

基于 EM 算法的高斯混合模型参数估计

余爱华

(正德职业技术学院, 南京 211106)

摘要: 讨论在一般的混合分布条件下, 用 EM 算法, 在最小熵原理的优化准则下的数据拟合问题。简单推导有限混合高斯分布的 EM 算法, 并针对其收敛速度慢的缺点设计一种有效的选取参数初始值的方法。实验结果表明, 该方法有助于 EM 算法以较快的速度在参数真值附近收敛。

关键词: 混合模型; 极大似然估计; EM 算法

0 引言

EM 算法是一种从“不完全数据”中求解模型参数的极大似然估计的方法, 所谓“不完全数据”一般有两种情况: ①由于观察过程本身的限制或者错误, 造成观察数据成为错漏的不完全数据; ②参数的似然函数直接优化十分困难, 而引入额外的参数 (隐含的或丢失的) 后就比较容易优化。于是定义原始观察数据加上额外数据组成“完全数据”, 原始观察数据自然就成为“不完全数据”。

EM 算法基本原理可以表述如下: 我们可以观察到的数据是 Y , 完全数据 $X=(Y, Z)$, Z 是缺失数据, θ 是模型参数。 θ 关于 Y 的后验分布 $p(\theta|Y)$ 均很复杂, 难以进行各种不同统计计算。假如缺失数据 Z 已知, 则可能得到一个关于 θ 的简单的添加后验分布 $p(\theta|y, z)$, 利用 $p(\theta|y, z)$ 的简单性我们可以进行各种统计计算。然后, 我们可以对 Z 的假定作检查和改进, 如此进行, 我们将一个复杂的极大化或抽样问题转化为一系列简单的极大化或抽样问题。

1 极大似然估计 (MLE)

极大似然估计是一种常用的参数估计方法, 它以观测值出现的概率最大作为准则。

设 x 为连续随机变量, 其分布密度函数为 $p(x|\theta)$,

$\theta=\{\theta_1, \cdots, \theta_M\}$, 这个密度函数由参数 θ 完全决定。已知 N 个观测值 x_1, \cdots, x_N , 假设它们是从分布密度为 $p(x|\theta)$ 的总体中独立抽取的。记 $X=\{x_1, \cdots, x_N\}$, 则:

$$p(X|H)=\prod_{i=1}^N p(x_i|\theta) \doteq L(\theta|X) \quad (1)$$

函数 $L(\theta|X)$ 称为似然函数。当 X 固定时, $L(\theta|X)$ 是 θ 的函数。极大似然参数估计的实质就是求出使 $L(\theta|X)$ 达到极大时 θ 值, 即:

$$\theta=\operatorname{argmin}_{\Theta}(L(\theta|X))$$

其中 Θ 表示参数空间。为了便于求出使 $L(\theta|X)$ 达到极大的 θ , 通常对式 (1) 两边取对数, 即:

$$\ln(L(\theta|X))=\sum_{i=1}^N \ln(p(x_i|\theta)) \quad (2)$$

将式 (2) 分别对 θ_i 求偏导, 令偏导数等于零, 得方程组:

$$\frac{\partial}{\partial \theta_i} \ln(L(\theta|X))=0, i=1, \cdots, M \quad (3)$$

求解方程组 (3), 可以得到极大似然估计值 θ 。

但是, 极大似然估计存在的问题是, 对于许多具体问题不能构造似然函数解析表达式, 或者似然函数的表达式过于复杂而导致求解方程组 (3) 非常困难, 因此须借助于其他方法。下面将要介绍的 EM 算法就是实

收稿日期: 2011-06-09 修稿日期: 2011-07-13

作者简介: 余爱华 (1978-), 女, 福建人, 讲师, 从事领域为计算机软件与理论、计算机网络教学与研究工作

际应用中的一种有效方法。

2 EM 算法

EM 算法是进行极大似然估计的一种有效方法,它主要应用于下面两种非完全数据参数估计:①观测数据不完全,这是由于观测过程的局限性所导致;②似然函数不是解析的,或者似然函数的表达式过于复杂从而导致极大似然函数的传统估计方法失效,第二种情况在模式识别中经常遇到。

假设 X 是服从某一分布的非完全观测数据集,且存在一个完全数据集 $Z=(X, Y)$, 则 Z 的密度函数为:

$$p(z|\theta)=p(x, y|\theta)=p(y|x, \theta)p(x|\theta) \quad (4)$$

从式(4)可以看出,密度函数 $p(z|\theta)$ 是由边沿密度函数 $p(x|\theta)$ 、隐变量 y 的假设、参数 θ 初始估计值以及隐变量与观测变量之间的关系决定。

