

# 基于高斯噪声模型的马尔可夫网络构建算法

杨 博, 张军英

(西安电子科技大学计算机学院, 陕西 西安 710071)

**摘要:** 针对小样本集构建稀疏马尔可夫网络计算量大和求解精度不高的问题, 提出一种基于高斯噪声模型的迭代噪声消减(iterative noise reduction, INR)算法。该算法首先利用回归误差的高斯特性筛选相关变量, 然后通过 boosting 方法的自回归更新策略逐步改进学习能力, 最后采用赤池信息准则(Akaike information criterion, AIC)避免出现过拟合。此外, 给出了自回归更新公式, 实现了可控的学习错误率并分析了计算复杂度。实验结果表明, INR 能有效构建高维稀疏网络, 在学习效率和精度方面具有明显优势。

**关键词:** 人工智能; 迭代噪声消减; 网络推理; 马尔可夫网络; 高斯噪声

**中图分类号:** TP 18

**文献标志码:** A

**DOI:** 10.3969/j.issn.1001-506X.2012.05.33

## Gaussian noise model based algorithm to construct Markov network

YANG Bo, ZHANG Jun-ying

(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

**Abstract:** To solve the difficulties of high calculation quantity and low precision in constructing sparse Markov network with a small set of samples, an iterative noise reduction (INR) algorithm based on the Gaussian noise model is proposed. The algorithm firstly picks out the related variables through employing statistic test to regression residuals. After that, a learning ability is gradually improved through the autoregressive update strategy similar as boosting method. Finally, Akaike information criterion (AIC) is used to avoid overfit. In addition, the iterative update formula is provided and the error rate controlling is realized. Furthermore, the computational complexity of the proposed algorithm is analyzed. The experimental results show that INR can effectively construct the high dimensional sparse network and has obvious advantages on learning precision and efficiency.

**Keywords:** artificial intelligence; iterative noise reduction (INR); network inference; Markov network; Gaussian noise

## 0 引 言

马尔可夫网络描述多变量间的复杂关系, 是机器学习、人工智能等领域的一个重要模型。典型的马尔可夫网络包括高斯图模型、马尔可夫随机场等, 它们被广泛应用于多元变量数据分析、决策支持系统、图像处理、因果推理等领域。目前, 通过数据挖掘方法构建稀疏马尔可夫网络是非常热门的研究问题<sup>[1]</sup>。

近年来, 求解该问题出现了许多新颖的算法, 例如, 应用随机搜索算法进行变量选择并构建马尔可夫网络<sup>[2-3]</sup>, 这种方法虽然也能生成网络关系图, 但计算复杂度大于采用 L1 范数作为罚项的 lasso 方法<sup>[4]</sup>。lasso 方法比岭回归<sup>[5]</sup>更倾向得到稀疏解, 但它对数据的协方差矩阵较敏感, 稳定性

和鲁棒性较差, 时间复杂度较高。由于样本容量受制于实验技术和样本供体, 许多研究需要从小样本高维数据进行网络结构推理。文献[6]提出的 glasso 通过矩阵分块坐标下降法提高马尔可夫网络学习的效率, 但这种优化技术往往得不到最优解, 导致求解精度低。文献[7]则提出 Boosti-Graph 算法构建高维稀疏网络, 采用回归方法生成邻接矩阵, 但迭代过程中易产生过拟合, 解的稳定性差。此外, 核关联分析<sup>[8-9]</sup>、贝叶斯推理<sup>[10]</sup>、进化计算<sup>[11]</sup>等方法进行网络构建大多只能应用于小样本低维数据, 处理高维数据时计算量很大, 往往不能有效完成学习任务。

当前, 通过小样本集推理高维马尔可夫网络仍然比较困难, 主要是算法求解精度不高, 运行时间较长, 不能很好地改善结构估计的错误率。本文假设数据具有独立高斯分

收稿日期: 2011-06-13; 修回日期: 2012-03-01。

基金项目: 国家自然科学基金(61070137, 60371044); 国家自然科学基金重点项目(60933009)资助课题

作者简介: 杨博(1978-), 男, 讲师, 博士研究生, 主要研究方向为机器学习与数据挖掘。E-mail: boyang@mail.xidian.edu.cn

布的加性噪声,较合理地抽象高维样本的随机特性和测量误差,提出一种基于高斯噪声模型的新算法来分析误差的非高斯性,并结合 boosting 自回归更新及 AIC 约束两种策略实现快速构建稀疏马尔可夫网络,提高了学习精度,降低了错误率。

### 1 问题描述

构建马尔可夫网络是指在给定样本集下,寻找一个能最好匹配该样本集的网络结构。该结构用马尔可夫网络来表示,并且用边来反映节点(对应数据的属性变量)间潜在的概率依赖关系或条件独立关系<sup>[12]</sup>。

