

修正 Gibbs 采样的有限混合模型无监督学习算法

刘伟峰, 韩崇昭, 石勇

(西安交通大学电子与信息工程学院, 710049, 西安)

摘要: 针对传统有限混合模型无监督学习算法不能处理参数维数变化的问题,提出了一种基于修正 Gibbs 采样的无监督学习算法. 该算法的关键是,在每一次完全采样之后引入分布元的合并和剔除技术,即将利用均值、协方差矩阵差值的 2 范数作为合并的判断准则,最小且小于阈值的分布元权重作为剔除规则. 仿真实验表明,所提算法对于参数初值的选择是不敏感的,对于分布元个数的先验信息要求得更少,它不仅可以处理维数变化问题,而且不必计算跳变概率,同时能够很好地估计出分布元个数及其参数.

关键词: 无监督学习;有限混合模型;参数维数变化;跳变;分布元管理

中图分类号: TP391 **文献标志码:** A **文章编号:** 0253-987X(2009)02-0015-05

Unsupervised Learning for Finite Mixture Models Via Modified Gibbs Sampling

LIU Weifeng, HAN Chongzhao, SHI Yong

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Since the conventional algorithm can not deal with the variable parameter dimension in the unsupervised learning of finite mixture models (FMM), an unsupervised learning algorithm based on the modified Gibbs sampling scheme is proposed. The key for the proposed algorithm is to adopt the component management techniques that include component combination and elimination after each complete iterative step. The 2-norm of the differences in the mean and covariance are used for the component combination rule, and the component elimination rule is that the component that has the least weight and is less than certain threshold will be discarded. Simulation results show that the proposed algorithm is robust for the parameter initialization and requires fewer prior information for the number of components. The proposed algorithm can deal with the variable dimension and avoid the calculation of the jump probability. Moreover, it can estimate the number of the components and parameters effectively.

Keywords: unsupervised learning; finite mixture model; parameter dimension change; jump; component management

有限混合模型(FMM)^[1]可解决半参数建模问题,在模式识别^[2]、图像处理^[3]、人工智能、天文学等方面得到了广泛的应用.

对于一组异质观察数据,即属于不同随机源的数据,每一个随机源称为一个分布元. FMM 无监督学习的目标是估计这些分布元的参数以及混合权

重. 目前,主要有两种算法解决这个问题,一是期望最大化(EM)算法,另一是 Bayes 随机采样算法. EM 算法用来解决极大似然准则下的参数估计问题,时至今日它仍然是 FMM 参数估计问题的标准算法之一. 但是,EM 算法有两个不足之处,一是参数估计结果依赖于初值的选择,二是容易陷入局部

极值点^[4-6]. Bayes 随机采样方法主要解决后验分布估计的问题,而大部分基于 Bayes 随机采样方法的混合估计问题均假设分布元个数已知且固定^[7-8],在不确定分布元个数问题上也有相应的解决方法,如可逆跳的 MCMC 方法(RJMCMC)^[9],但基于 RJMCMC 方法的 FMM 参数估计方法^[10]需要在高维空间采样来计算接受概率.

综上,本文提出了一种基于修正 Gibbs 采样的算法.

1 问题描述

1.1 有限混合模型

具有 k -分布元的 FMM 可以表示为

$$f(\mathbf{y}_i | \theta) = \sum_{j=1}^k \omega_j f_j(\mathbf{y}_i | \theta_j) \quad (1)$$

式中: $\mathbf{y}_i \in \mathbf{R}^d$ 是观测数据; θ_j 是第 j 个分布元的参数; ω_j 是混合权重并且满足 $\omega_j \geq 0$, $\sum_{j=1}^k \omega_j = 1$. 假设有 n 个独立观察数据 $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, 那么参数似然函数可表示为

$$L(\mathbf{y} | \theta) = \prod_{i=1}^n \sum_{j=1}^k \omega_j f_j(\mathbf{y}_i | \theta_j) \quad (2)$$

FMM 参数 $\omega = \{\omega_1, \dots, \omega_k\}$, $\theta = \{\theta_1, \dots, \theta_k\}$. 根据式(2), FMM 的参数后验分布可表示为

$$p(\omega, \theta | \mathbf{y}) \propto L(\mathbf{y} | \omega, \theta) p(\omega, \theta) \quad (3)$$