下面讨论密度函数 $p(z|\theta)$ 的具体形式。

由式(4)给出的密度函数可以定义一个新的似然函数:

$$L(\theta|Z)=L(\theta|X, Y) \doteq p(X, Y|\theta) \quad (5)$$

称此函数为完全数据似然函数。由于隐变量 Y 未知,因此似然函数 $L(\theta|Z)$ 是随机的,且由隐变量 Y 所决定。

EM 算法的第一步 E-step: 即给定观测 X 值和当前参数估计值,计算完全数据对数似然函数 $\log p(X, Y|\theta)$ 关于未知数据 Y 的期望。为此,定义对数似然函数的期望:

$$Q(\theta, \theta^{(i-1)})=E[\log p(X, Y|\theta)|X, \theta^{(i-1)}] \quad (6)$$

其中 $\theta^{(i-1)}$ 为已知的当前参数估计值。

在式(6)中, X 和 $\theta^{(i-1)}$ 为常数, θ 为待优化的参数。 Y 为一随机变量,并假设它服从某一分布 $f(\cdot)$,

$$y \sim f(y|X, \theta^{(i-1)}) \quad (7)$$

因此,式(6)可写为:

$$\begin{aligned} Q(\theta, \theta^{(i-1)}) &= E[\log p(X, Y|\theta)|X, \theta^{(i-1)}] \\ &= \int_{y \in D} \log p(X, y|\theta) f(y|X, \theta^{(i-1)}) dy \end{aligned} \quad (8)$$

其中 $f(y|X, \theta^{(i-1)})$ 是不可观测数据 Y 的边沿分布密度函数,并且依赖于观测数据 X 和当前参数 $\theta^{(i-1)}$, D 为 y 的取值空间。在一些特殊情况下,边沿分布 $f(y|X, \theta^{(i-1)})$ 是 X 和 $\theta^{(i-1)}$ 的简单解析函数,但通常这个函数很难得到。

由乘法公式,得:

$$f(y, X|\theta^{(i-1)})=f(y|X, \theta^{(i-1)})f(X|\theta^{(i-1)}) \quad (9)$$

由于因子 $f(X|\theta^{(i-1)})$ 与 θ 无关,所以在实际问题处理中,用 $f(y, X|\theta^{(i-1)})$ 代替 $f(y|X, \theta^{(i-1)})$ 不影响式(8)中似然函数的最优化。

定义二元函数:

$$h(\theta, Y) \doteq \log L(\theta|X, Y) \quad (10)$$

其中 y 服从某一分布 $f_Y(y)$ 。那么:

$$E_Y[h(\theta, Y)] = \int_Y h(\theta, Y) f_Y(y) dy \doteq q(\theta) \quad (11)$$

从式(11)可知 $E_Y[h(\theta, Y)]$ 是关于 θ 的函数,以通过简单的最优化方法得到参数 θ 的估计值 $\hat{\theta}$ 。期望值 $E_Y[h(\theta, Y)]$ 的计算也就是 EM 算法的 E-step。

EM 算法的第二步 M-step: 最大化期望值 $Q(\theta, \theta^{(i-1)})$, 即找到一个 θ^i , 满足:

$$\theta^i = \underset{\theta \in \Theta}{\operatorname{argmin}} (Q(\theta, \theta^{(i-1)})) \quad (12)$$

其中 Θ 代表参数空间。

EM 算法是一种迭代算法,每一步迭代都能保证似然函数值增加,并且收敛到一个局部极大值。

3 混合密度参数估计

设样本集 $X=\{x_1, \dots, x_N\}$, 每个样本的类别未知,但知道它们是从混合密度为 $p(x|\theta)$ 的总体中独立抽取的,

$$p(x|\theta) = \sum_{i=1}^M \alpha_i p_i(x|\theta_i) \quad (13)$$

其中参数 $\theta=(\alpha_1, \dots, \alpha_M, \theta_1, \dots, \theta_M)$, 且 $\sum_{i=1}^M \alpha_i=1$ 。观测

样本的对数似然函数表达式为: $\log(L(\theta|X)) = \log \prod_{i=1}^N p(x_i|\theta)$

$$\begin{aligned} (x_i|\theta) &= \sum_{i=1}^N \log p(x_i|\theta) \\ &= \sum_{i=1}^N \log \left[\sum_{j=1}^M \alpha_j p_j(x_i|\theta_j) \right] \end{aligned} \quad (14)$$

