文章编号: 1001-0920(2008)02-0133-07

连续 PH 分布数据拟合的确定性退火 EM 算法

黄 卓, 王文峰, 郭 波
(国防科技大学 信息系统与管理学院, 长沙 410073)

摘 要: 针对目前连续 PH 分布数据拟合 EM(Expectation-Maximization)算法存在的初值敏感问题,提出运用确定性退火 EM 算法进行连续 PH 分布数据拟合,给出了详细的理论推导,并通过两个拟合实例与标准 EM 算法进行了对比. 对比结果表明所提出的方法可以有效地避免初值选择的不同对 EM 算法结果的影响,减小陷入局部最优的可能性,能得到比标准 EM 算法更好的结果.

关键词: PH 分布; 混合 Erlang 分布; 数据拟合; 确定性退火 EM 算法
中图分类号: O211.1 **文献标识码**: A

Fitting data with continuous phase-type distributions via deterministic annealing EM algorithm

HUANG Zhuo, WANG Wen-feng, GUO Bo
(Systems Engineering Department, National University of Defense Technology, Changsha 410073, China. Correspondent: HUANG Zhuo, E-mail: hz-nudt-edu@yahoo.com.cn)

Abstract: To overcome the initial parameters sensitive problem in the conventional expectation-maximization (EM) algorithm for phase-type distributions (PH) data fit, a PH distributions data fit method based on the deterministic annealing EM algorithm is proposed. The detailed theoretical inference process of the method is given. The method is compared with standard EM algorithm through two benchmarks. Contrast results show that the proposed method can obtain better estimates free of the initial parameter values and reduce the possibility of getting into local maximum.

Key words: Phase type distributions; Hyper-erlang distributions; Data fit; Deterministic annealing EM algorithm

1 引 言

连续 PH 分布定义为一个有限状态 Markov 过程的吸收时间分布,其分布函数为 $F(x) = 1 - \boldsymbol{\alpha}\exp(\boldsymbol{T}x)\boldsymbol{e}$,其中 $\boldsymbol{T}$ 为 m 阶方阵,$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)$ 为其瞬态的初始概率向量,$\boldsymbol{e}$ 为元素均为 1 的 m 阶列向量,$(\boldsymbol{\alpha}, \boldsymbol{T})$ 称为该 PH 分布的 m 阶表示. PH 分布中的每一个瞬态称为位相,因此 PH 分布又称为位相型分布. 许多常见分布都是 PH 分布的子集,如指数分布、Erlang 分布、混合指数分布、混合 Erlang 分布等.

PH 分布具有很多良好的特性,如 PH 分布类在 $[0, +\infty)$ 上全部概率分布类中稠密. 稠密性的理论意义在于,当研究分析若干个 $[0, +\infty)$ 上的一般分布 $F_j(j = 1, 2, \cdots, n)$ 的随机模型时,对于这些分布的一个连续泛函 $\Phi(F_1, F_2, \cdots, F_n)$ 进行数学处理是很困难的. 由于 PH 分布在相当大的程度上保持了指数分布的易于进行解析运算的性质,证明连续泛函的某种关系 Φ 对 PH 分布成立相对容易. 如果连续泛函的某种关系 Φ 对 PH 分布成立,而证明过程又不明显依赖于 PH 分布的特殊结构,则可断言关于 Φ 的结论在 F_j 均为一般分布时仍然成立. 运用 PH 分布进行各类随机问题的建模能有效地反映问题的内在规律,建立的模型具有很好的适用性,因此 PH 分布有广泛的应用领域. 此外,PH 分布还具有封闭性,PH 分布对于有限混合和卷积等运算封闭.

目前 PH 分布的理论与应用研究在国外开展的很多,PH 分布已经成功应用于许多科学研究领域,如生物统计学、地震预测分析、统计信号分析、通信系统设计与评估、交通系统分析等. 国内在 PH 分布的理论与应用方面的研究与国外相比较少. 田乃硕[1] 针对 PH 分布在休假排队系统方面的理论进行了深入而细致的研究;李泉林[2] 讨论了 PH 分布在

收稿日期: 2006-11-01; **修回日期**: 2007-01-16.
基金项目: 国家自然科学基金项目(70501031).
作者简介: 黄卓(1980—),男,江西新余人,博士生,从事系统可靠性分析、系统效能评估与优化等研究; 郭波(1962—),男,武汉人,教授,博士生导师,从事系统可靠性分析、装备系统工程等研究.

算法上的意义. 对于 PH 分布的统计分析理论与方法,国外开展了大量的研究工作,而国内尚未见到有文献进行相关研究,仅文献[2]对 Asmussen[3] 的工作进行了阐述.

