$$\bar{\pi}_i = \text{expected frequency (number of times) in state } S_i \text{ at time } (t = 1) = \gamma_1(i)$$

$$\bar{a}_{ij} = \frac{\text{expected number of transitions from state } S_i \text{ to state } S_j}{\text{expected number of transitions from state } S_i}$$

$$= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\bar{b}_j(k) = \frac{\text{expected number of times in state } j \text{ and observing symbol } v_k}{\text{expected number of times in state } j}$$

$$= \frac{\sum_{\substack{t=1 \\ \text{s.t. } O_t = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}.$$

解决方法：问题三

a) E步骤：由根据以下公式，计算期望值$\xi_t(i,j)$和$\gamma_t(i)$

$\xi_t(i,j)$表示在给定HMM和观察序列，在时间t位于状态i，时间$t+1$位于状态j的概率：

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda)$$
$$= \frac{P(q_t = S_i, q_{t+1} = S_j, O \mid \lambda)}{P(O \mid \lambda)}$$
$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O \mid \lambda)}$$
$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}$$

$\gamma_t(i)$表示在给定*HMM和观测序列，在时间t位于状态i的概率*

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i,j)$$

解决方法：问题三

b) M步骤：用E步骤所得到的期望值，根据以下公式重新估计π_i，a_{ij}，$b_j(k)$，得到模型λ_{i+1}

$$\pi_i = q_1为S_i的概率 = \gamma_1(i)$$

$$a_{ij} = \frac{Q中从状态q_i转移到q_j的期望次数}{Q中从状态q_i转移到另一状态(包括q_i本身)的期望次数} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

$$b_j(k) = \frac{Q中由状态q_j输出v_k的期望次数}{Q到达q_j的期望次数} = \frac{\sum_{t=1}^{T}\gamma_t(j)\times\delta(O_t, v_k)}{\sum_{t=1}^{T}\gamma_t(j)}$$

解决方法：问题三

已经证明， $P(O/\bar{\hat{M}})>P(O/M)$ ，即调整后的似然概率大于调整前

3. 重复步骤2，直至收敛。

Introduction

一、HMM基础理论

二、HMM应用于语音识别

三、HMM用于语音识别的改进

HMM用于语音识别

1. 引入HMM
2. 基本过程
3. 下溢问题
4. 参数初始化问题
5. 训练多组观测序列时的Baum-Welch算法
6. B矩阵为连续型或半连续型时的训练