极大似然估计就是求使公式(14)中的函数 $L(\theta|X)$ 达到极大时参数 θ 的值。最大化这个似然函数是很困难的,因为它包含和式取对数。我们已经提到,通过假设隐变量的存在可以大大简化似然函数方程的求解。下面讨论基于 EM 算法的混合密度参数估计问题。

假设 X 为非完全数据, 并且存在一个不可观测的数据 $Y = \{y_i\}_{i=1}^N$, 它的取值表示某一个观测数据来自某一类, 由此隐变量假设可知, $y_i \in \{1, \dots, M\}$, $y_i = k$ 表示第 i 个观测数据属于第 k 类。如果知道 y 的取值, 那么:

$$\begin{aligned} \log(p(X, Y|\theta)) &= \log \prod_{i=1}^N p(x_i, y_i|\theta) \\ &= \sum_{i=1}^N \log(p(x_i, y_i|\theta_{y_i})) \\ &= \sum_{i=1}^N \log(p(x_i, y_i|\theta_{y_i}) p_{y_i}(y_i|\theta_{y_i})) \\ &= \sum_{i=1}^N \log(\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})) \end{aligned} \quad (15)$$

如果知道类条件密度函数形式, 那么估计问题可以得到解决。但是目前这个问题还不能解决, 问题就在于不知道 y 的取值。下面确定 y 分布。设初始的参数估计值为:

$$\theta^g = (\alpha_1^g, \dots, \alpha_M^g, \theta_1^g, \dots, \theta_M^g) \quad (16)$$

由贝叶斯公式, 得:

$$p(y_i|x_i, \theta^g) = \frac{\alpha_{y_i}^g p_{y_i}(x_i|\theta_{y_i}^g)}{p(x_i|\theta^g)} = \frac{\alpha_{y_i}^g p_{y_i}(x_i|\theta_{y_i}^g)}{\sum_{k=1}^M \alpha_k^g p_k(x_i|\theta_k^g)} \quad (17)$$

进而可得:

$$p(y|X, \theta^g) = \prod_{i=1}^N p(y_i|x_i, \theta^g) \quad (18)$$

其中 $Y = \{y_1, \dots, y_N\}$ 是隐变量 Y 的一次样本实现, 且独立同分布。由此可知, 如果给出参数初始估计值, 并且假设存在隐变量 y , 由式(18)就得到 y 的边沿分布密度函数。由式(8)、(15)、(18)可知, 完全数据的似然函数为:

$$\begin{aligned} Q(\theta, \theta^g) &= \sum_{y \in D} \log(L(\theta|X, y)) p(y|X, \theta^g) \\ &= \sum_{y \in D} \sum_{i=1}^N \log[\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})] \prod_{j=1}^N p(y_j|x_j, \theta^g) \\ &= \sum_{y=1}^M \dots \sum_{y_N=1}^M \sum_{i=1}^N \log[\alpha_{y_i} p_{y_i}(x_i|\theta_{y_i})] \prod_{j=1}^N p(y_j|x_j, \theta^g) \\ &= \sum_{y=1}^M \dots \sum_{y_N=1}^M \sum_{i=1}^N \sum_{l=1}^M \delta_{l, y_i} \log[\alpha_l p_l(x_i|\theta_l)] \prod_{j=1}^N p(y_j|x_j, \theta^g) \\ &= \sum_{i=1}^N \sum_{l=1}^M \log[\alpha_l p_l(x_i|\theta_l)] \sum_{y=1}^M \dots \sum_{y_N=1}^M \sum_{j=1}^N \delta_{l, y_j} \prod_{j=1}^N p(y_j|x_j, \theta^g) \end{aligned} \quad (19)$$

对于 $l \in \{1, \dots, M\}$,

$$\begin{aligned} &\sum_{y=1}^M \dots \sum_{y_N=1}^M \delta_{l, y_i} \prod_{j=1}^N p(y_j|x_j, \theta^g) \\ &= \left[\sum_{y=1}^M \dots \sum_{y_N=1}^M \sum_{j=1}^N \prod_{j=1, j \neq i}^N p(y_j|x_j, \theta^g) \right] p(l|x_i, \theta^g) \\ &= \prod_{j=1, j \neq i}^N \left[\sum_{y=1}^M p(y_j|x_j, \theta^g) \right] p(l|x_i, \theta^g) \\ &= p(l|x_i, \theta^g) \end{aligned} \quad (20)$$