对每一个观测数据 $\mathbf{y}_i$, 相应的缺失变量是一个 k 维向量, $\mathbf{e}_i = \{e_{i,1}, \dots, e_{i,k}\}$, $e_{i,j} \in \{0, 1\}$, 并且满足 $\sum_{j=1}^k e_{i,j} = 1$, 其中 $e_{i,j} = 1$ 表示数据 $\mathbf{y}_i \in f_j(\cdot)$. 本文假设混合分布是高斯混合的, 这样相应的参数集就是 $\theta_j = (\boldsymbol{\mu}_j, \boldsymbol{\sigma}_j)$.

1.2 分布元个数估计

通过最小化费用函数, EM 算法可以估计的分布元个数

$$\hat{k} = \arg \min_k c(\theta, k) \quad (4)$$

式中: $c(\theta, k) = -\ln L(\mathbf{y} | \theta) + L_M(k)$, $\ln L(\mathbf{y} | \theta)$ 是自然对数似然函数, $L_M(k)$ 是信息长度. 信息长度的类型包括: Schwarz Bayes 信息准则(BIC), 最小信息长度(MML), Akaike 信息准则(AIC), 等. 一般来说, 费用函数最小并不能保证似然函数最小, 它仅是似然函数与信息长度之间的一种折中.

当分布元个数不确定时, 参数的维数也是不确定的. RJMCMC 算法基于的核心思想是 Markov 转

移核平衡性条件, 其接受-拒绝率可以表示成测度比的形式, 即

$$\gamma_m(\mathbf{x}, \mathbf{x}') = \min \left\{ 1, \frac{\pi(\mathrm{d}\mathbf{x}') q_m(\mathbf{x}', \mathrm{d}\mathbf{x})}{\pi(\mathrm{d}\mathbf{x}) q_m(\mathbf{x}, \mathrm{d}\mathbf{x}')} \right\} \quad (5)$$

式中: $q_m(\cdot)$ 是跳变类型的概率; $\pi(\cdot)$ 是后验分布; $\mathbf{x}'$ 是采样点; $\mathrm{d}(\cdot)$ 是微分算子. Recharadson 进一步给出了全 Bayes 的混合估计方法^[10], 而“全”指的是联合估计分布元个数和参数.

1.3 Gibbs 采样算法

假设 $\varphi(\mathbf{x}) = \varphi(\mathbf{x}_1, \dots, \mathbf{x}_m)$ 指的是 m 维联合分布, $\varphi(\mathbf{x}_i | \mathbf{x}_{-i})$ 是全条件分布, 其中 $\mathbf{x}_{-i} \triangleq \{\mathbf{x}_j, j \neq i\}$. 那么, 传统的 Gibbs 采样算法描述如下.

(1) 首先假设初始值是 $\mathbf{x}^{(0)} = (\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_m^{(0)})$, 然后从条件密度 $\varphi(\mathbf{x}_i | \mathbf{x}_{-i})$ 采样.

(2) $\mathbf{x}_1^{(1)}$ 从 $\varphi(\mathbf{x}_1 | \mathbf{x}_{-1}^{(0)})$ 中采样, $\mathbf{x}_2^{(1)}$ 从 $\varphi(\mathbf{x}_2 | \mathbf{x}_{-2}^{(0)})$ 中采样, 依次类推, $\mathbf{x}_m^{(1)}$ 从 $\varphi(\mathbf{x}_m | \mathbf{x}_{-m}^{(0)})$ 中采样.

(3) 完成一次从 $\mathbf{x}^{(0)}$ 到 $\mathbf{x}^{(1)}$ 的转移.

(4) 重复以上各步直到 $\{\mathbf{x}^{(t)}\}$ 达到稳态.

基于以上迭代过程, 可以获得一条 Markov 链.

2 修正的 Gibbs 采样算法

2.1 FMM 参数后验分布

基于以下参数分布, 可以利用 Gibbs 采样算法.

