

有向网络的混合模型新退火算法研究

王静红^{1,2}, 柴变芳³, 李 笔⁴

(1 河北师范大学信息技术学院, 河北省网络与信息安全重点实验室, 河北石家庄 050000;

2. 伊利诺伊大学香槟分校, 伊利诺伊乌尔班纳 6180;

3. 河北地质大学信息学院, 河北石家庄 050001; 4. 河北师范大学商学院, 河北石家庄 050240)

摘要: 混合模型的传统期望最大化(EM)算法可以有效地探索网络的结构规律性,但它总是陷入局部最大值.为此提出了确定性退火期望最大化(NMEM)算法来解决这个问题,该算法不仅能够防止局部最优,而且提高了收敛速度,因此 NMEM 算法适用于估计混合模型的参数.该算法通过经验设置其初始参数 β_0 , 设计了有向网络的混合模型新退火算法,并设计了 β_0 的参数选择方法.

关键词: 混合模型;退火算法;收敛速度;有向网络

中图分类号: TP391 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2018.06.001

引用格式: 王静红,柴变芳,李笔. 有向网络的混合模型新退火算法研究[J]. 中国科学技术大学学报, 2018, 48(6):433-439.

WANG Jinghong, CAI Bianfang, LI Bi. Research on a new annealing algorithm for mixed model of directed networks[J]. Journal of University of Science and Technology of China, 2018, 48(6):433-439.

Research on a new annealing algorithm for mixed model of directed networks

WANG Jinghong^{1,2}, CAI Bianfang³, LI Bi⁴

(1. College of Information Technology, Hebei Normal University; Hebei Key Lab Network and Information Security, Shijiazhuang 050000, China;

2. University of Illinois at Urbana Champaign, Illinois 61801, USA;

3. Department of Information Engineering, Hebei GEO University, Shijiazhuang 050031, China;

4. Business College, Hebei Normal University, Shijiazhuang 050024, China)

Abstract: Although the traditional expectation-maximization (EM) algorithm of the mixed model can effectively explore the structural regularity of the network, it always gets stuck in some local maximum. A deterministic annealing expectation maximization (NMEM) algorithm is proposed to solve this problem, which not only prevents local optimum but also improves convergence speed and is thus used to estimate the parameters of the hybrid model. The algorithm always sets its initial parameters β_0 through experience. If β_0 is too small, the results are meaningless, or if β_0 is too large, it will converge to the local maximum more frequently. Furthermore, a new hybrid model of directional network and a parameter selection method of β_0 were designed.

Key words: mixture model; annealing algorithm; convergence rate; directed network

收稿日期: 2017-06-12; 修回日期: 2017-07-14

基金项目: 国家自然科学基金(61672206); 河北自然科学基金(F2013205192); 河北省教育厅项目(ZD2018023)资助.

作者简介: 王静红,女,博士/教授,研究方向:机器学习与数据挖掘,复杂网络. E-mail: wangjinghong@126.com

通讯作者: 柴变芳,博士/副教授. E-mail: wangjinghong6301@163.com

0 引言

社区检测^[1]是网络分析中一类流行的自动识别技术,如抑制整个网络的复杂性和识别网络中的关键节点等.到目前为止,已经开发了大量用于社区检测任务的方法,包括层次聚类,分割聚类,基于模块化的方法等,它们仅仅集中于检测紧密连接的子图,而复杂网络可能具有许多其他类型的结构,包括核心-外围,分层、多分体结构或它们的混合等.最近提供了一些模型来检测除紧密连接的子图之外的更多种结构,这些模型主要分两大类.一类是基于随机块模型(SBM)^[2]的算法,其算法通过 Gibbs 抽样方法^[3]来估计参数,变分 EM 算法^[4-6],变分贝叶斯方法^[7],传播方法^[8]等.这些算法的时间复杂度近似为 $O(mc^2)$,其中 m 和 c 分别表示边和簇的数量.另一类是用于网络探索的混合模型^[9],其时间复杂度为 $O(mc)$.相比之下,混合模型的 EM 算法(EMMM)^[9]比基于 SBM 的模型的算法更有效.