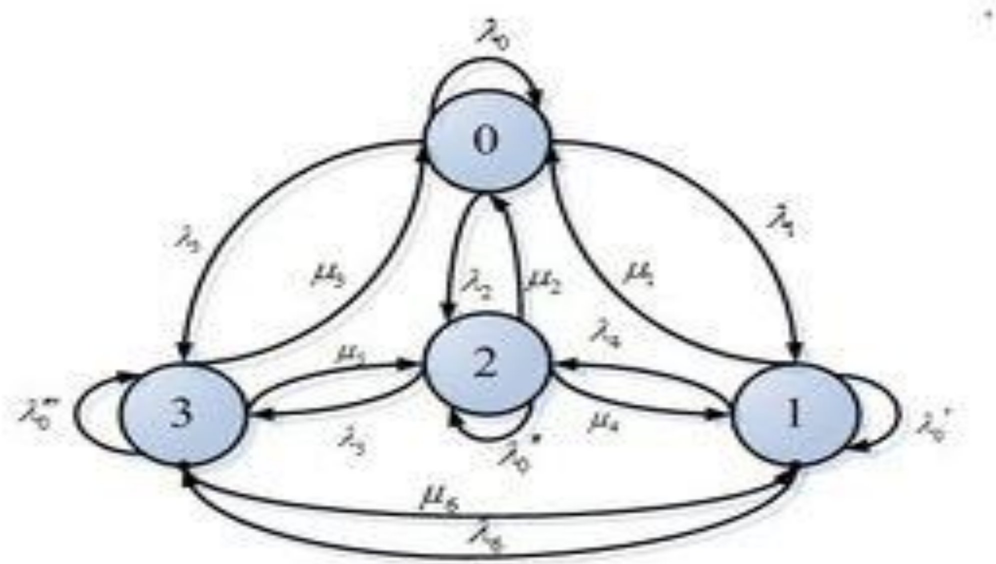
引入HMM

1. 语音具有短时平稳性，对平稳信号处理（数学建模，特征提取）的技术较为成熟。所以对于短时的语音信号处理也较为成熟

2. 但整体上来说，语音信号是时变的，非平稳的。如何处理？

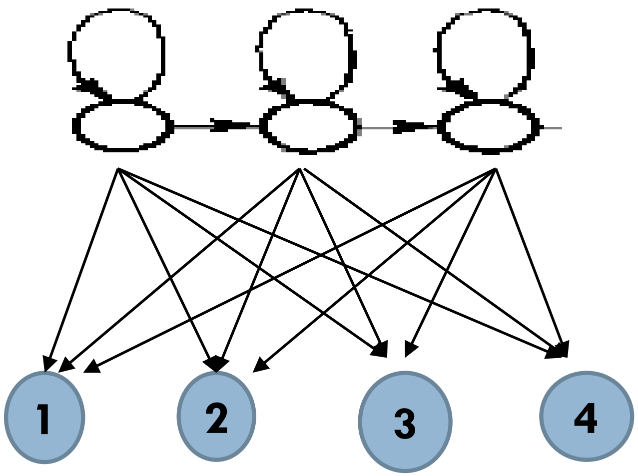
马尔科夫链可以给语音信号建模

虽然马尔科夫链可以描述语音信号，但不是最佳和最有效的



引入HMM

- 3. 因此引入HMM



- 用可观测层对基元发音的声学变化建模
- 用隐藏层对基元音发音速率建模

基本过程

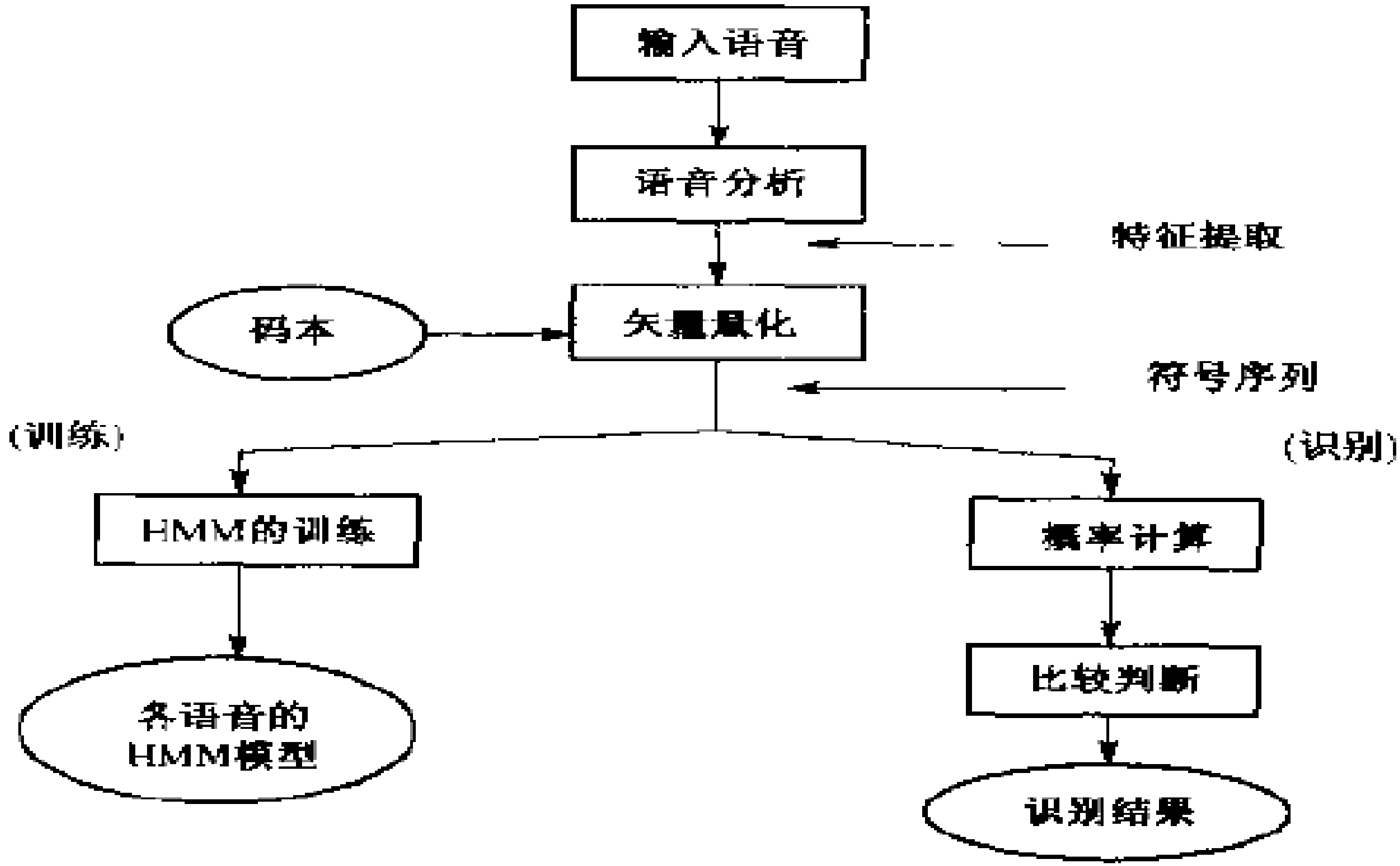
HMM处理三种问题，我理解为三种功能：

1. 给出 O (可观测序列o1 o2 o3... ot），求P(O | hmm)

2. 给出 O，求出使P(O, H | hmm)最大的 H （H为与O对应的隐藏层时序序列）

3. 给出 O，求出使P(O | hmm) 最大的 hmm（参数调整）

基本过程


输入语音
语音分析
特征提取
码本
矢量量化
符号序列
(训练)
(识别)
HMM的训练
概率计算
各语音的
HMM模型
比较判断
识别结果

基本过程

应用于孤立词语音识别：

（训练：功能三）
1. 给出词汇 a ，b ，c 的 mfcc 作为每个词汇的 O 序列，记为Oa，Ob，Oc；
2. 基于上述的HMM的第三个功能，以Oa，Ob，Oc作为条件，分别获得三个HMM，记为HMMa，HMMb，HMMc；

（识别：功能一）
3. 给出未知词汇 x 的mfcc ， 记为Ox；
4. 基于HMM的第一个功能，将 Ox 输入到HMMa，HMMb，HMMc中获得三个概率，记为Pa，Pb，Pc；
5. 比较Pa，Pb，Pc,获取最大的Py(y = a 或 b 或 c ）
6 x = y;