(1) 混合权重 $\{\omega_j\}$. 它满足 Dirichelet 分布, 即

$$(\omega_1, \dots, \omega_k) \sim D(a_1 + l_1, \dots, a_k + l_k) \quad (6)$$

式中: $a_j > 0$ 是常数; l_j 是属于第 j 个分布元的观测数据个数.

(2) 缺失变量 $\{e_{i,j}\}$. 它可以根据以下的 Bayes 公式估计, 即

$$\hat{e}_{i,j} = (\omega_j f_j(\mathbf{y}_i | \theta_j)) / \sum_{j=1}^k \omega_j f_j(\mathbf{y}_i | \theta_j) \quad (7)$$

$$l_j = \sum_{i=1}^n \hat{e}_{i,j} \quad (8)$$

(3) 方差 $\{\sigma_j\}$. 它服从逆伽玛分布, 即

$$\boldsymbol{\sigma}_j^{-2} \sim \Gamma(\alpha_0 + l_j/M_0, \boldsymbol{\beta}_0 + \boldsymbol{\Sigma}_j/N_0)$$

$$\boldsymbol{\Sigma}_j = \left(\sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}_j)(\mathbf{y}_i - \boldsymbol{\mu}_j)^T \hat{e}_{i,j} \right) / \sum_{i=1}^n \hat{e}_{i,j} \quad (9)$$

式中: $\alpha_0, \boldsymbol{\beta}_0$ 是正常数; $M_0, N_0 > 0$ 起调节作用.

(4) 均值 $\{\boldsymbol{\mu}_j\}$. 它满足参数为 $\boldsymbol{\xi}_j, \boldsymbol{\sigma}_j$ 的高斯分布, 可以从下式中采样

$$\boldsymbol{\mu}_j = N(\boldsymbol{\xi}_j, \boldsymbol{\sigma}_j^2) \quad (10)$$

$$\boldsymbol{\xi}_j = \left(\sum_{i=1}^n \mathbf{y}_i \hat{e}_{i,j} \right) / \sum_{i=1}^n \hat{e}_{i,j} \quad (11)$$

2.2 分布元管理技术

当获得了完全的采样点之后, 便可应用分布元

的管理技术.

2.2.1 分布元合并 如果2个分布元具有相同的方差(协方差)和均值,那么可以直接进行合并.通常情况下,这个条件很难满足.对于方差相同、均值不同的分布元,可按照椭圆距离准则合并,即

$$(\boldsymbol{\mu}_{j1} - \boldsymbol{\mu}_{j2})^T \mathbf{P}_j^{-1} (\boldsymbol{\mu}_{j1} - \boldsymbol{\mu}_{j2}) \leq \lambda_{\text{threshold}} \quad (12)$$

式中: $\lambda_{\text{threshold}}$ 是阈值.但是,这种假设也难以满足,更何况协方差矩阵相等,因此可取其次,即选择大权重方差法,即

$$(\boldsymbol{\mu}_{j1} - \boldsymbol{\mu}_{j2})^T \mathbf{P}_{\max_j(\omega_{j1}, \omega_{j2})}^{-1} (\boldsymbol{\mu}_{j1} - \boldsymbol{\mu}_{j2}) \leq \lambda_{\text{threshold}} \quad (13)$$

这个方法的缺点是,当2个高斯元均值接近时,式(13)会接近0(见图1),而实际上这2个高斯元差别很大.因此,可同时利用均值、协方差矩阵差值的2范数作为合并判断准则,即

$$\|\boldsymbol{\mu}_{j1} - \boldsymbol{\mu}_{j2}\|_2 \leq \lambda_\mu, \quad \|\boldsymbol{\sigma}_{j1} - \boldsymbol{\sigma}_{j2}\|_2 \leq \lambda_\sigma \quad (14)$$