从 PH 分布的理论到实际应用,PH 分布的统计分析方法起着重要的桥梁作用. PH 分布的统计分析方法主要分两类:矩估计方法和极大似然估计方法. 矩估计方法[4-9] 的研究大多集中在寻找如何将一个一般分布映射为一个 PH 分布的算法,一般都考虑使得一般分布和映射的 PH 分布有相同的前三阶矩. 矩匹配算法的评价标准通常有:匹配阶数的数量、匹配算法的效率、算法的通用性和匹配得到 PH 分布的阶数. PH 分布拟合数据的极大似然估计方法[10-14] 的研究大多是利用标准 EM 算法进行的;中国科学院 Wang[15] 针对混合 Erlang 分布,采用 D&C-EM 算法拟合网络流量中的长尾数据. Thümmler[16] 针对混合 Erlang 分布,运用标准 EM 算法进行数据拟合. 这些研究工作没有涉及如何解决运用 EM 算法,进行 PH 分布数据拟合时存在的初值敏感问题,本文研究的目的就是寻找解决这个问题的方法.

美国加州理工学院的 Rose[17] 于 1990 年首先提出确定性退火算法,该算法是根据自然法则计算的一个重要分支. 其根据退火过程,将求解优化问题的最优点转化为求一系列随温度变化的物理系统的自由能函数的极小,它能够使算法避开局部最优解而得到全局最优解,具有广泛的应用前景. 杨广文等[18] 首次对确定性退火技术的物理背景作了详细的描述,并在理论上证明了当自由能函数满足一定条件时,自由能函数的全局最优解是温度的一个连续映射,从而为确定性退火技术提供了可靠的理论依据. 确定性退火 EM(DAEM) 算法由 Ueda[19] 提出,DAEM 与 EM 算法相比较而言,可以有效地减少初始值选择对最终结果的影响,以避免局部最优.

本文研究 DAEM 算法与 PH 分布拟合相结合的理论问题,给出了详细的理论推导并进行验证.

2 PH 分布的子集 ——混合 Erlang 分布

混合 Erlang 是 PH 分布的一个子类,其定义如下:

定义 1 称一个随机变量 X 服从混合 Erlang 分布,如果该随机变量 X 由若干个 Erlang 分布按一定比例混合而成,其概率密度分布函数可表达为

$$f_X(x) = \sum_{i=1}^{M} \alpha_i \frac{(\lambda_i x)^{k_i-1}}{(k_i-1)!}\lambda_i e^{-\lambda_i x},$$

且满足 $\sum_{i=1}^{M}\alpha_i = 1$,其中 M 表示 Erlang 分布数;组成混合 Erlang 分布的每一个 Erlang 分布称为分支或组成分布.

定理 1[20] 设 H 表示全体混合 Erlang 分布的集合,设 F 表示所有连续非负随机变量的集合,则 H 为 F 的一个稠密集合,即 F 中的任意随机变量可由 H 中的随机变量任意逼近.

由定理 1 可知,选择混合 Erlang 分布进行数据拟合具有普遍的适用性,理论上能够很好地拟合各种分布函数,因此本文选择 PH 分布的子类 ——混合 Erlang 分布进行数据拟合的研究.

3 混合分布数据拟合的 EM 算法

3.1 EM 算法

EM 算法是一种在不完全数据情况下计算极大似然估计或者后验分布的迭代算法,最初由 Dempster[3] 等提出. EM 算法在每一迭代循环过程中交替执行两个步骤:E 步(Expectation step,期望步)和 M 步(Maximization step,极大化步). EM 步在给定完全数据和前一次迭代所得到的参数估计的情况下,计算完全数据对应的对数似然函数的条件期望;M 步极大化完全数据对数似然函数以确定参数的值,并用于下步的迭代. 算法在 E 步和 M 步之间不断迭代直至收敛,即两次迭代之间的参数变化小于一个预先给定的阈值时结束.

一般地,以 $p(\theta|Y)$ 表示 θ 的基于观测数据的分布密度函数,称为观测后验分布;$p(\theta|Y,Z)$ 表示在添加数据 Z 后得到的关于 θ 的后验分布密度函数,称为添加后验分布;$p(Z|\theta,Y)$ 表示在给定 θ 和观测数据 Y 时潜在数据 Z 的条件分布密度函数,EM 算法为极大似然估计方法,其目的是计算 $p(\theta|Y)$ 的参数 θ. EM 算法按如下步骤进行,记 $\theta^{(j)}$ 为第 $i+1$ 次迭代开始时的参数估计值,则第 $i+1$ 次迭代的两步为:

1) E 步:将 $p(\theta|Y,Z)$ 或 $\log p(\theta|Y,Z)$ 关于 Z 的条件分布求期望,从而把 Z 通过积分去掉,即

$$Q(\theta|\theta^{(j)},Y) = \int \log p(\theta|Y,Z)\, p(Z|\theta^{(j)},Y)\,\mathrm{d}Z. \quad (1)$$

2) M 步:将 $Q(\theta|\theta^{(j)},Y)$ 极大化,即寻找一个点 $\theta^{(i+1)}$,使

$$Q(\theta^{(i+1)}|\theta^{(j)},Y) = \max_{\theta} Q(\theta|\theta^{(j)},Y), \quad (2)$$

如此形成了一次迭代过程 $\theta^{(j)} \to \theta^{(i+1)}$.