由式(19)和式(20), 得:

$$\begin{aligned} Q(\theta, \theta^g) &= \sum_{l=1}^M \sum_{i=1}^N \log[\alpha_l p_l(x_i|\theta_l)] p(l|x_i, \theta^g) \\ &= \sum_{l=1}^M \sum_{i=1}^N \log(\alpha_l) p(l|x_i, \theta^g) + \sum_{l=1}^M \sum_{i=1}^N \log(p_l(x_i|\theta_l)) p(l|x_i, \theta^g) \\ &= I_1 + I_2 \end{aligned} \quad (21)$$

由式(21)明显看到, 第一项 I_1 只含有参数 α_l , 第二项 I_2 只含有参数 θ_l , 因此最大化似然函数 $Q(\theta, \theta^g)$ 可分别最大化式(20)中的两项 I_1 和 I_2 。

为了得到参数估计 $\hat{\alpha}_l$, 只需最大化式(20)中的项 I_1 , 这是一个条件极值问题, 因此需要引入一个拉格朗日乘子 λ , 解方程:

$$\frac{\partial}{\partial \alpha_l} \left[\sum_{i=1}^M \sum_{i=1}^N \log(\alpha_l) p(l|x_i, \theta^g) + \lambda \left(\sum_{l=1}^M \alpha_l - 1 \right) \right] = 0, \quad l=1, \dots, M \quad (22)$$

得:

$$\hat{\alpha}_l = \frac{1}{N} \sum_{i=1}^N p(l|x_i, \theta^g), \quad l=1, \dots, M \quad (23)$$

为了得到参数估计 $\hat{\theta}_l$, 须知道 x 的条件概率密度函数形式。这里假设 x 服从高斯分布, 均值为 μ_l , 方差为 $\sum l$, 即 $\theta_l = (\mu_l, \sum l)$ 。这样有,

$$\begin{aligned} &p_l(x|\mu_l, \sum l) = \\ &= \frac{1}{(2\pi)^{d/2} (\det \sum l)^{1/2}} \exp \left[-\frac{1}{2} (x_l - \mu_l)^T \left(\sum l \right)^{-1} (x_l - \mu_l) \right] \end{aligned} \quad (24)$$

忽略与 θ_l 无关的项, 将式(24)代入式(3), 得:

$$\sum_{l=1}^M \sum_{i=1}^N \log(p_l(x_i|\theta_l)) p_l(l|x_i, \theta^g) =$$

$$\sum_{i=1}^M \sum_{l=1}^N \left[-\frac{1}{2} (\det \sum l) - \frac{1}{2} (x_i - \mu_l)^T \sum_{l=1}^{-1} (x_i - \mu_l) \right] p(l|x_i, \theta^e) \quad (25)$$

式(25)两边对 μ_l 求导,并令导函数等于零,得:

$$\sum_{i=1}^N \sum_{l=1}^{-1} (x_i - \mu_l) p(l|x_i, \theta^e) = 0 \quad (26)$$

进而,得:

$$\hat{\mu}_l = \frac{\sum_{i=1}^N x_i p(l|x_i, \theta^e)}{\sum_{i=1}^N p(l|x_i, \theta^e)} \quad (27)$$

式(25)两边对 $\sum_{l=1}^{-1}$ 求导,并令导函数等于零,得:

$$\sum_{i=1}^N p(l|x_i, \theta^e) (\sum_{l=1}^{-1} l - N_{l,i}) = 0 \quad (28)$$

其中 $N_{l,i} = (x_i - \mu_l)(x_i - \mu_l)^T$ 。进而,得:

$$\hat{\sum}_{l=1}^{-1} = \frac{\sum_{i=1}^N p(l|x_i, \theta^e) (x_i - \mu_l)(x_i - \mu_l)^T}{\sum_{i=1}^N p(l|x_i, \theta^e)} \quad (29)$$

综上所述,参数的估计方程为式(23)、式(27)和式(29)。

4 在 Gaussian 混合模型中应用 EM 算法进行参数估计

Gaussian 混合模型结构就是利用一定数量的高斯函数

通过线性混合来逼近某变量的概率分布的方法,现在考虑将 EM 算法应用到混合分布的情形。下面我们以一个二维混合正态分布为例在 MatLab 中做模拟实验。

本实验中,混合分布样本数据的产生方法如下:

设 X 是属于给定的混合密度参数族 P 的一个随机变量,它具有形如式(4)的概率密度函数。设 x 是对 X 的一个观察值, Z 是 x 的分类随机变量,其取值 $Z=i$ ($i=1, 2, \dots, m$);那么在给定 $Z=i$ 的条件下, $X=x$ 的条件概率为 $p_i(x|\theta_i)$ ($i=1, 2, \dots, m$),而其无条件概率为 $p(x|\theta)$ 。

X 的一个观察样本产生过程如下:

(1)按分布 $P(Z=i)=\lambda_i$ ($i=1, 2, \dots, m$) 产生 Z 的一个观察值:首先产生一个 $[0,1]$ 区间的随机变量 r ,若 $\sum_{k=1}^{i-1}$

$\lambda_k \leq r \leq \sum_{k=1}^i \lambda_k$, 则 $Z=i$;

(2)若 $Z=i$,则按 X 的第 i 分支对应的分布产生一个观察值 $x, i=1, 2, \dots, m$ 。

重复(1)、(2)这一过程 n 次,可得到关于 x 的一组容量为 n 的简单随机样本观测值。

实验步骤如下:

(1)利用上述方法产生一个二维混合正态分布的 2000 个样本点;

(2)确定算法需要的参数初始值:混合比例 α 采用平均分配策略,各分支的期望的元素是介于各样本元

表 1 Gaussian 混合模型参数真实值与估计值比较(一)

参数	真实值	初始值	估计值
α	(0.66 0.22 0.12)'	(1/3 1/3 1/3)'	(0.6667 0.2742 0.0591)'
μ_1	(1.5 0.1)'	(3.5181 7.5751)'	(1.5147 0.09214)'
Σ_1	$\begin{bmatrix} 3.2 & -0.1 \\ -0.1 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 8.1621 & 2.2129 \\ 2.2129 & 7.0368 \end{bmatrix}$	$\begin{bmatrix} 3.2837 & -0.1221 \\ -0.1221 & 0.4180 \end{bmatrix}$
μ_2	(7.2 7.2)'	(9.4981 5.5791)'	(7.0375 7.1076)'
Σ_2	$\begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}$	$\begin{bmatrix} 5.2206 & 2.1335 \\ 2.1335 & 2.2804 \end{bmatrix}$	$\begin{bmatrix} 0.4558 & -0.1326 \\ -0.1326 & 0.6744 \end{bmatrix}$
μ_3	(6.4 6.9)'	(0.1421 7.9621)'	(6.4259 6.9406)'
Σ_3	$\begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 4.4964 & -2.3566 \\ -2.3566 & 3.5572 \end{bmatrix}$	$\begin{bmatrix} 0.1194 & 0.1397 \\ 0.1397 & 0.1700 \end{bmatrix}$

表 2 高斯混合模型参数真实值与估计值比较(二)

参数	真实值	初始值	估计值
α	$(0.66 \ 0.22 \ 0.12)^T$	$(1/3 \ 1/3 \ 1/3)^T$	$(0.6667 \ 0.2742 \ 0.0591)^T$
μ_1	$(1.5 \ 0.1)^T$	$(3.5181 \ 7.5751)^T$	$(1.5147 \ 0.09214)^T$
Σ_1	$\begin{bmatrix} 3.2 & -0.1 \\ -0.1 & 0.5 \end{bmatrix}$	$\begin{bmatrix} 8.1621 & 2.2129 \\ 2.2129 & 7.0368 \end{bmatrix}$	$\begin{bmatrix} 3.2837 & -0.1221 \\ -0.1221 & 0.4180 \end{bmatrix}$
μ_2	$(7.2 \ 7.2)^T$	$(9.4981 \ 5.5791)^T$	$(7.0375 \ 7.1076)^T$
Σ_2	$\begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}$	$\begin{bmatrix} 5.2206 & 2.1335 \\ 2.1335 & 2.2804 \end{bmatrix}$	$\begin{bmatrix} 0.4558 & -0.1326 \\ -0.1326 & 0.6744 \end{bmatrix}$
μ_3	$(6.4 \ 6.9)^T$	$(0.1421 \ 7.9621)^T$	$(6.4259 \ 6.9406)^T$
Σ_3	$\begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}$	$\begin{bmatrix} 4.4964 & -2.3566 \\ -2.3566 & 3.5572 \end{bmatrix}$	$\begin{bmatrix} 0.1194 & 0.1397 \\ 0.1397 & 0.1700 \end{bmatrix}$