如果满足这2个条件,则合并这2个分布元.文献[10]给出如下的合并方法

$$\omega_j^* = \omega_{j1} + \omega_{j2} \quad (15)$$

$$\omega_j^* \boldsymbol{\mu}_j^* = \omega_{j1} \boldsymbol{\mu}_{j1} + \omega_{j2} \boldsymbol{\mu}_{j2} \quad (16)$$

$$\omega_j^* (\boldsymbol{\mu}_j^{*2} + \boldsymbol{\sigma}_j^{*2}) = \omega_{j1} (\boldsymbol{\mu}_{j1}^2 + \boldsymbol{\sigma}_{j1}^2) + \omega_{j2} (\boldsymbol{\mu}_{j2}^2 + \boldsymbol{\sigma}_{j2}^2) \quad (17)$$

对多个分布元的合并,可建立分布元合并矩阵

$$\mathbf{C} = \begin{bmatrix} 1 & c_{1,2} & \cdots & c_{1,k} \\ 0 & 1 & \cdots & c_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (18)$$

式中: $c_{j,j}=1$ 且

$$c_{p,q} = \begin{cases} 1 & (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q)^T \mathbf{P}_{\max}^{-1} (\boldsymbol{\mu}_p - \boldsymbol{\mu}_q) \leq \lambda_{\text{threshold}} \\ 0 & \text{其他} \end{cases}$$

其中 $p \leq q, p=1, \dots, k-1; q=2, \dots, k$. 所有的分布元可按照列来合并,合并规则为

$$\omega_j^* = c_{j,j} \omega_j + c_{j,j+1} \omega_{j+1} + \cdots + c_{j,k} \omega_k \quad (19)$$

$$\omega_j^* \boldsymbol{\mu}_j^* = c_{j,j} \omega_j \boldsymbol{\mu}_{jj} + \cdots + c_{j,k} \omega_k \boldsymbol{\mu}_{jk} \quad (20)$$

$$\omega_j^* (\boldsymbol{\mu}_j^{*2} + \boldsymbol{\sigma}_j^{*2}) = c_{j,j} \omega_j (\boldsymbol{\mu}_{jj}^2 + \boldsymbol{\sigma}_{jj}^2) + \cdots + c_{j,k} \omega_k (\boldsymbol{\mu}_{jk}^2 + \boldsymbol{\sigma}_{jk}^2) \quad (21)$$

合并完成后,合并各项被删除,并用新的合并分布元取代.

2.2.2 分布元剔除 如果分布元的权重小于阈值 λ_ω ,那么该分布元可被剔除.本文的剔除规则为

$$\omega_j^{\min} = \min_j (\omega_j) \delta(\omega_j < \lambda_\omega) \quad (22)$$

也就是说权重小于阈值,分布元被剔除.

事实上,在每一步分量的采样过程中,本文的修

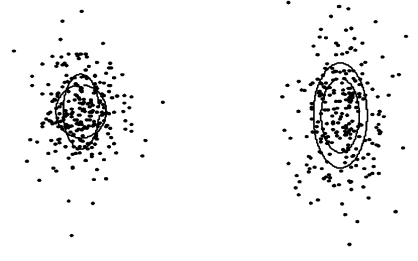


图1 相同均值、不同协方差的高斯元分布

正 Gibbs 采样算法的参数维数并没有发生变化,这意味着不必考虑 RJMCMC 中遇到的维数变化问题,因为在新的完全采样点获得之前,分布元的合并、剔除已经完成,此时分布元的参数维数是定值.

3 算法执行

3.1 参数初值选择

(1)分布元参数.假设观测数据是 $\mathbf{y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{y}_i \in \mathbf{R}^d$,那么数据空间的大小可以根据各个分量的最大、最小值所张成的空间来确定,即

$$\mathbf{V}_{\max}^l = \max\{\mathbf{y}_1^l, \dots, \mathbf{y}_n^l\} \quad (23)$$

$$\mathbf{V}_{\min}^l = \min\{\mathbf{y}_1^l, \dots, \mathbf{y}_n^l\} \quad l = 1, \dots, d \quad (24)$$

这样的数据空间为

$$\mathbf{V} = [\mathbf{V}_{\min}, \mathbf{V}_{\max}] \quad (25)$$

而均值 $\{\boldsymbol{\mu}_j\}$ 的初值是在该数据空间随机采样选择的,不过均值初值最好是均匀分布在数据空间中.对于协方差 $\{\boldsymbol{\sigma}_j\}$,建议其初值取等方差,当然也可以随机取值,在本文的算例3中就是随机取值的.