下溢问题

前向变量和后向变量

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(O_{t+1}), \qquad 1 \le t \le T-1$$

$$1 \le j \le N.$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j),$$

$$t = T-1, T-2, \cdots, 1, 1 \le i \le N.$$

随着递归，$\alpha_{t+1}(j)$ 和 $\beta_t(i)$ 在逐步减小，当 t 足够大时，两者将趋于0，在计算时可能造成下溢问题。

下溢问题

首先考虑前向变量 $\alpha_t(i)$。在计算 $\alpha_t(i)$时，在每个时刻 t，按照递归公式计算该时刻的值，要同时乘上下面的定标因子：

$$c_t = \frac{1}{\sum_{i=1}^{N} \alpha_t(i)} \tag{5-46}$$

类似于在每个时刻都进行一次归一化

通过归纳法可知：

$$\tilde{\alpha}_{t-1}(j) = \left(\prod_{\tau=1}^{t-1} c_\tau\right)\alpha_{t-1}(j)$$

对后向变量进行类似处理

下溢问题

经过上面的处理后，$\hat{a}_{ij}$变为如下的形式：

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \tilde{\alpha}_t(i) a_{ij} b_j(O_{t+1}) \tilde{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} \tilde{\alpha}_t(i) a_{ij} b_j(O_{t+1}) \tilde{\beta}_{t+1}(j)}$$