传统的 EM 算法总是收敛到差的局部最大值,文献^[10]提出 DAEM 算法来克服局部最大值问题.它从 $\beta = \beta_0 \cong 1$ 开始,并缓慢增加 β 到 1,在每个 β , DAEM 算法执行 EM 算法.这增加了收敛时间,特别是对于具有倾斜的混合系数和簇之间的重叠的数据.EM 算法^[11]从 $\beta > 1$ 开始缓慢减小到 1.它通过限制重叠量来提高速度,但更频繁地趋于收敛到较差的局部最优.DAAEM 算法^[11],不仅能防止 EM 算法陷入局部最优,而且提高了算法的收敛速度.它从一个小参数 $\beta = \beta_0 \cong 0$ 开始,缓慢增加 β 并超过 1,直到选定的上限,最后慢慢地降低 β 到 1.

为了使混合模型的参数估计算法^[9]收敛到局部最优或有效地近似全局最优,我们使用 DAAEM 算法框架来估计混合模型的参数,构造新混合模型的 DAAEM 算法简称为 NMEM.在 DAAEM 算法需要设置初始 β_0 ,在文献^[10-11]中,非上标,初始值 β_0 总是由实验设定的.Ueda 等^[10]证实 $\beta_0 = 0.1$ 可能足够小,新迭代中的 β 被设置为常数(最后迭代中 β 的 1.1~1.5)倍;Naim 等^[11]认为有必要为复杂数据选择较小的 β_0 ,而对于简单数据选择较大的 β_0 ,目前还没有关于如何在理论上设置它的研究,如果我们设置太小的 β_0 ,NMEM 算法可能将节点划分为无意义的集群^[12],例如每个节点属于具有相等概率的每个集群的分区.一旦算法收敛到这种情况,在随后的迭代中它是不能从该点逃逸的.如果我们设置

太大的 β_0 ,NMEM 算法更容易收敛到局部最大值.从收敛性质方面来说,选择的初始参数 β_0 开始的算法不应该收敛到一个无意义的聚类点,这应该是 NMEM 算法的一个不稳定的固定点.

通过收敛性分析有助于研究算法的固定点,收敛速率用于测量固定点是否稳定.计算收敛速率的两个流行工具是 Hessian 矩阵和 Jacobian 矩阵.Xu 等^[13]提出了混合参数收集的高斯混合的对数似然函数的 Hessian 矩阵.Xiong 等^[14]通过计算模糊 c -means 算法的参数映射的 Hessian 矩阵来执行最优测试. Hessian 矩阵的公式太复杂,无法分析收敛性,雅可比矩阵具有与 Hessian 矩阵相同的能力来判断算法是否收敛到局部最大值.Chaomurilige 等^[15]提供了一种使用雅可比矩阵分析在 Gustafson-Kessel 算法中选择模糊性参数的理论方法,已经证明,雅可比矩阵对于收敛分析来说是一个更有效的工具,其用于通过使无意义的固定点处的收敛速率不小于 1 来选择算法的参数,但是没有关于如何使用雅可比矩阵选择 β_0 参数的 DAAEM 或 DAEM 算法的研究,本文为 NMEM 算法的初始参数 β_0 提供一个理论参数选择方法.NMEM 算法的收敛速率是基于后验概率的雅可比矩阵计算的,根据收敛速度分析固定点的稳定性来获得参数 β_0 的理论下限,根据该理论下限值设定初始参数 β_0 ,实验通过将理论下限向上舍入到最接近的十分之一来选择 β_0 .

本文使用 DAAEM 算法来估计网络探索混合模型参数,设计一个方法来获得基于雅可比矩阵分析的 NMEM 算法的理论最小 β_0 .

1 混合模型的 EM 算法

我们首先描述在网络中探索性分析的混合模型.然后推导其参数估计(EMMM)在定向网络中的 EM 算法^[9].

基于混合模型的结构检测^[9]旨在通过将模型拟合到观察的网络来推导网络中节点的分配.这里我们专注于有向网络的混合模型,它很容易扩展到无向或加权网络的情况.如果存在从节点 i 到节点 j 的边缘,则具有 N 个节点的网络由邻接矩阵 A 表示,元素 $A_{ij} = 1$,否则为 0.假设网络的节点落入 c 个社区并且模型参数被指定为三元组 $(\{g_i\}, \{\pi_r\}, \{\theta_{ri}\})$,其中隐藏变量 g_i 表示节点 i 的组分配, π_r 节点的分数在组 r 中, θ_{ri} 是从组 r 的节点到节点 i 存在有向边的概率.