将上述 E 步和 M 步进行迭代直至 $\| Q(\theta^{(i+1)}|\theta^{(j)},Y) - Q(\theta^{(j)}|\theta^{(j)},Y) \|$ 或 $\| \theta^{(i+1)} - \theta^{(j)} \|$ 充分小时停止.

EM 算法本质上是梯度法,但与其他的数值方法相比,EM 算法具有简单、易实现、数值稳定、保证

收敛的优点,因此本文选择 EM 算法进行数据拟合.

3.2　混合分布数据拟合的 EM 算法

设混合概率密度函数为 $p(x\mid\Theta)=\sum_{l=1}^{M}a_l p_l(x\mid\theta_l)$,参数 $\Theta=(a_1,\cdots,a_M,\theta_1,\cdots,\theta_M)$,同时满足条件 $\sum_{l=1}^{M}a_l=1$.

设 $X=(x_1,\cdots,x_N)$ 为对 $p(x\mid\Theta)$ 进行观测得到的样本数据,此时的对数似然函数为

$$\log(L(\Theta\mid X))=\sum_{i=1}^{N}\log\Big(\sum_{l=1}^{M}a_l p_l(x_i\mid\theta_l)\Big). \quad (3)$$

由于存在和的对数运算,求解使式(1)极大的 Θ 是很困难的.为此考虑 X 为不完全数据,假设存在不可观测数据项 $Y=(y_1,\cdots,y_N)$,y_i 表示第 i 个观测数据 x_i 由混合分布的第 y_i 个概率分布函数产生,则 $y_i\in(1,\cdots,M)$,若 $y_i=k$ 则表示第 i 个观测数据 x_i 由混合分布的第 k 个概率分布函数产生.

如果能观测到数据 $Y=(y_1,\cdots,y_N)$,则可把 (X,Y) 作为完全数据,此时可以构造完全数据对数似然函数为

$$\log(L(\Theta\mid X,Y))=\sum_{i=1}^{N}\log(a_{y_i}p_{y_i}(x_i\mid\theta_{y_i})). \quad (4)$$

下面构造 EM 算法的 Q 函数,即完全数据对数似然函数关于丢失数据边缘概率分布密度函数的期望.在已知参数 $\Theta^{(g)}=(a_1^{(g)},\cdots,a_M^{(g)},\theta_1^{(g)},\cdots,\theta_M^{(g)})$ 的条件下,由 Bayes 公式可得

$$p(y_i\mid x_i,\Theta^{(g)})=\frac{a_{y_i}^{(g)}p_{y_i}(x_i\mid\theta_{y_i}^{(g)})}{\sum_{l=1}^{M}a_l^{(g)}p_l(x_i\mid\theta_l^{(g)})}. \quad (5)$$

令 $Y=(y_1,\cdots,y_N)$ 表示 N 个独立采样数据来自混合分布的那个组成分布,则 Y 边缘分布密度函数为

$$p(Y\mid X,\Theta^{(g)})=\prod_{i=1}^{N}p(y_i\mid x_i,\Theta^{(g)}), \quad (6)$$

则 EM 算法的 Q 函数为

$$Q(\Theta,\Theta^{(g)})=\sum_{Y\in\Omega}\log(L(\Theta\mid X,Y))\,p(Y\mid X,\Theta^{(g)}), \quad (7)$$

其中 Ω 表示 Y 的取值空间.

对式(7)进行化简可得混合分布的 Q 函数为

$$Q(\Theta,\Theta^{(g)})=\sum_{l=1}^{M}\sum_{i=1}^{N}(\log(a_l p_l(x_i\mid\theta_l))\,p(l\mid x_i,\Theta^{(g)}), \quad (8)$$

其中 $p(l\mid x_i,\Theta^{(g)})$ 表示数据 x_i 从第 l 个组成分布产生的概率.

关于式(7)化简的具体过程请参见文献[21].

4　混合 Erlang 分布数据拟合的确定性退火 EM 算法

4.1　确定性退火 EM 算法及其全局最优特性

由于 EM 算法的结果依赖于开始迭代的初始参数值,EM 算法一般只能收敛到局部最优值解.在 EM 算法中,未知数据后验概率密度函数在 M 步中起重要作用,而该函数在迭代的初始阶段是很不可靠的[19].为此,基于最大熵准则,文献[19]给出了一个新的后验概率密度函数

$$f(x_{\mathrm{mis}}\mid x_{\mathrm{obs}},\Theta)=\frac{1}{Z}\exp\{-\beta(-L_c(x;\Theta))\}. \quad (9)$$

其中:x,x_{mis},x_{obs} 分别表示完全数据、观测数据和丢失数据;$L_c(x;\Theta)$ 表示完全数据的似然函数;$Z=\int\exp\{-\beta(-L_c(x;\Theta))\}\mathrm{d}x_{\mathrm{mis}}$,$Z$ 称为配分函数;$\frac{1}{\beta}$ 类似于退火的温度.值得注意的是,当 $\beta=1$ 时,由式(9)得到的后验概率密度与式(5)相同,即 $\beta=1$ 时的解空间与原问题的解空间相同,因此随着温度的降低,β 将逐渐接近并最终等于 1.