其中的 $\tilde{\alpha}_t(i)$和 $\tilde{\beta}_{t+1}(j)$由下式确定：

$$\tilde{\alpha}_t(i) = \left[\prod_{s=1}^{t} c_s\right]\alpha_t(i) = C_t \alpha_t(i)$$

$$\tilde{\beta}_{t+1}(j) = \left[\prod_{s=t+1}^{T} c_s\right]\beta_{t+1}(j) = D_{t+1}\beta_{t+1}(j)$$

将以上两式带入重估公式得：

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} C_t \alpha_t(i) a_{ij} b_j(O_{t+1}) D_{t+1} \beta_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} C_t \alpha_t(i) a_{ij} b_j(O_{t+1}) D_{t+1} \beta_{t+1}(j)}$$

下溢问题

$$C_t D_{t+1} = \prod_{s=1}^{t} C_s \prod_{s=t+1}^{T} C_s = \prod_{s=1}^{T} C_s = C_T \tag{5-56}$$

因此，由于 C_T 与时间无关，因此，在重估公式中，分子和分母中的重估公式可以相互约掉，从而实现定标的目的。

参数初始化

- 1. 通过调整公式，如果参数初值为0，则始终为零，即左右型经过调整仍是左右型

- 2. π和 A的初始设置对识别率影响不大，参数B较前两者影响更大。

- 3.

需要说明的是：语音识别一般采用从左到右型 HMM，所以初始状态概率 π_i 不需要估计，总是设定为：

$$\pi_1 = 1;$$

$$\pi_i = 0 \quad (i = 2, \cdots, N)。 \tag{5-21}$$

参数初始化

对于离散性HMM，一般采用均匀或随机赋值

均匀赋值：

A:给予从状态 i 转移出去的每条弧相等的转移概率

$$a_{ij}=\frac{1}{\text{从状态 } i \text{ 转移出去的弧的条数}}$$

B:给予每一个输出观察符号相等的输出概率初始值

$$b_{ij}(k)=\frac{1}{\text{码本中码字的个数}}$$

并且每条弧上给予相同的输出概率距阵；

参数初始化

- 一、小语音单位
- 可采用手工对输入语音进行状态划分并统计出相应的概率分布作为初值

- 二、大语音单位：
- 普遍采用分段**k-means** 算法。

训练多组观测序列时的Baum-Welch算法

《语音信号处理》提供方法

2)给定一个(训练)观察值符号序列 $O=o_1,o_2,\cdots,o_T$,由初始模型计算 $\gamma_t(i,j)$等,并且,由上述重估公式,计算 $\hat{a}_{ij}$和$\hat{b}_{ij}(k)$;

3)再给定一个(训练)观察值符号序列 $O=o_1,o_2,\cdots,o_T$,把前一次的 $\hat{a}_{ij}$和$\hat{b}_{ij}(k)$作为初始模型计算 $\gamma_t(i,j)$等,由上述重估公式,重新计算 $\hat{a}_{ij}$和$\hat{b}_{ij}(k)$;

4)如此反复,直到 $\hat{a}_{ij}$和$\hat{b}_{ij}(k)$收敛为止。

训练多组观测序列时的Baum-Welch算法

- 论文《A Tutorial on Hidden Markov Models》的方法是

$$\overline{a_{ij}} = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\, a_{ij} b_j(O_{t+1}^{(k)})\, \beta_{t+1}^k(j)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\, \beta_t^k(i)}$$

$$P(O|\lambda) = \prod_{k=1}^{K} P(O^{(k)}|\lambda) = \prod_{k=1}^{K} P_k.$$

$$\overline{b_j}(\ell) = \frac{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{\substack{t=1 \\ \text{s.t. } O_t = v_\ell}}^{T_k-1} \alpha_t^k(i)\, \beta_t^k(i)}{\sum_{k=1}^{K} \frac{1}{P_k} \sum_{t=1}^{T_k-1} \alpha_t^k(i)\, \beta_t^k(i)}$$