模型参数 $\{\pi_r\}$ 和 $\{\theta_{r,i}\}$ 满足归一化条件:

$$\sum_{r=1}^c \pi_r = 1, \sum_{i=1}^N \theta_{r,i} = 1.$$

假设网络的边缘由底层概率分布的混合生成. 从 i 到 j 的每个边都是独立生成的. 首先, 边 $\langle i, j \rangle$ 的开始节点 i 以概率 π_{g_i} 选择其组 G_i . 然后, 组 G_i 中的节点 i 通过概率 $\theta_{r,i}$ 链接到末端节点 j . 观察到的网络 A 的可能性可以写为

$$\begin{aligned} \Pr(A, g | \pi, \theta) &= \Pr(A | g, \pi, \theta) \Pr(g | \pi, \theta) \\ \Pr(A | \pi, \theta) &= \prod_{ij} \Pr(A_{ij} | \pi, \theta) = \\ &= \prod_i \sum_{r=1}^c [\pi_r \prod_j (\theta_{r,j})^{A_{ij}}] \end{aligned} \quad (1)$$

在公式(1)中似然的对数是

$$M = \log \Pr(A | \pi, \theta) = \sum_i \log \sum_{r=1}^c [\pi_r \prod_j (\theta_{r,j})^{A_{ij}}] \quad (2)$$

我们经常使用 EM 算法通过最大化对数似然 M 来估计参数, M 的下限由等式(2)的 Jensen 不等式公式来计算. 标记为 \bar{M} .

$$\bar{M} \sum_{ir} q_{ir} [\log \pi_r + \sum_j A_{ij} \log \theta_{rj}] - \sum_{ir} q_{ir} \log q_{ir} \quad (3)$$

式中, q_{ir} 为节点 i 属于集群 r 的后验概率, 满足归一化条件 $\sum_r q_{ir} = 1$.

通过引入拉格朗日乘子以在 E 步中合并约束 q , 目标函数变为

$$\bar{M}' = \bar{M} + \lambda (1 - \sum_r q_{ir}) \quad (4)$$

令在 E 步中 \bar{M}' 的导数为零, 用于 q_{ir} 的更新方程推导如下:

$$q_{ir} = \frac{\Pr(A_i, g_i = r | \pi, \theta)}{\Pr(A''_i | \pi, \theta)} = \frac{\pi_r \prod_j \theta_{r,j}^{A_{ij}}}{\sum_s \pi_s \prod_j \theta_{s,j}^{A_{ij}}} \quad (5)$$

通过引入拉格朗日乘数以合并归一化条件 $\sum_r \pi_r = 1$ 和 $\sum_j \theta_{rj} = 1$, 目标函数变为

$$MM = \bar{M} + \alpha (1 - \sum_r \pi_r) + \sum_j \beta_j (1 - \sum_j \theta_{rj}) \quad (6)$$

通过使 π_r 或 θ_{rj} 的 MM 导数在 M 步中为 0, 获得模型参数的更新方程如下:

$$\pi_r = \frac{1}{N} \sum_i q_{ir}, \theta_{rj} = \frac{\sum_i A_{ij} q_{ir}}{\sum_i d_i q_{ir}} \quad (7)$$

式中, d_i 表示节点 i 的偏度.

根据 EM 算法的框架, 混合模型的 EM 算法更新 q 和 π, θ 由方程 (5) 和 (7) 迭代直到算法收敛.

2 混合模型的确定性抗退火期望最大化算法

为了分析 EM 算法在一般结构检测任务中的混合模型的收敛性质, 我们在几个小型网络上运行 EM 算法 80 次后测试收敛对数似然值, 包括 karate 网络, dolphin 网络和邻接网络. EM 算法重新启动 80 次. 每个 EM 算法迭代, 直到两个连续迭代之间的对数似然的相对误差小于阈值 10^{-10} . 测试表明, 混合模型的 EM 算法总是收敛到几个局部最优. 实验结果列于表 1~2 中.