采用式(9)的未知数据后验概率密度函数后,EM 算法的完全数据似然函数的条件期望变为

$$f(x_{\mathrm{mis}}\mid x_{\mathrm{obs}},\Theta)=\frac{\exp\{-\beta(-L_c(x;\Theta))\}}{\int\exp\{-\beta(-L_c(x;\Theta))\}\mathrm{d}x_{\mathrm{mis}}}. \quad (10)$$

从式(10)可以看出,参数 β 对未知数据后验概率密度函数具有平滑作用.β 很小时,$f(x_{\mathrm{mis}}\mid x_{\mathrm{obs}},\Theta)$ 为 $[0,1]$ 上的均匀分布,这个全局最小可以很容易通过传统的 EM 算法找到.随着 β 的增大(相当于降低温度),后验概率密度函数的作用逐渐增强.由于参数 β 的平滑作用,在两个相邻的温度值之间,新的全局最小值与前一个全局最小值很接近,可把前一个全局最小值作为下一个温度下的初始参数值,然后运行 EM 算法,就容易找到新的全局最小值.不断降低温度,直到 $\beta=1$,理论上最终可以得到全局最优值[18].

4.2　混合 Erlang 分布拟合数据的确定性退火 EM 算法

由定义 1 可知,混合 Erlang 分布密度函数为

$$p(x\mid\Theta)=\sum_{l=1}^{M}\alpha_l\frac{(\lambda_l x)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x}, \quad (11)$$

其中 $\Theta=(\alpha_1,\cdots,\alpha_M,\lambda_1,\cdots,\lambda_M)$.

令 $p(x_i\in l)=p(l\mid x_i,\Theta^{(g)})$,由式(8)可知混合 Erlang 分布 EM 算法的 Q 函数为

$$Q(\Theta,\Theta^{(g)})=$$

$$\sum_{l=1}^{M}\sum_{i=1}^{N}\left(\log\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)\right)p(x_i\in l).$$

其中:$\Theta^{(g)}$ 表示 EM 算法进行 E 步计算时采用的参数,Θ 表示 M 步需要计算的新参数.

根据 DA EM 算法,对于 $p(x_i\in l)$ 没有任何事先确知的信息,采用极大熵原理求 $p(x_i\in l)$.

目标:使熵 $H(p)=-\sum_{l=1}^{M}\sum_{i=1}^{N}\left(\log p(x_i\in l)\right)p(x_i\in l)$ 极大.约束为

$$\begin{cases} Q(\Theta,\Theta^{(g)})= \\ \sum_{l=1}^{M}\sum_{i=1}^{N}\left(\log\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)\right)p(x_i\in l), \\ \sum_{i=1}^{N}\left(\sum_{l=1}^{M}p(x_i\in l)-1\right)=0, \end{cases}$$

其中 $Q(\Theta,\Theta^{(g)})$ 理解为内能,在某一个确定的温度值时为常数.

为极大化 $H(p)$,引入拉格朗日乘数 β 和 λ,根据拉格朗日乘数法,可定义

$$\begin{aligned}\tilde{F}=&-\sum_{l=1}^{M}\sum_{i=1}^{N}\left(\log p(x_i\in l)\right)p(x_i\in l)+\\ &\beta\left(\sum_{l=1}^{M}\sum_{i=1}^{N}\left(\log\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)\right)p(x_i\in l)-\right.\\ &\left.Q(\Theta,\Theta^{(g)})\right)+\lambda\left(\sum_{i=1}^{N}\left(\sum_{l=1}^{M}p(x_i\in l)-1\right)\right). \quad (12)\end{aligned}$$

下面求解使 $\tilde{F}$ 取得极大值的 $p(x_i\in l)$.

将式(12)对 $p(x_i\in l)$ 求偏导数并令其为 0,可得

$$\begin{aligned}&\beta\log\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)-\\ &\log p(x_i\in l)-1+\lambda=0. \quad (13)\end{aligned}$$

化简式(13)可得

$$p(x_i\in l)=\frac{\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)^{\beta}}{\exp(1-\lambda)}. \quad (14)$$

由 $\sum_{l=1}^{M}p(x_i\in l)=1$,有

$$\sum_{l=1}^{M}p(x_i\in l)=\frac{\sum_{l=1}^{M}\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)^{\beta}}{\exp(1-\lambda)}=1, \quad (15)$$

则有

$$\exp(1-\lambda)=\sum_{l=1}^{M}\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)^{\beta}. \quad (16)$$

将式(16)代入(14),可得

$$p(x_i\in l)=\frac{\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)^{\beta}}{\sum_{l=1}^{M}\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)^{\beta}}. \quad (17)$$

式(17)中当 $\beta=1$ 时,$p(x_i\in l)$ 与标准 EM 算法计算未知数据期望值的方法相同.