Introduction

一、HMM基础理论

二、HMM应用于语音识别

三、**HMM用于语音识别的改进**

HMM用于语音识别的改进

1. 提高HMM描述语音动态特性的能力
2. MCE(minimum classification Error Method)
3. 直接利用状态持续时间

提高HMM描述语音动态特性的能力

- HMM基于以下假设：
- 1. 状态转移概率(A矩阵)只与前一时刻的状态有关，与（过去的）观察序列无关，且时不变。
- 2. 状态转移概率密度函数(B矩阵)与过去的状态无关，与过去的观察序列无关。

- 但是语音时发声系统连续变化产生的，有很强的相关性。
- 以上假设是不合理的。

提高HMM描述语音动态特性的能力

- 改进方法一：（改进特征）

- 利用语音的静态特征参数，通过差分获取动态特性

$$\Delta X(t) = \frac{\sum_{i=-w}^{w} iX(t+i)}{\sum_{i=-w}^{w} i^2}$$

提高HMM描述语音动态特性的能力

- 改进方法二：（模型改进）

- 复数帧段输入HMM：
- 1. 将相继N帧的特征参数矢量按序连接成一个特征矢量。
- 2. 帧移为一帧
- 3. 这样称为N帧帧宽的复数帧输入HMM

提高HMM描述语音动态特性的能力

- 改进方法二：（模型改进）
- 1. 随着连接，特征向量维数将很大，计算资源消耗大
- 2. 数据不足时，模型精度反而下降

- 常采用的解决方法：
- 1. 使用修正高斯分布函数（MGDF）和径向基（RBF）函数代替传统高斯函数
- 2. 通过PCA，K-L变换降维。

MCE

HMM模型的代表性训练方法是Baum-Welch的最大似然（ML）法。它追求的是似然函数最大化。

即最大化

$$P(O|\lambda)$$

和他不同的另一种训练方法是MCE，追求分类误差最小化。

$$L(X,\vec{\theta})$$

它的问题在于如何定义这个误差。

提高HMM描述语音动态特性的能力

- 识别函数：

$$g(X_{k,n},\vec{\theta}) = \log\left[\left\{\sum_{all\ S} P(X_{k,n}, S_c \mid \theta_c)^{\xi}\right\}^{\frac{1}{\xi}}\right]$$

- 误差函数：

$$d(X_{k,n},\vec{\Theta}) = -g(X_{k,n},\vec{\theta}_k) + \left[\frac{1}{M}\sum_{p(p\neq k)} g(X_{k,n},\vec{\theta}_p)^{\eta}\right]^{\frac{1}{\eta}}$$

- 损失函数：

$$l(d(X_{k,n},\vec{\Theta})) = \frac{1}{1+\exp(-\alpha d)(X_{k,n},\vec{\Theta}))}$$

- 总损失：

$$L(X,\vec{\Theta}) = \sum_{k=1}^{K}\sum_{n=1}^{N_k} l(d(X_{k,n},\vec{\Theta}))$$

MCE

方法：梯度下降

$$\vec{\Theta}(n+1)=\vec{\Theta}(n)-\varepsilon_{\vec{\Theta}(n)}\nabla L(X,\vec{\Theta}(n))$$

均值：

$$\Delta\mu(n+1)=\varepsilon_{\mu}(n)\frac{\partial L(\mu(n))}{\partial\mu(n)}+m\Delta\mu(n)$$

方差：

$$\Lambda=P^{-1}\Sigma^{-1}P=diag(\lambda_1,\cdots,\lambda_d,\cdots,\lambda_D)$$

$$\Sigma^{-1}=P\Lambda P^{-1}$$

$$\lambda(n+1)=\lambda(n)-\varepsilon_{\lambda}(n)\frac{\partial L(X,\lambda(n))}{\partial\lambda(n)}$$

MCE

当上式计算量太大，可用下列方法优化。

(1) $n=1$。

(2)根据 $\vec{\Theta}(n)$选择 D 个值 $\varepsilon_1,\cdots,\varepsilon_D$，由式(5-78)求出 $\vec{\Theta}(n+1),\cdots,\vec{\Theta}(n+1)_D$。