表 1 运行 EM 算法 80 次的测试结果

log-likelihood	average-iterations	times
-499.657 3	62	40
-491.1758	66	14
-501.9945	36	20

表 2 重启后运行 EM 算法 80 次的测试结果

network	log-likelihood	average-iterations	times
dolphin	-1 211,648 6	126	80
karate	-499,554 8	77	80
adjacency	-3 544,663 8	397	80

确定性抗退火期望最大化(DAAEM)算法是关于最大似然估计的一种知名算法. 它不仅解决了传统 EM 算法中的局部最优问题, 而且提高了收敛速度, 适用于估计混合模型的参数. 作为 DAEM 算法的框架^[10], NMEM 算法引入了对应于反温度的参数 β , 传统的 EM 算法是 NMEM 算法在温度 $\beta = 1$ 时的特殊情况, NMEM 算法引进了 Eq.(5) 中 q 的后验概率中的温度参数 β , 那被 β 参数化的新的后验概率可以重写为

$$p_{ir} = \frac{(\pi_r \prod_j \theta_{r,j}^{A_{ij}})^\beta}{\sum_s (\pi_s \prod_j \theta_{s,j}^{A_{ij}})^\beta} \quad (8)$$

式中, $1/\beta$ 为相应的温度.

当后验概率已知时, 使用该后验概率以与 EM 算法相同的方式来估计模型参数. 算法从 $\beta = \beta_0 \ll 1$ (高温) 开始, 并缓慢增加 β 直到 $\beta_{\max} > 1$, 然后减小回

到 $\beta = 1$. 在最终温度下, NMEM 算法迭代直到收敛, NMEM 算法流程如下算法 2.1 所示.

算法 2.1 网络探索混合模型的确定性抗退火 EM 算法

初始化 $\beta = \beta_0$ ($0 < \beta_0 \ll 1$) 并且设定初始参数

$$\Theta^{(0)} = (\pi^{(0)}, \theta^{(0)});$$

设置 $T, t = 0$;

迭代直至收敛

- E-step: 计算估计后验概率

$$P_{ir} = \frac{(\pi_r \prod_j \theta_{ij}^{\Delta_{ij}})^{\beta}}{\sum_s (\pi_s \prod_j \theta_{ij}^{\Delta_{ij}})^{\beta}} \quad (9)$$

- M-step: 计算估计 Θ^{new}

$$\sum_i \theta_{ij} = \frac{\sum_i A_{ij} q_{ir}}{\sum_i d_i q_{ir}} \quad (10)$$

增加 β

如果 $\beta \leq \beta_{\max}, t = t + 1$,

返回到步骤 3

降低 β

迭代直至收敛

- E-step: 估计式(9)后验概率.

- M-step: 估计式(10) Θ^{new} .

如果 $\beta \geq 1, t = t + 1$, 返回到步骤 6

停止程序

在 3 个网络中 NMEM 算法重新启动 80 次. 表的标签与表 1 相同.

从结果中, 我们发现 EM 算法收敛到的局部最优较多, 特别是当一个网络变得更复杂的时候. 这表明混合模型的 EM 算法总是收敛到许多差的局部最大值^[16].

3 NMEM 算法基于收敛速率的参数选择方法

本文设计一种方法来获得基于雅可比矩阵分析的 NMEM 算法的理论最小 β_0 . 选择参数的原则是使聚类算法在无意义群集固定点上变得不稳定, 而一种算法的稳定性可以用它的收敛速率测量. 本文方法首先基于由 Jacobain Matrix 提出的 NMEM 算法的收敛速率, 在无意义群集固定点上计算之后的收敛速率, 当收敛速率小于 1 时, 就可以得出一个理论上较小的 β_0 值范围. 任何小于这个范围的 β_0 值都符合标准, 可以使得 NMEM 算法避免无意义的群集结果.

NMEM 算法的分割矩阵符合后加入概率 $\{p_{ir}\} (i \in (1, \dots, N), r \in (1, \dots, c))$, 通过分割矩阵, 我们很容易判断一个群集结果是否有意义, 进

而通过分割矩阵来定义两个形式化的无意义群集结果.

无意义分区: $U_{\pi} = [\pi_1, \pi_2, \dots, \pi_c]^T \otimes \mathbf{1}_{1 \times N}$ 无意义分区, 其中 \otimes 表示克罗内克积, $\mathbf{1}$ 表示所有 1 的向量. 这个分区表示所有节点属于具有相同概率 π_r 的聚类 $r (r = 1, \dots, c)$.