(2)分布元个数.选择一个适当大的值 $k_{\max}$ 作为分布元个数的初值,惟一的条件是它需大于实际分布元个数.如果分布元个数过大,虽然不会影响最后的结果,但更新参数会增加,计算速度就会大大降低.反之,如果它低于实际数目,则估计精度不高.

3.2 阈值选择

(1)分布参数.对于分布元参数 $\theta = \{\theta_1, \dots, \theta_k\}$,可利用2范数通过判断参数之间的接近程度来求取,即

$$d(\theta_{j1}, \theta_{j2}) = \|\theta_{j1} - \theta_{j2}\|_2 \quad (26)$$

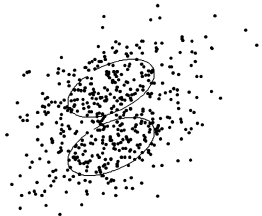
对于高斯分布参数,可以利用式(14)来计算.需要指出的是,这2个条件必须同时满足,其原因除了图1的情况外,另一种情况就是方差相同、均值不同(见图2).本文给出的方法对这2种情况均适用.

均值阈值为

$$\lambda_\mu = \frac{c(k) \max_j \{\|\boldsymbol{\mu}_j\|_2\}}{k} \quad (27)$$

式中: $0 < c(k) \leq 1$ 是一个取决于分布元的常数,分布元越分散,取值越大,相反则取值越小; $\max_j \{ \|\boldsymbol{\mu}_j\|_2 \}$ 表示最大均值范围,其与 k 的比值表示高斯元均值的

图 2 相同协方差、不同均值的高斯元分布



方差阈值为

$$\lambda_{\Sigma} = s(k) \|\boldsymbol{\Sigma}_{\arg \max_j \{ \det \mathbf{P}_j \}}\|_2 \tag{28}$$

或者
$$\lambda_{\sigma} = s(k) \boldsymbol{\sigma}_{\arg \max_j \{ \sigma_j \}} \tag{29}$$

式中: $0 < s(k) \leq 1$ 是系数. 方差阈值应选择行列式最大的协方差作为参考.

(2)权重阈值. 可根据式(22)、式(23)确定权重阈值. 如果权重阈值选择得当,分布元可以快速地剔除,所以需精细地调节这 2 个公式中的参数.

4 算 例

算例 1 该算例取自文献[10]. 数据分布:星系分布、核酸分布和血液中的酶分布为如图 3~图 5 所示. 初始的分布元数选择 $k_{\max} = 12$. 依据本文算法,最终的分布元个数分别为 8、5、4.

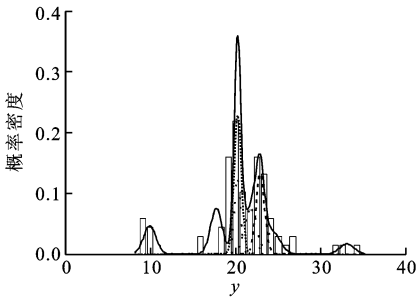


图 3 星系分布

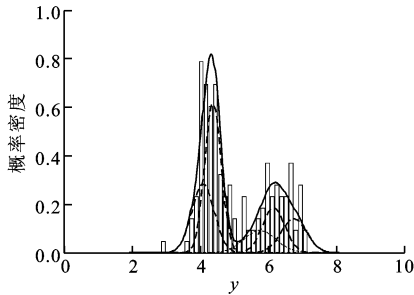


图 4 核酸分布

算例 2 比较算法的收敛情况. 假设初始点为 $(-1.3, 8.8)$ 、 $(2.3, 11.1)$, 实际的观测数据由 2 个高斯混合元组成,如图 6、图 7 所示. 从图中可以看出,EM 算法会收敛到局部极值点 $(1.56, 3.04)$ 和