(3)求出损失函数 $L(X,\vec{\Theta}(n+1)_1),\cdots,L(X,\vec{\Theta}(n+1)_D)$。

(4)在 3)中使损失函数最小的参数集作为新的参数集 $\vec{\Theta}(n+1)$。

$$\vec{\Theta}(n+1)=\operatorname{argmin}\{L(X,\vec{\Theta}(n+1)_d)\}$$

(5)对于阈值 δ 满足下例条件时转向 2)。

$$|L(X,\vec{\Theta}(n))-L(X,\vec{\Theta}(n+1))|>\delta$$

(6)结束。

直接利用状态持续时间

HMM在隐状态i处持续n帧的概率可表示为：

$$d_i(n)=a_{ii}^{n-1}(1-a_{ii})$$

即一个音持续时间越长，其概率呈指数型下降，这与实际不符。

直接利用状态持续时间

方法一：增加HMM的状态数

增加状态数，然后采取状态捆绑，相同的稳定态，用相同的参数

方法二：采用后处理方法

1. 求出 $P(X_1, X_2, \cdots, X_T)$（问题一）
2. 用viterbi算法求出该输出可观测序列X的最可能隐状态序列S
3. 从S中计算出在i状态停留时间 τ_i 的似然度 $\log d_i(\tau_i)$
4. 修正 $P(X_1, X_2, \cdots, X_T)$

$$\log P(X) = \log P(X) + w \sum_{i} \log d_i(\tau_i)$$

其中 $\sum_{\tau} d_i(\tau) = 1$

直接利用状态持续时间

方法三：采用状态持续时间分布的HMM系统

前后向概率计算

$$\alpha_t(i) = \sum_j \sum_{\tau \leq t} \alpha_{t-\tau}(j) a_{ji} d_i(\tau) \prod_{k=1}^{\tau} b_{ji}(X_{t-k+1})$$

$$\beta_t(i) = \sum_j \sum_{\tau \leq T-t} a_{ij} d_j(\tau) \prod_{k=1}^{\tau} b_{ij}(X_{t-k+1}) \beta_{t+\tau}(j)$$

修正后输出X的概率计算

$$P(X) = \sum_i \sum_j \sum_{t'-t \leq \tau} \sum_{t \leq t'} \alpha_{t'-\tau}(i) a_{ij} d_j(\tau) \prod_{k=1}^{\tau} b_{ij}(X_{t'-k+1}) \beta_{t'}(j)$$

直接利用状态持续时间

离散型参数重估：

$$\hat{a}_{ij} = \frac{\sum_t \sum_{\tau \le t} \alpha_{t-\tau}(i) a_{ij} d_j(\tau) \prod_{k=1}^{\tau} b_{ij}(X_{t-k+1}) \beta_t(j)}{\sum_t \alpha_t(i) \beta_t(i)}$$

$$\hat{b}_{ij}(h) = \frac{\sum_t \sum_{\tau \le t} \alpha_{t-\tau}(i) a_{ij} d_j(\tau) c_h(\tau) \prod_{k=1}^{\tau} b_{ij}(X_{t-k+1}) \beta_t(j)}{\sum_t \sum_{\tau \le t} \alpha_{t-\tau}(i) a_{ij} d_j(\tau) \tau \prod_{k=1}^{\tau} b_{ij}(X_{t-k+1}) \beta_t(j)}$$

直接利用状态持续时间

连续性参数重估：

$$\hat{a}_{ij} = \frac{\sum_{t}\sum_{\tau\leqslant t}\alpha_{t-\tau}(i)a_{ij}d_j(\tau)\prod_{k=1}^{\tau}b_{ij}(X_{t-k+1})\beta_t(j)}{\sum_{t}\alpha_t(i)\beta_t(i)}$$

$$\hat{\mu}_{ij} = \frac{\sum_{t}\sum_{\tau\leqslant t}\alpha_{t-\tau}(i)a_{ij}d_j(\tau)\prod_{k=1}^{\tau}b_{ij}(X_{t-k+1})\beta_t(j)\left[\sum_{k=1}^{\tau}X_{t-k+1}\right]}{\sum_{t}\sum_{\tau\leqslant t}\alpha_{t-\tau}(i)a_{ij}d_j(\tau)\tau\prod_{k=1}^{\tau}b_{ij}(X_{t-k+1})\beta_t(j)} \qquad (5\text{-}101)$$