绝对无意义分区: $U_{\frac{1}{c}} = \left[\frac{1}{c} \right]_{c \times N}$ 被称为绝对无意义分区. 此分区表示所有节点属于具有相同概率的每个聚类 $\frac{1}{c}$ 绝对无意义分区是无意义分区的一种特殊况.

本文提供如何在直接网络事件中计算出 NMEM 算法的收敛速率, 这种计算可以很容易扩展到间接网络中. 通常有两种方法来计算收敛速率, 海森矩阵和雅可比矩阵, 在海森矩阵^[17-18]中, 需要计算出在 NMEM 算法中利用方程(7)全部模型参数的海森矩阵. 海森矩阵不同公式的参数增加了分析收敛速度的难度. 雅可比矩阵同样具有分析算法的收敛性质的能力, 而且雅可比矩阵比海森矩阵更容易计算. Chaomurilige 等^[15] 已经开始使用雅可比矩阵来计算 GK 算法的收敛速率.

根据 Olver^[19] 的推论, 如果算法收敛到稳定的固定点 p^* , 则在点 p^* 处的雅可比矩阵的谱半径 (或最大特征值) 应当小于 1.

本文使用雅可比矩阵来计算出 NMEM 算法的收敛速率.

雅可比矩阵通项公式为

$$f'(p^*) = \left. \frac{\partial f^{(a\beta)}(p)}{\partial p} \right|_{p=p^*} \quad (11)$$

式中, $\{p^{(0)}, p^{(1)}, p^{(2)}, \dots, p^{(n)}, \dots, p^{(*)}\}$ 是一个参数收敛序列. 通过公式(9), 可以得到雅可比矩阵.

为了得到雅可比矩阵, 首先需要明确 NMEM 算法的参数映射.

通过以下方法, NMEM 算法的迭代公式可以是映射 $p^{t+1} = F^{(a\beta)}(p)^{(t)}$.

$$p_{ir}^{(t+1)} = F^{(a\beta)}(p_{ir}^{(t)}) = \frac{(\pi_r^{(t+1)} \prod_j (\theta_{rj}^{(t+1)}) A_{ij})^a}{\sum_s (\pi_s^{(t+1)} \prod_j (\theta_{rj}^{(t+1)}) A_{ij})^{\beta}} \quad (12)$$

$$\pi_r^{(t+1)} = \frac{1}{N} \sum_i p_{ir}^{(t)} \quad (13)$$

$$\theta_{rj}^{(t+1)} = \frac{\sum_i A_{ij} p_{ir}^{(t)}}{\sum_i d_i p_{ir}^{(t)}} \quad (14)$$

$\{p^{(t)}, t = 0, 1, \dots, n, \dots\}$ 是 NMEM 算法的一个收敛序列. 这个序列起始于点 p^0 , 而且将会收敛到一个点 p^* , 这个收敛点是一个固定点, 也就是说 $p^* = F^{(\alpha\beta)}(p^*)$.

根据雅可比矩阵在 $(\partial F)^{(\alpha\beta)}(p_{ir}/\partial p_{js})|_{p=p^*}$ 的点上的最大特征值不小于 1, 我们可以得出结论: 算法不能稳定地收敛到这些无意义的点上. NMEM

算法有着一个限制, 即 $\sum_{i=1}^c p_{ir} = 1$, 之后是 $p_{ic} = 1 =$

$\sum_{i=1}^{c-1} p_{ir}$. 雅可比矩阵 $[\partial F^{(\alpha\beta)}(p_{ir}/\partial p_{js})]_{ijrs}$ 的容积可表示为

$$(c-1) \times (c-1) \times N \times N \times N$$

$$(i, j \in \{1, \dots, N\}, r, s \in \{1, \dots, c-1\}),$$

其 $(r-1) \times N + i$ 行和 $(s-1) \times N + j$ 列处的元素矩阵可以由引理 3.1 中的统一表达式表示.

这里关注的是从理论上说明初始参数设置为相等的情况, 即 $\alpha \neq \beta$, 后期我们进一步讨论两个参数不相等的情况, 被加入了随机扰动的起始值.