下面求使 $Q(\Theta,\Theta^{(g)})$ 极大的 $a_l(l=1,2,\cdots,M)$ 和 $\lambda_l(l=1,2,\cdots,M)$,约束条件 $\sum_{l=1}^{M}a_l=1$.根据拉格朗日乘数法,引入拉格朗日乘数 λ,并定义

$$\begin{aligned}F=&\sum_{l=1}^{M}\sum_{i=1}^{N}\left(\log\left(a_l\frac{(\lambda_l x_i)^{k_l-1}}{(k_l-1)!}\lambda_l e^{-\lambda_l x_i}\right)\right)p(x_i\in l)+\\ &\lambda\left(\sum_{l=1}^{M}a_l-1\right). \quad (18)\end{aligned}$$

将式(18)对 a_l 求偏导数得

$$\frac{\partial F}{\partial a_l}=\frac{1}{a_l}\sum_{i=1}^{N}p(x_i\in l)+\lambda=0, \quad (19)$$

则有

$$a_l=-\frac{1}{\lambda}\sum_{i=1}^{N}p(x_i\in l). \quad (20)$$

因 $\sum_{l=1}^{M}a_l=1$ 有 $\sum_{l=1}^{M}\left(-\frac{1}{\lambda}\sum_{i=1}^{N}p(x_i\in l)\right)=1$,则 $\lambda=-N$,所以

$$a_l=\frac{1}{N}\sum_{i=1}^{N}p(x_i\in l). \quad (21)$$

将式(18)对 λ_l 求偏导数可得

$$\begin{aligned}&\partial F/\partial\lambda_l=\\ &\sum_{i=1}^{N}\left\{(k_l-1)\frac{1}{\lambda_l}+\frac{1}{\lambda_l}-x_i\right\}p(x_i\in l)=0, \quad (22)\end{aligned}$$

化简式(22)可得

$$\lambda_l=\frac{k_l\sum_{i=1}^{N}p(x_i\in l)}{\sum_{i=1}^{N}x_i p(x_i\in l)}. \quad (23)$$

由式(17),(21)和(23)可以得到混合 Erlang 分布拟合数据的确定性退火 EM 算法的步骤如下:

Step1:选择混合 Erlang 分布的初始参数 $\Theta^{(g)}$.

Step2:给定 β 初始值(温度的倒数)及温度降低的系数 $C(C>1)$.

Step3:定义停止 EM 算法迭代的收敛标准(本文标准为相邻两步结果的对数似然度之差小于某个值).

Step4:当不满足 EM 算法迭代的收敛标准时,重复执行以下 3 步:

Step4.1:E 步:在当前温度值和参数下根据式(17)计算 $p(x_i\in l)(l=1,\cdots,M,i=1,\cdots,N)$;

Step4.2:M 步:根据式(21)和(23)计算新参数值 a_l 和 $\lambda_l(l=1,\cdots,M)$;

Step4.3:更新当前参数值 $a_l^{(g)}=a_l(l=1,\cdots,M)$,$\lambda_l^{(g)}=\lambda_l(l=1,\cdots,M)$.

Step5: 降低温度值,令 $\beta = C \times \beta$.

Step6: 当 $\beta \leq 1$ 时,执行 Step4;否则算法结束.

4.3 β初始值及降温系数 C 的选择

文献[19]的实验结果表明:β的初始值一般选取为 0.1;降温系数 C 一般在区间[1.1,1.5]内选取. 本文在第 5 节的对比分析中,β取值为 0.1,C 取值为 1.2.

4.4 混合分布确定性退火 EM 算法与标准 EM 算法的算法时间复杂度分析

一般情况下,标准 EM 算法需要从多个不同起始参数分别进行迭代,并从这些结果中选择最优解. 设标准 EM 算法选择不同初始参数个数为 R,则标准 EM 算法的时间复杂度为

$$O(R \times M \times N), \tag{24}$$

其中 M 为混合 Erlang 分布的分支数,N 为样本数.

确定性退火 EM 算法需要在每个温度值进行 EM 迭代,当 $\beta = 1$ 时停止. 设 k 表示确定性退火 EM 算法的降温次数,则有

$$\beta \times C^k = 1, \tag{25}$$

对式(25)两边去对数并化简可得

$$k = -\ln \beta / \ln C, \tag{26}$$

则有确定性退火 EM 算法的时间复杂度为

$$O\left(-\frac{\ln \beta}{\ln C} \times M \times N\right). \tag{27}$$

对比式(24)和(27)可以看出,确定性退火 EM 算法和标准 EM 算法的时间复杂度基本相同. 因为标准 EM 算法选择初始参数是随机的、盲目的,所以在许多实际问题中可以认为标准 EM 算法会耗费更多的时间.

