GMM-HMM

1. GMM-HMM 思想

   1）Hmm 是隐状态是离散的，观测状态也是离散的。hmm 中发射概率 B 是离散的。

   2）Gmm-hmm 就是把发射概率替换成 gmm（高斯混合模型）。

   3）gmm 参数求解是用 em 算法

   4）hmm 参数求解也是用 em 算法

   5）Gmm-hmm 参数求解也是这样的...

   6）拿语音识别为例，隐状态是三音素。语音识别的目的是将一串语音识别成一串音素，相当于 hmm 的解码的任务。首先会用 mfcc 提语音的特征，这样每一帧的语音都会对应成一个 13 维的向量表示。目标就是将 13 维的向量所对应的隐状态（三音素，常用三音）求出来。当然一串语音是由一连串的帧表示成的。所以一串语音最终解码成一串三音素。 这个任务很适合用 hmm 来做。但是 hmm 的发射概率是离散的，也就是说观测状态是有限的，而语音识别中的观测状态是无限的 13 维向量。所以需要连续型 hmm 来做。将发射概率，也就是一个三音素（隐状态）的发射概率变成 gmm。gmm 中高斯的个数可以自由设定，每个高斯都是 13 维的，就是（X1,X2,...,X13)服从联合正态分布。 那如果知道这个三音素服从的高斯混合模型的参数，将这个帧对应的 mfcc 特征向量代入这个 gmm 就可以得到这个三音素转移到这个状态的发射概率。

   值得注意的是一个三音素就要对应一个 gmm（里面有很多参数），如果有 n 个三音素，那么就要有 n 个 gmm 模型，每个 gmm 模型都对应很多个参数。

   参数个数计算：13 个均值，13 个方差，k 个高斯模型对应 k 个权重。就是 26*k+k 个参数。 如果有 n 个三音素，则有 n*k*27 个参数

为了方便学习算法，不拿 13 维高斯模型举例子，太复杂，拿 1 维。同时 1 维和 13 维的计算方法是相似的。

Continuous Density HMM

$$b_j(o) = \sum_{k=1}^{M} c_{jk} N(o; \mu_{jk}, U_{jk})$$

N( ): Multi-variate Gaussian

μ_{jk}: mean vector for the k-th mixture component

U_{jk}: covariance matrix for the k-th mixture component

$\sum_{k=1}^{M} c_{jk} = 1$ for normalization

2. GMM-HMM 参数求解

1）gmm-hmm 中的参数有 hmm 中的参数：A、B、Pi 和 gmm 中的参数：多个高斯模型的均值、方差还有每个高斯的权重。

首先观察离散 hmm 中构造的两个变量：

$$\gamma_t(i)=\frac{\alpha_t(i)\,\beta_t(i)}{\sum_{i=1}^{N}[\alpha_t(i)\,\beta_t(i)]}=\frac{P(\bar{O},q_t=i|\lambda)}{P(\bar{O}|\lambda)}=P(q_t=i|\bar{O},\lambda)$$

$$\varepsilon_t(i,j)=\frac{\alpha_t(i)\;a_{ij}\;b_j(o_{t+1})\;\beta_{t+1}(j)}{\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_t(i)\;a_{ij}\;b_j(o_{t+1})\;\beta_{t+1}(j)}$$

$$=\frac{P(\bar{O},q_t=i,q_{t+1}=j|\lambda)}{P(\bar{O}|\lambda)}=P(q_t=i,q_{t+1}=j|\bar{O},\lambda)$$

思考：

虽然 A 的转移概率是没有变的，但是 B 的发射概率变了。一个隐状态要对应一个 gmm，而 gmm 中的参数并没有在上述变量中进行体现。

构造出来的两个变量：

$\gamma_t(j,k)=\gamma_t(j)$ but including the probability of o_t evaluated in the k-th mixture component out of all the mixture components

$$=\left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\right]\left[\frac{c_{jk}\,N(o_t;\,\mu_{jk},\,U_{jk})}{\sum_{m=1}^{M}c_{jm}N(o_t;\,\mu_{jm},\,U_{jm})]}\right]$$

$$\varepsilon_t(i,j)=\frac{\alpha_t(i)\;a_{ij}\;b_j(o_{t+1})\;\beta_{t+1}(j)}{\sum_{j=1}^{N}\sum_{i=1}^{N}\alpha_t(i)\;a_{ij}\;b_j(o_{t+1})\;\beta_{t+1}(j)}$$

$$=\frac{P(\bar{O},q_t=i,q_{t+1}=j|\lambda)}{P(\bar{O}|\lambda)}=P(q_t=i,q_{t+1}=j|\bar{O},\lambda)$$

因为待求解的参数必须都要由这两个变量可以计算出来：

待求解变量：

$$\bar{\pi}_i=\gamma_1(i)$$

$$\bar{a}_{ij}=\frac{\sum_{t=1}^{T-1}\varepsilon_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

$$\bar{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j, k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j, k)}$$

$$\bar{\mu}_{jk} = \frac{\sum_{t=1}^{T} [\gamma_t(j, k) \bullet o_t]}{\sum_{t=1}^{T} \gamma_t(j, k)}$$

$$\bar{U}_{jk} = \frac{\sum_{t=1}^{T} [\gamma_t(j, k)(o_t - \mu_{jk})(o_t - \mu_{jk})']}{\sum_{t=1}^{T} \gamma_t(j, k)}$$

Gmm-hmm 算法的伪代码：

1. 初始化 gmm-hmm 中参数
2. 计算构造出来的那两个变量
3. 计算待求参数
4. 循环 2，3 直到收敛

从语音识别角度理解 gmm-hmm 训练过程：

1. 将语音使用 mfcc 抽取特征
2. 初始化 gmm-hmm 参数
3. 根据初始化参数计算构造出来的两个变量
4. 计算带球参数
5. 循环 3，4 直到收敛