引理 3.1

当 $i = 1, \dots, N, r = 1, \dots, c-1$ 且 $j = 1,$

$\dots, N, s = 1, \dots, c-1$ 时, 雅可比矩阵 $\frac{\partial F^{(\beta)}(p)}{\partial p}$ 的

每一个要素 $\frac{\partial F^{(\beta)}(p_{ir})}{\partial p_{js}}$ 都可以被以下方程式运算:

$$\begin{aligned} \frac{\partial F_{ir}^{(\beta)}}{\partial p_{js}} &= \beta \times \delta_{rs} \left(\frac{p_{ir}}{N_{\pi r}} + p_{ir} \sum_{l=1}^N \frac{A_{il}}{\theta_{rl}} \frac{A_{jl} - d_j \theta_{rl}}{\sum_{k=1}^N d_k p_{kr}} \right) - \\ & p_{ir} \times \beta \times \left(\frac{p_{ir}}{N_{\pi s}} + p_{ir} \sum_{l=1}^N \frac{A_{il}}{\theta_{sl}} \frac{A_{jl} - d_j \theta_{sl}}{\sum_{k=1}^N d_k p_{ks}} \right) + \\ & p_{ir} \times \beta \times \left(\frac{p_{ic}}{N_{\pi c}} + p_{ic} \sum_{l=1}^N \frac{A_{il}}{\theta_{cl}} \frac{A_{jl} - d_j \theta_{cl}}{\sum_{k=1}^N d_k p_{kc}} \right) \end{aligned} \quad (15)$$

$$\delta_{rs} = \begin{cases} 1, & r = s \\ 0, & r \neq s \end{cases}, \text{ 和 } \frac{\partial \theta_{rl}}{\partial p_{js}} = \frac{A_{jl} - d_j \theta_{rl}}{\sum_{k=1}^N d_k p_{kr}}$$

为是克罗内克 delta 函数.

4 实验

本文通过实验验证由 NMEM 算法提出的参数选择方法计算的初始参数的理论下限 β 理论, 进行 NMEM 算法的性能分析, 从 β_0 开始, 通过向上舍入

到理论到最接近的十分之一. 此外, 研究收敛速度和网络结构之间的关系.

4.1 实验集和分析

真实网络数据采用经典网络: Books on politics (美国政论著作关系网络)、Karate (Zachary 空手道俱乐部网络^[14]) 和 Football (美国大学足球赛网络) (数据采集参照 <http://www.orgnet.com/>). 选用的经典网络都有清晰的社区结构, 介绍如下: Books on politics 由 46 个节点 139 条边构成, 每个节点表示一本书, 每本书均为亚马逊网站出售的关于美国政论的著作, 两节点间的边表示顾客同时购买了两本书, 整个网络分成两派 (两种社区), 节点之间的连边表示两本书之间关联程度. Karate 由 38 个节点 220 条边组成, 每个节点表示一个俱乐部的成员, 两节点之间的边表示两成员之间的社会交往关系, 该俱乐部主要分为两个社区, 一个为俱乐部主管一个为校长, 因此俱乐部成员就以俱乐部主管和校长为中心分成了两个社区. Football 由 120 个节点 1 221 条边组成, 每个节点表示一支队伍, 两节点之间的表示连接此边的两个队伍正在比赛, 全部队伍被分为 12 个联盟, 即 12 个社区, 与联盟内球队进行比赛的概率大于与联盟外球队比赛的概率, 测试结果见 3.

表 3 真实网络上的 NMEM 算法的 λ_{\max} , $\beta_{\text{theoretical}}$ 和 β_0

Tab.3 Values of λ_{\max} , $\beta_{\text{theoretical}}$ and β_0 by NMEM algorithm on real network

network	$\lambda_{\max}(B)$	$\beta_{\text{theoretical}}$	β_0
Polbook	13.527	0.089 1	0.1
Risk	4.395 7	0.275	0.3
Karate	7.509	0.161 2	0.2
Dolphin	6.903	0.175 8	0.2
Wiki-Vote	90.280 9	0.013 2	0.1
socSlash	115.251	0.012	0.1
Web	193.586	0.006 5	0.1
Football	8.821	0.138	0.2
Lesmis	11.478	0.106	0.1
Adjnoun	7.171	0.168 7	0.2

所有算法在满足停止标准时终止: $|L(\theta^{t+1} - L(\theta^T))|/L(\theta^{t+1}) < \tau$, 其中 (θ^T) 代表的对数似然值. 在第 t 次迭代, τ 是容差参数. 公差设置为 10^{-6} . 所有测试执行 10 次试验, 并使用平均值. 所有三种算法的初始模型参数是随机给出的. 根据文献[11]的经