5 确定性退火 EM 算法与 EM 算法拟合结果对比分析

本节通过两个实例验证本文提出的混合 Erlang 分布拟合数据的 DAEM 算法在避免初值影响方面的有效性,并与标准 EM 算法得到的结果进行比较. DAEM 算法和 EM 算法均用 Matlab 实现.

从不同初始点开始得到的不同结果可以认为是不同的局部最优解. 混合 Erlang 分布的参数多,解空间的具体形态很难描述,因此一个解是否为全局最优解就无法确切知道. 事实上,当标准 EM 算法从足够多个不同的随机初始点开始迭代,得到的最好结果可以认为是近似全局最优解或全局最优解.

5.1 DAEM 和 EM 算法拟合 Weibull 分布数据的结果对比分析

Weibull 分布一般为双参数分布,调整其尺度参数 α和形状参数β可以得到很多分布曲线形状以满足数据拟合的要求. 其在许多领域被广泛采用,极具代表性. 这里选择尺度参数 $\alpha = 3.5$,形状参数 $\beta = 4$,对 Weibull 分布随机采样 100 个数据,从不同的初始参数开始,分别用 DAEM 算法和 EM 算法进行 5 次拟合,拟合结果见表 1. 表 1 中混合 Erlang 分布起始参数的每一行代表一个 Erlang 分布的参数,依次为 α_i,λ_i,k_i.

从表 1 可以看出,从不同的混合 Erlang 分布的起始参数出发,标准 EM 算法得到的拟合 Weibull 分布数据的结果差异很大,易陷入不同的局部最优;而 DAEM 算法从不同的起始参数出发可以得到近乎相同的结果,有效地避免了初始参数值对最终结果

表 1　Weibull 分布拟合结果对比

次数	起始参数值	EM 算法结果		DAEM 算法结果	
		拟合结果	对数似然度	拟合结果	对数拟然度
1	0.2 0.1 4 0.3 0.2 2 0.4 1 3 0.1 1 3	0.113 4 1.243 4 4 0.126 0 0.642 8 2 0.608 5 0.970 1 3 0.152 1 0.970 1 3	-613.705 9	0.474 5 1.280 6 4 0.073 1 0.652 2 2 0.226 2 0.967 6 3 0.226 2 0.967 6 3	-335.337 4
2	0.2 0.9 4 0.3 0.2 2 0.4 1 3 0.1 1 3	1.000 0 1.286 6 4 0.000 0 0.667 1 2 0.000 0 0.977 7 3 0.000 0 0.977 7 3	-300.039 3	0.474 5 1.280 6 4 0.073 1 0.652 2 2 0.226 2 0.967 6 3 0.226 2 0.967 6 3	-335.337 4
3	0.2 0.2 4 0.3 0.2 2 0.4 1 3 0.1 1 3	0.000 9 1.222 8 4 0.144 3 0.640 3 2 0.683 9 0.965 8 3 0.171 0 0.965 8 3	-743.108 0	0.474 5 1.280 6 4 0.073 1 0.652 2 2 0.226 2 0.967 6 3 0.226 2 0.967 6 3	-335.337 4
4	0.2 0.3 4 0.3 0.2 2 0.4 1 3 0.1 1 3	0.033 1 1.232 2 4 0.138 8 0.641 1 2 0.662 5 0.967 2 3 0.165 6 0.967 2 3	-697.024 9	0.474 5 1.280 6 4 0.073 1 0.652 2 2 0.226 2 0.967 6 3 0.226 2 0.967 6 3	-335.337 4
5	0.2 0.5 4 0.3 0.2 2 0.4 1 3 0.1 1 3	0.120 1 1.205 2 4 0.130 0 0.622 7 2 0.599 9 0.945 2 3 0.150 0 0.150 0 3	-610.690 4	0.472 5 1.246 7 4 0.072 9 0.631 8 2 0.227 3 0.940 8 3 0.227 3 0.940 8 3	-341.543 4

的影响. EM 算法得到的平均对数似然度为 - 592.913 7,DAEM 算法得到的平均对数似然度为 - 336.578 6. 从拟合结果可以看出,本文提出的算法优于 EM 算法的结果.