素的最大值和最小值之间随机产生,各分支的协方差矩阵采用对称矩阵,产生的参数初始值见表 1 所示。

在本实验中 EM 算法迭代终止条件为:连续两次所得参数估计值满足 $\left| \frac{\log p(x|\Theta^t) - \log p(x|\Theta^{t+1})}{\log p(x|\Theta^t)} \right| \leq 0.00001$ 时迭代停止,本次试验经过 81 次迭代后程序终止。结果见表 1 所示。

再取一组远离参数真实值的初值进行迭代,经过 140 次迭代后程序终止,迭代的结果如表 2 所示。

从所做的大量数值计算来看,用 EM 算法对混合正态模型进行密度参数估计,能收敛到各参数的极大似然估计量,并且样本越大,估计结果越接近参数真值。

5 K 均值算法与 EM 算法的结合

EM 算法的结果可以保证收敛到似然函数的稳定点,EM 算法能得到了广泛的应用也正是因为它的可行性和稳定性。初始值对 EM 算法的迭代速度影响很大,而且,估计模型参数极大似然的 EM 算法本身是一个局部最优搜索算法,常常会陷入局部最优解,这就使得它的搜索结果对初始化敏感。而一般采用的随机初始化方法来初始化 EM 算法过于简单,实用的效果也不太好。

通过试验研究表明 EM 算法的时间复杂度高于 K 均值算法,且随着聚类个数的增加,EM 算法和 K 均值算法的时间耗比值呈明显上升趋势。但是,EM 算法相比 K 均值算法,聚类的正确性和稳定性更好。

基于 HCM 算法和 EM 算法各自的优缺点,我们可以把 K 均值算法的结果作为 EM 算法的初始值,来提高 EM 算法的迭代速度。在本试验中把 K 均值算法的结果作为 EM 算法的初始值: $\alpha=(0.6636 \ 0.2221 \ 0.113 \ 3)$; $\mu_1=(1.5130 \ 0.09240)^T$; $\Sigma_1=[3.2922 \ -0.1811; -0.1811 \ 0.4610]$; $\mu_2=(7.0355 \ 7.1026)^T$; $\Sigma_2=[0.4422 \ -0.1806; -0.1806 \ 0.6634]$; $\mu_3=(6.4233 \ 6.9426)^T$; $\Sigma_3=[0.1205 \ 0.1321; 0.1321 \ 0.3755]$;该结果和真值已经很接近,最终用了 5 次迭代可达到上述收敛结果。

从上面的计算中我们可以发现采用 K 均值算法对 EM 算法进行初始化,明显提高了 EM 算法的迭代速度。然而 K 均值算法本质上是一种局部搜索技术,它采用了所谓的梯度下降法来寻找最优解。因此容易陷入局部最小值,而得不到全局最优解,并且从某种程度加剧了 EM 算法局部收敛的缺陷。

参考文献

- [1]张士峰. 混合正态分布参数极大似然估计的 EM 算法. 飞行器测控学报, 2004,23(4):20~25
- [2]刘宝亮,温艳清. 有限混合模型的参数估计. 山东理工大学学报(自然科学版),2008,22(5):108~110
- [3]彭锋. MatLab 6.x 符号运算及其应用. 北京:机械工业出版社,2003
- [4]楼顺天,陈生潭,雷虎民等. MatLab 5.x 程序设计语言. 西安:西安电子科技大学出版社,2000

(下转第 31 页)

Problems Existed on Computer Basis Teaching of Independent Colleges and Its Solution

ZHANG Yuan^{1,2} , HUANG Lei²

(1. College of Software, Tongji University Shanghai 200092;

2. Yinchuan College, China University of Mining Technology, Yinchuan 750011)

Abstract: The college students' computer level is one of the means students' comprehensive ability, and to independent college students employment is very important. According to years of teaching experience, proposes independent college computer problems in basic education and how to solve these problems, puts forward its own proposals.

Keywords: Independent Colleges; Computer Basis; Solution

(上接第 7 页)

Design of Gaussian Mixture Model Parameter Based on EM Algorithm

YU Ai-hua

(Zhengde Vocational and Technical College, Nanjing 211106)

Abstract: Mainly deals with the data mining problem in the optimization principle of minimum entropy with expectation maximization algorithm in the mixed-conditional circumstances. The EM iterative formulas for Gaussian mixtures are derived, and proposes an efficient initialization method for these iterative formulas to accelerate the EM algorithm's speed of convergence. Experience of numerical simulation shows that the proposed initial method contributes to accelerate the algorithm's speed of convergence, the more accurate results can be obtained.

Keywords: Mixture Model; Maximum Likelihood Estimation; EM Algorithm