验, NMEM 算法中的 β_{\max} 被设置为 1.15. 我们还测试了不同 $\beta_{\max} = \{1.2, 1.3, 1.4, 1.5\}$ 的影响. 在 $\beta_{\max} = 1.15$ 时, 算法总是收敛的. 从 10 个运行以相同的 β_0 开始, 并且不同网络上的所有 β_0 根据表 3 中的 β_0 设置. 在 NMEM 算法中, 将 β 更新为 $\beta_{\text{new}} \leftarrow \beta_{\text{last}} + 0.1$, 其中 β_{new} 在当前迭代中为 β , β_{last} 在最后一次迭代中为 β .

此外, 表 3 中的结果表明不同的网络具有不同的 $\beta_{\text{theoretical}}$. NMEM 算法的初始 β_0 不应小于对应的 $\beta_{\text{theoretical}}$. 表 3 中最后一列的值 β_0 大于对应的 $\beta_{\text{theoretical}}$, 这是我们在实验中使用的初始 β_0 . 如果 β 理论值 < 0.1 , 我们将其舍入到最接近的十分之一, 并且初始 $\beta_0 = 0.1$.

4.2 NMEM 算法的性能

为了演示具有适当 β_0 的 NMEM 算法的性能, 我们执行两种测试. 首先, 我们调查与高于海豚, Adjnoun, 足球网的网络的理论值 $\beta_{\text{theoretical}}$ 不同的初始 β_0 值的影响. 其次, 几种算法进行比较.

图 1 展示出了以高于理论下限的不同 β_0 开始的 NMEM 算法的准确度. 它表明所选择的 β_0 在 $\{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ 中具有最佳性能. 根据我们的参数选择方法, 三个网络上的算法的初始 β_0 都被选择为 0.2. 对于 Book 网络, 在所选 $\beta_0 = \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ 处的与在 β_0 处的 NMEM 算法相同的精度. 这是因为 Book 网络的分类简单且易于检测. Adjnoun 网络比 Book 网络更复杂. 在 Adjnoun 网络上的 $\beta_0 = \{0.6, 0.7, 0.8, 0.9\}$ 处的精度值低于 $\beta_0 = \{0.2, 0.3, 0.4, 0.5\}$ 处的精度值. 足球网络是 3 个网络中最复杂的, 其在 β_0 的算法具有比其他 β_0 更好的性能. 随着初始 β_0 增加, 精度值开始减小. 原因是复杂网络的目标函数具有更多的局部最大值, 较大的初始 β_0 使得算法容易更快地收敛到局部最大值, 因此对于复杂网络的初始 β_0 值, NMEM 算法的性能随着 β_0 的增加而减小. 总而言之, 根据我们的参数选择方法从选择的 β_0 开始的 NMEM 算法具有更好的性能.

图 1 的比较结果表明, 具有不同 β_0 的 NMEM 算法的精度高于 3 个网络上的理论值. 每个曲线和误差栏分别代表 10 次试验的平均精度和标准偏差.

5 结论

为了有效地探索网络的结构, 使用确定性抗退火期望最大化算法来估计混合模型的参数. 根据提

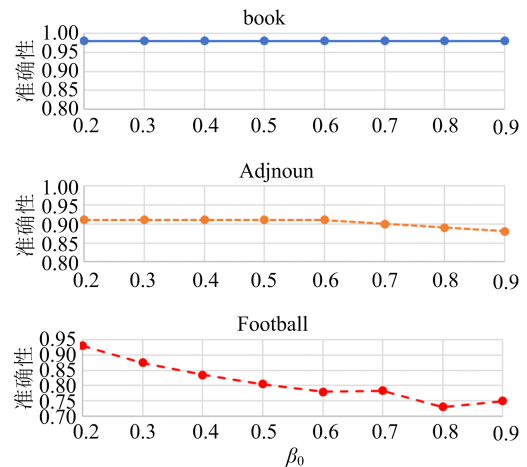


图 1 3 个网络的准确性比较

Fig.1 Accuracy comparison of three networks

供的参数选择方法的理论下限选择初始 β_0 , 由基于雅可比矩阵的算法的收敛速率计算. 只要 β_0 不小于下限, NMEM 算法将不会收敛到定义的无意义结果. 同时, 我们使用的 β_0 不应该大于下限; 否则算法将更频繁地收敛到局部最优. 本文实验中, β_0 是下限被四舍五入到最接近 10 的值. 对合成网络和实际网络的实验结果表明, NMEM 算法和提出的参数选择方法是有效的, NMEM 算法的收敛速度受网络程度的分类混合的影响.