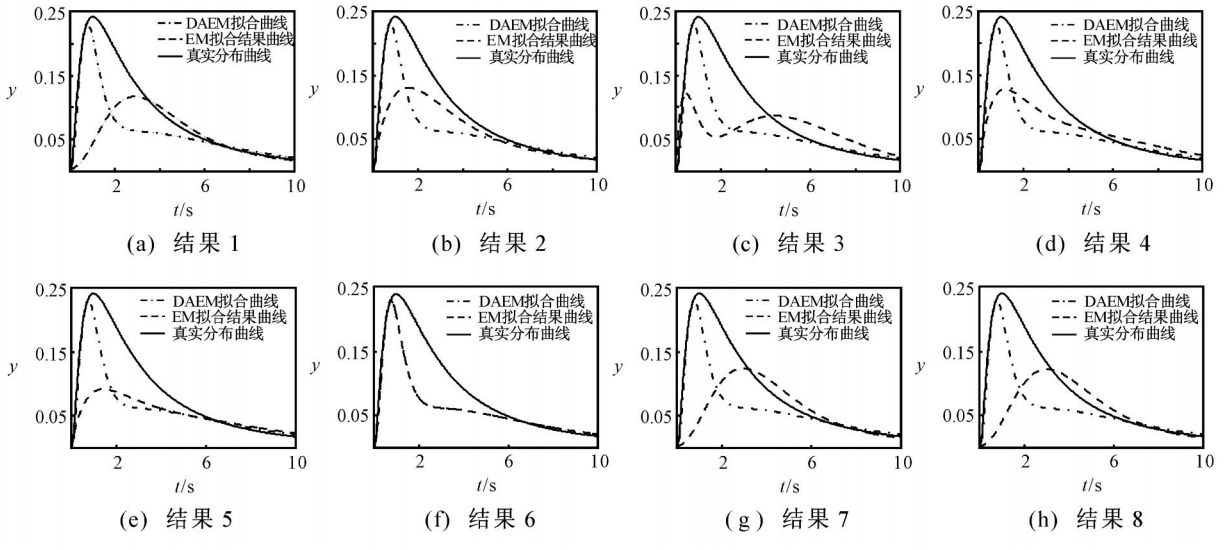
5.2 DAEM 和 EM 算法拟合对数正态分布数据的结果对比分析

随机变量的自然对数服从均值为 μ 和标准差为 δ 的正态分布,称为对数正态分布. 这里选择 $\mu = 1$, $\delta = 1$,对这个对数正态分布随机采样 200 个数据,分别用 DAEM 算法和 EM 算法进行 8 次拟合,拟合结果见表 2,拟合效果见图 1.

从表 2 可以看出,从不同的混合 Erlang 分布的起始参数出发,EM 算法得到的拟合对数正态分布的结果差异很大;而DAEM算法从不同的起始参数

表 2 对数正态分布拟合结果对比

次数	起始参数值	EM 算法结果		DAEM 算法结果		拟合曲线
		拟合结果	对数似然度	拟合结果	对数拟然度	
1	0.5 0.9 4 0.3 0.2 2 0.1 1 3 0.1 1 3	0.378 5 1.062 1 4 0.273 0 0.267 3 2 0.174 3 1.979 2 3 0.174 3 1.979 2 3	- 613.195 6	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 1
2	0.1 0.9 4 0.6 0.8 2 0.2 1 3 0.1 1 3	0.171 5 0.413 5 4 0.539 8 0.648 5 2 0.159 4 0.854 5 3 0.129 3 3.453 4 3	- 543.611 6	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 2
3	0.5 0.9 4 0.3 0.8 2 0.1 1 3 0.1 4 3	0.388 8 0.554 9 4 0.361 3 0.929 2 2 0.109 6 1.191 8 3 0.140 4 3.015 5 3	- 535.366 7	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 3
4	0.3 0.9 4 0.3 0.8 2 0.2 1 3 0.2 4 3	0.308 6 0.505 4 4 0.305 7 0.776 6 2 0.199 1 1.033 2 3 0.186 7 2.973 4 3	- 582.172 8	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 4
5	0.5 5 4 0.3 0.2 2 0.1 1 3 0.1 1 3	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 5
6	0.5 0.9 4 0.3 0.2 2 0.1 0.5 3 0.1 1 3	0.436 2 1.025 9 4 0.240 5 0.273 1 2 0.192 6 2.249 0 3 0.130 8 1.413 9 3	- 609.175 7	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 6
7	0.5 0.9 4 0.3 0.2 2 0.1 3 3 0.1 1 3	0.444 9 1.012 6 4 0.230 8 0.261 0 2 0.206 5 3.285 4 3 0.117 8 1.257 1 3	- 612.907 5	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 7
8	0.5 0.9 4 0.3 5 2 0.1 3 3 0.1 1 3	0.568 4 0.679 7 4 0.136 8 2.419 6 2 0.158 8 2.434 2 3 0.136 0 1.192 4 3	- 574.217 6	0.217 6 4.101 7 4 0.496 7 0.330 9 2 0.142 8 1.096 9 3 0.142 8 1.096 9 3	- 542.798 6	见图 1 结果 8



(a) 结果 1 (b) 结果 2 (c) 结果 3 (d) 结果 4

(e) 结果 5 (f) 结果 6 (g) 结果 7 (h) 结果 8

图 1 拟合结果对比

出发可以得到基本相同的结果,有效地避免了初始参数值对最终结果的影响. EM 算法得到的平均对数似然度为 - 576.603 1,DAEM 算法得到的平均对数似然度为 - 542.798 6. 从拟合结果可以看出,本文提出的算法优于 EM 算法的结果.