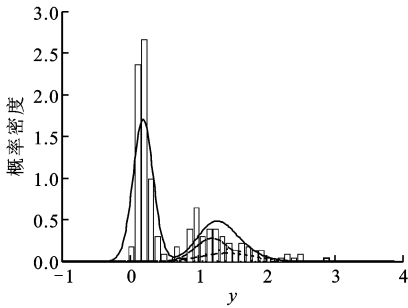


图 5 血液中的酶分布

$(1.61, 3.08)$, 而本文算法会收敛到正确的位置 $(2.01, 5.0)$ 和 $(1.01, 1.11)$. 图中的虚线给出的是参数的学习(收敛)路径. 实际上,一旦初始值给定,EM 算法的收敛路径就已固定,这是因为 EM 算法是一种确定性算法. 本文算法对于相同的初始值,每次收敛路径均有所不同,这主要是由于它是一种随机方法,但它始终会在全局最大值处进入稳态.

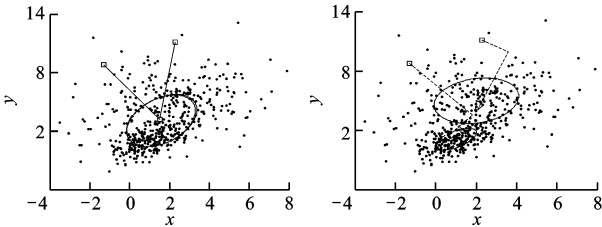


图 6 EM 算法收敛到局部极小值

图 7 本文算法收敛到全局最大值

算例 3 该算例取自文献[5]. 实际的混合模型为 $f(\mathbf{y}) = \sum_{j=1}^4 \omega_j N(\mathbf{y}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, 其中 $\omega_1 = \omega_2 = \omega_3 = 0.3$, $\omega_4 = 0.1$, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = [-4, -4]$, $\boldsymbol{\mu}_3 = [2, 2]$, $\boldsymbol{\mu}_4 = [-1, -6]$.

协方差矩阵

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}; \mathbf{C}_2 = \begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix}$$

$$\mathbf{C}_3 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}; \mathbf{C}_4 = \begin{bmatrix} 0.125 & 0 \\ 0 & 0.125 \end{bmatrix}$$

初始分布元个数为 10, 初始均值、协方差在数据空间随机分布,分布元参数的学习过程如图 8 所示. 最终的参数估计值如下: $\hat{\omega}_1 = 0.299$, $\hat{\omega}_2 = 0.286$, $\hat{\omega}_3 = 0.299$, $\hat{\omega}_4 = 0.11$; $\hat{\boldsymbol{\mu}}_1 = [-4.06, -3.76]$, $\hat{\boldsymbol{\mu}}_2 = [-3.86, -4.03]$, $\hat{\boldsymbol{\mu}}_3 = [2.1, 2.01]$, $\hat{\boldsymbol{\mu}}_4 = [-1.06, -6.06]$;

$$\hat{\mathbf{C}}_1 = \begin{bmatrix} 0.90 & 0.53 \\ 0.53 & 1.10 \end{bmatrix}$$

$$\hat{\mathbf{C}}_2 = \begin{bmatrix} 4.90 & -1.76 \\ -1.76 & 5.50 \end{bmatrix}$$

$$\hat{\mathbf{C}}_3 = \begin{bmatrix} 1.80 & -0.71 \\ -0.71 & 1.60 \end{bmatrix}$$
$$\hat{\mathbf{C}}_4 = \begin{bmatrix} 0.200 & 0 & 0.023 & 2 \\ 0.023 & 2 & 0.100 & 0 \end{bmatrix}$$

可以看出:开始有 10 个分布元,随机分布在数

据空间中;在第 8 s 时,通过合并、剔除后剩下 8 个分布元;在第 15 s 时,仅剩下 5 个分布元;在第 30 s 时,一对分布元被合并,剩下 4 个分布元.可见,分布元参数逐渐收敛到稳态值.