通过参数选择任务的收敛速率分析算法以进行网络探索, 它很容易扩展到其他算法的社区发现. 在未来, 我们将进一步研究初始参数设置为不相等的情况 $\alpha \neq \beta$, 以及基于我们提出的方法的社区发现的其他算法.

参考文献 (References)

- [1] FORTUNATO S. Community detection in graphs[J]. Physics Report, 2010, 486(3): 75-174.
- [2] HOLLAND P W, LASKEY K B, LEINHARDT S. Stochastic blockmodels: First steps [J]. Social Networks, 1983, 5(2): 109-137.
- [3] SNIJDERS T A B, NOWECKI K. Estimation and prediction for stochastic blockmodels for graphs with latent block structure [J]. Journal of Classification, 1997, 14 (1): 75-100.S
- [4] DAUDIN J J, PICARD F, ROBIN S. A mixture model for random graphs [J]. Statistics and Computing, 2008, 18(2): 173-183.
- [5] SHEN W H, CHENG X Q, GUO J F. Exploring the structural regularities in networks [EB/OL]. [2011-05-04]. <https://doi.org/10.1103/PhysRevE.84.056111>.

- [6] CHAI B F, YU J, JIA C Y, et al. Combining a popularity-productivity stochastic block model with a discriminative-content model for general structure detection [J]. *Physical Review E*, 2013, 88 (1): 012807.
- [7] LATOUCHE P, BIRMELE E, AMBROISE C. Variational Bayesian inference and complexity control for stochastic block models[J]. *Statistical Modelling*, 2012, 12(1): 93-115.
- [8] DECELLE A, KRZAKALA F, MOORE C, et al. Inference and phase transitions in the detection of modules in sparse networks [J]. *Physical Review Letters*, 2011, 107(6): 065701.
- [9] YANG B, HE H, HU X. Detecting community structure in networks via consensus dynamics and spatial[J]. *Statistical Mechanics and its Applications*, 2017, 483(1): 156-170.
- [10] UEDA N, NAKANO R. Deterministic annealing em algorithm[J]. *Neural Network*, 1998, 11(2): 271-282.
- [11] NAIM I, GILDEA D. Convergence of the EM algorithm for Gaussian mixtures with unbalanced mixing coefficient [C]// *Proceedings of the 29th International Conference on Machine Learning*. Edinburgh, UK: ACM Press, 2012: 1427-1431.
- [12] VON LUXBURG U, BEN-DAVID S. Towards a statistical theory of clustering[M]// *Pascal Workshop on Statistics and Optimization of Clustering*, 2005.
- [13] XU L, JORDAN M I. On convergence properties of the EM algorithm for Gaussian mixtures[J]. *Neural Computation*, 1996, 8 (1): 129-151.
- [14] XIONG H, SHANG P. Weighted multifractal cross-correlation analysis based on Shannon entropy [J]. *Science and Numerical Simulation*, 2016, 30 (1): 268-283.
- [15] CHAOMURILIGE C, YU J, YANG M S. Analysis of parameter selection for Gustafson-Kessel fuzzy clustering using Jacobian matrix [J]. *IEEE Transactions on Fuzzy Systems*, 2015, 23 (6): 2329-2342.
- [16] WU C F J. On the convergence properties of the EM algorithm[J]. *Annual Statistics*, 1983, 11(1): 95-103.
- [17] 陈宇, 许莉薇, 江露, 等. 成像流型辨识算法[J]. *哈尔滨理工大学学报*, 2014, 19(4): 111-116.
CHEN Y, XU L W, JIANG L, et al. Electrical capacitance tomography identification algorithm based on GMM model[J]. *Journal of Harbin University of Science and Technology*, 2014, 19(4): 111-116.
- [18] MA J W, FU S Q. On the correct convergence of the EM algorithm for Gaussian mixtures [J]. *Pattern Recognition*, 2005, 38(12): 2602-2611.
- [19] OLVER P J. Lecture notes on numerical analysis[EB/OL]. [2008-05-18]. <http://www.math.umn.edu/olver/num.html>.