6 结　　论

本文提出采用确定性退火 EM 算法进行 PH 分布的数据拟合,给出了混合 Erlang 分布拟合数据的确定性退火 EM 算法的理论推导过程和算法的具体执行步骤,并用 Matlab 实现了混合 Erlang 分布拟合数据的确定性退火 EM 算法和标准 EM 算法. 通过对两种算法拟合 Weibull 分布和对数正态分布结果的对比分析可以看出,DAEM 算法有效地解决了 EM 算法对初值的敏感性问题,能通过迭代得到较优的结果;而 EM 算法不能避免初值所带来的影响,拟合得到的结果不稳定. 本文提出的算法优于目前常用于 PH 分布拟合的标准 EM 算法. 下一步的研究工作是如何将 DAEM 算法应用于一般 PH 分布的数据拟合.

参考文献(References)

[1] 田乃硕. 拟生灭过程与矩阵几何解[M]. 北京: 科学出版社, 2002.
(Tian Nai-shuo. Quasi birth-death process and matirx geometric solutions[M]. Beijing: Science Press, 2002.)

[2] 李泉林. 随机模型的算法研究[D]. 北京: 中国科学院自动化研究所, 1999.
(Li Quan-lin. Research on the algorithms of stochastic models[D]. Beijing: Institute of Automation, Chinese Academy of Science, 1999.)

[3] Asmussen S, Nerman O, Olsson M. Fitting phase-type distributions via the EM algorithm[J]. Scandinavian J of Statistics, 1996, 23 (4): 419-441.

[4] Bobbio A, Horvath A, Telek M. Matching three moments with minimal acyclic phase type distributions [J]. Stochastic Models, 2005, 21(2/3): 303-326.

[5] Johnson M A, Taaffe M R. Matching moments to phase distributions: Mixtures of Erlang distributions of common order [J]. Communications in Statistics-Stochastic Models, 1989, 5(6): 711-743.

[6] Johnson M A, Taaffe M R. Matching moments to phase distributions: Density function shapes [J]. Communications in Statistics-Stochastic Models, 1990, 6(2): 283-306.

[7] Johnson M A, Taaffe M R. Matching moments to phase distributions: Nonlinear programming approaches [J]. Communications in Statistics-Stochastic Models, 1990, 6(2): 259-281.

[8] Johnson M A, Taaffe M R. An investigation of phase-distribution moment-matching algorithms for use in queueing models[J]. Queueing Systems, 1991, 8(2): 129-147.

[9] Osogami T, Harchol-Balter M. Closed form solutions for mapping general distributions to quasi-minimal PH distributions[J]. Performance Evaluation, 2006, 63 (6): 524-552.

[10] Bobbio A, Cumani A. ML estimation of the parameters of a PH distribution in triangular canonical form[C]. Computer Performance Evaluation. Amsterdam: Elsevier, 1992: 33-46.

[11] Feldmann A, Whitt W. Fitting mixtures of exponentials to long-tailed distributions to analyze network performance models [J]. Performance Evaluation, 1998, 31(3/4): 245-279.

[12] Horvath A, Telek M. Approximating heavy tailed behavior with phase type distribution[C]. Proc of the 3rd Int Conf on Matrix-Analytic Methods in Stochastic Models. Leuven, 2000.

[13] Khayari R E, Sadre R, Haverkort B R. Fitting world-wide web request traces with the EM-algorithm[J]. Performance Evaluation, 2003, 52 (2/3): 175-191.

[14] Riska A, Diev C, Smirni E. An EM-based technique for approximating long-tailed data sets with PH distributions[J]. Performance Evaluation, 2004, 5 (2): 147-164.

[15] Wang J F, Zhou H X, Zhou M T, et al. A general model for long-tailed network traffic approximation[J]. J of Supercomputing, 2006, 38 (2): 155-172.

[16] Thummler A, Buchholz P, Telek M. A novel approach for phase-type fitting with the EM algorithm[J]. IEEE Trans on Dependable and Secure Computing, 2006, 3 (3): 245-258.

[17] Rose K, Gurewitz E, Fox G C. Statistical mechanics and phase transitions in clustering[J]. Physical Review Letters, 1990, 65(9): 945-948.

[18] 杨广文, 李晓明, 王义和. 确定性退火技术[J]. 计算机学报, 1998, 21 (8): 765-768.
(Yang Guang-wen, Li Xiao-ming, Wang Yi-he. Deterministic annealing technique [J]. Chinese J of Computers, 1998, 21(8): 765-768.)

[19] Ueda N, Nakano R. Deterministic annealing EM algorithm[J]. Neural Networks, 1998, 11 (2): 271-282.

[20] Yuguang F. Hyper-Erlang distribution model and its application in wireless mobile networks[J]. Wireless Networks, 2001, 7(2): 211-219.

[21] Bilmes J A. A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models [R]. Berkeley: International Computer Science Institute, 1997.