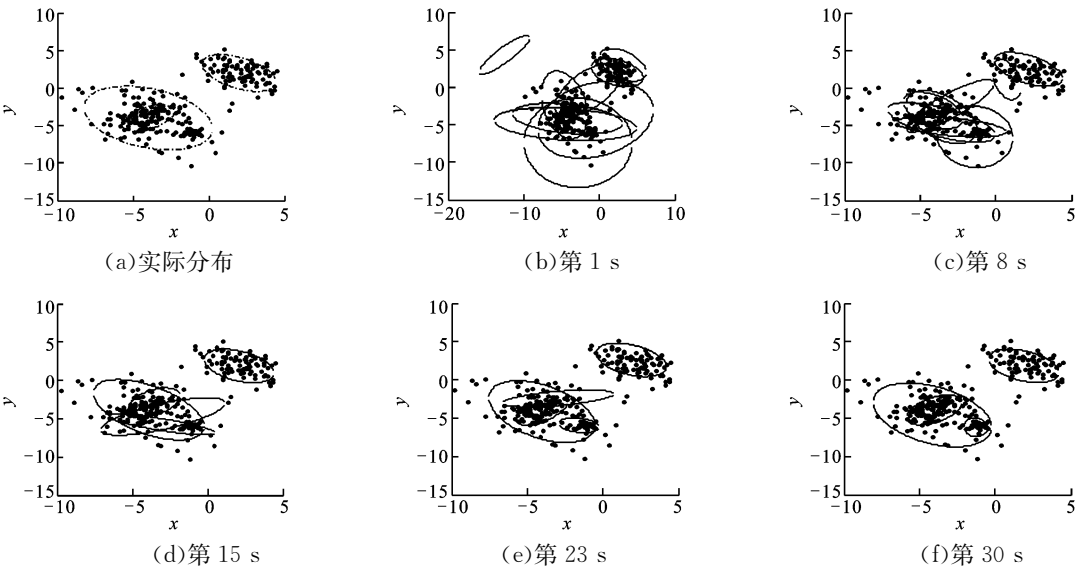


图 8 具有 10 个初始分布元的高斯混合模型的参数学习过程

5 结 论

本文给出了一种 FMM 无监督的学习算法,该算法基于修正 Gibbs 采样并且结合了分布元的合并、剔除两种管理技术. 它的参数初值在数据空间中是随机选择的,分布元的个数仅需取一个适当大的数即可. 这说明算法对初始值具有比较强的鲁棒性. 与 RJMCMC 算法相比,本文算法不需要考虑不同空间的参数跳变问题,所以无需计算跳变概率,仿真实验也表明本文算法是有效的.

参考文献:

[1] MCLACHLAN G, PEEL D. Finite mixture models [M]. New York, USA: John Wiley & Sons, 2000.

[2] CONSTANTINOPOULOS C, TITSIAS M K, BAYESIAN A L. Feature and model selection for Gaussian mixture models [J]. IEEE Trans Patt Anal Mach Intell, 2006, 28(6):1013-1018.

[3] BOUGUILA N, ZIOU D. A hybrid Sem algorithm for high-dimensional unsupervised learning using a finite generalized dirichlet mixture [J]. IEEE Trans Image Processing, 2006, 15(9):2657-2668.

[4] PEMKOPF F, BOUCHAFFRA D. Genetic based EM algorithm for learning Gaussian mixture models [J].

IEEE Trans Patt Anal Mach Intell, 2005, 27(8):1344-1348.

[5] FIGUEIREDO M A T, JAIN A K. Unsupervised learning of finite mixture models [J]. IEEE Trans Patt Anal Mach Intell, 2002, 24(3):381-396.

[6] REDDY C K, CHIANG Hsiao-Dong, RAJARATNAM B. TRUST-TECH-based expectation maximization for learning finite mixture models [J]. IEEE Trans Patt Anal Mach Intell, 2008, 30(7):1146-1157.

[7] DIEBOLT J, ROBERT C P. Estimation of finite mixture distributions through Bayesian sampling [J]. Journal of Royal Statistical Society: B, 1994, 56(2):363-375.

[8] FRUWIRTH-SCHNATTER S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models [J]. Journal of the American Statistical Association, 2001, 96(453):194-209.

[9] GREEN P J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination [J]. Biometrika, 1997, 82(4):711-732.

[10] RICHARDSON S, GREEN P J. On Bayesian analysis of mixtures with an unknown number of components [J]. Journal of Royal Statistical Society: B, 1997, 59(4):731-792.

(编辑 苗凌)