$$\hat{\Sigma}_{ij} = \frac{\sum_{t}\sum_{\tau\leqslant t}\alpha_{t-\tau}(i)a_{ij}d_j(\tau)\prod_{k=1}^{\tau}b_{ij}(X_{t-k+1})\beta_t(j)\left[\sum_{k=1}^{\tau}(X_{t-k+1}-\mu_{ij})(X_{t-k+1}-\mu_{ij})^t\right]}{\sum_{t}\sum_{\tau\leqslant t}\alpha_{t-\tau}(i)a_{ij}d_j(\tau)\tau\prod_{k=1}^{\tau}b_{ij}(X_{t-k+1})\beta_t(j)}$$

直接利用状态持续时间

i $d(\tau)$ i 可采用三种分布：

1. 离散型概率分布：

d[i, t] = S序列中在状态i持续时间为t的帧数/S序列中总帧数

2. 高斯分布：

$$d_i[\tau] = \frac{1}{\sqrt{2\pi}\rho_i}\exp\left[-\frac{1}{2\rho_i^2}(\tau-\mu_i)^2\right], \quad \begin{matrix} i = 1 \sim N \\ \tau \leqslant D \end{matrix}$$

3. Gamma分布

$$d_i[\tau] = \eta_i^{\gamma_i}\tau^{\gamma_i - 1}e^{-\eta_i\tau}/\Gamma(\gamma_i)$$

$$\Gamma(\gamma) = \frac{1}{\gamma}\prod_{n=1}^{\infty}\left(1+\frac{1}{n}\right)^{\gamma}\left(1+\frac{\gamma}{n}\right)^{-1} = \int_0^{\infty}e^{-t}t^{\gamma-1}dt \quad , \quad \begin{matrix} i = 1 \sim N \\ \tau \leqslant D \end{matrix}$$

参考文献

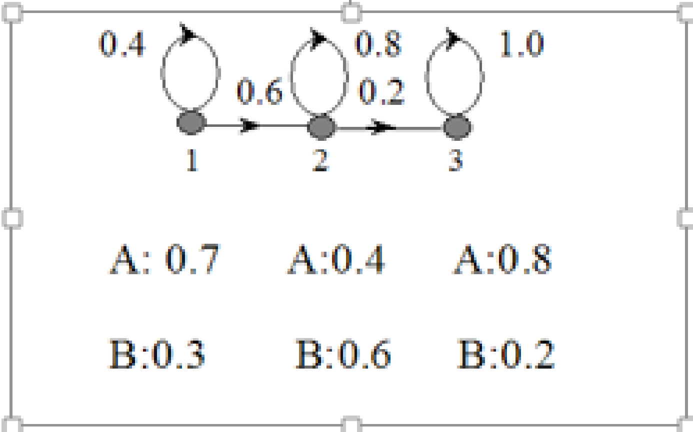
1. 论文《A Tutorial on Hidden Markov Models》

2. 《语音信号处理》-----赵立

3. “马尔科夫链” “HMM”-----维基百科

4.PPT《hmm_and_htk》----秦春来

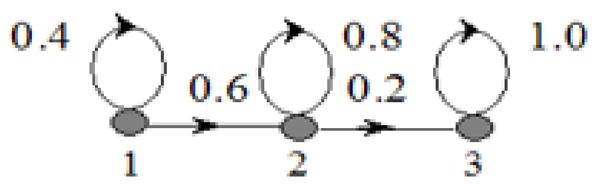
作业



- HMM模型如上图所示
- 观察序列$O = ABAB$
- 计算$P(O|\lambda)$？

注：题中A,B表示可观测状态

作业



A: 0.7　　A:0.4　　A:0.8

B:0.3　　B:0.6　　B:0.2

- HMM模型如上图所示
- 观察序列$O = ABAB$
- 求状态序列$S = s_1 s_2 \dots s_2$，使得观察序列出现的可能性最大

作业

- 说明：
- 1.不要求求出结果，但要求详细步骤。
- 2.可以写在纸上，拍张照片即可
- 3.详细步骤可以参考ppt《hmm_and_htk》
- 4.邮箱：y19941010@126.com
- 5.deadline : 2015-05-06（下周三）