对于  $n$  行  $p$  列观测数据  $\mathbf{X}=(x_1, x_2, \dots, x_p)$ ,假设满足多元变量高斯分布  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,即  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,且样本间独立同分布。通过马尔可夫网络  $G=(\mathbf{V}; \mathbf{E})$  表示  $\mathbf{X}$ ,则马尔可夫网络中每个节点对应一个变量,  $\mathbf{V}$  表示所有节点的集合,马尔可夫网络的边代表节点间交互作用或条件依赖关系,  $\mathbf{E}$  表示所有边的集合<sup>[13]</sup>。其中,  $n$  是观测样本的个数,  $p$  是变量或节点的个数,  $\boldsymbol{\mu}=(\mu_1, \dots, \mu_p)^T$  是均值向量,  $\boldsymbol{\Sigma}$  是协方差矩阵。为叙述方便,以下不区分变量和节点。

为学习节点的连接关系即推理邻接矩阵  $\mathbf{A}$ ,往往需要计算节点间的偏相关系数  $\rho$ ,它用来衡量节点间交互作用的大小。对于前述观测样本  $\mathbf{X}$ ,若已知精度矩阵  $\boldsymbol{\Omega}=\boldsymbol{\Sigma}^{-1}=(\omega_{ij}), 1 \leq i, j \leq p$ ,则  $\rho$  可表示为<sup>[14-15]</sup>:

$$\rho_{ij} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \quad (1)$$

这种计算  $\rho$  的方法需要首先估计逆协方差矩阵,然而,在高维数据下( $n \ll p$ ),样本协方差矩阵往往是病态或奇异矩阵,计算它的逆矩阵产生较大的均方误差。

为规避以上问题,提高学习精度并降低计算复杂度,本文采用文献[16]提出的式(2),通过回归系数  $\beta$  求偏相关系数  $\rho$ :

$$\rho_{ij} = \begin{cases} \text{sgn } \beta_j^i \cdot \min \{1, \sqrt{\beta_j^i \beta_i^j}\}, & \text{sgn } \beta_j^i = \text{sgn } \beta_i^j \\ 0, & \text{其他} \end{cases} \quad (2)$$

式中,  $\rho_{ij}$  表示节点  $i$  和  $j$  间的偏相关系数;  $\beta_j^i$  表示  $i$  作为自变量对  $j$  进行线性回归得到的系数;  $\text{sgn}$  表示符号函数。上述公式避免了高维情况下  $\beta_j^i$  和  $\beta_i^j$  可能异号的问题,并保证  $\rho_{ij} \in [-1, 1]$ 。

求得偏相关系数后,根据文献[17]提出的式(3)计算邻接矩阵  $\mathbf{A}$ 。

$$\mathbf{A} = \{a_{ij} \mid a_{ij} \in \{0, 1\}\} \\ a_{ij} = \begin{cases} 0, & \rho_{ij} = 0 \\ 1, & |\rho_{ij}| > 0 \end{cases} \quad (3)$$

本文所提算法通过线性回归拟合求得回归系数并计算偏相关系数,最后生成稀疏马尔可夫网络,详细过程在下节进行阐述。

## 2 基于高斯噪声模型的迭代噪声消减算法

### 2.1 高斯噪声模型

**定义 1** 高斯噪声模型。它是考虑高斯噪声影响的线性回归模型。对于变量  $x_i$ ,假设存在高斯噪声的影响,则变量  $x_j$  对变量  $x_i$  的线性回归模型为

$$x_i = \beta_{ij} x_j + \epsilon \quad (4)$$

式中,  $\beta_{ij}$  是变量  $x_j$  对变量  $x_i$  的回归系数;高斯噪声  $\epsilon \sim N(0, \sigma^2)$ ,满足独立高斯分布且噪声方差未知,  $\sigma^2$  的初始估计是观测数据的样本方差。

在高斯噪声模型假设下,若变量没有交互作用,则线性回归误差完全是高斯性噪声;当变量具有交互作用时,则回归误差除了高斯性噪声外,还包含非高斯性噪声。因此,对回归误差进行高斯性检验,并通过迭代更新过程筛选引发非高斯性噪声的相关变量是本文算法的核心。

### 2.2 回归误差统计检验

本文通过检验回归误差  $x_i - \hat{x}_i$  的均值  $\bar{\epsilon}_i$  是否服从正态分布  $N(\mu_0, \sigma^2)$  来判定回归误差的高斯性。因此,可作假设  $H_0: \mu = \mu_0; H_1: \mu \neq \mu_0$  并检验统计量  $T = \frac{\bar{\epsilon}_i - \mu_0}{\sigma/\sqrt{n}}$  是否服从自由度为  $n-1$  的  $t$  分布。对于给定的显著性水平值  $\alpha$ ,则

$$P \left\{ \left| \frac{\bar{\epsilon}_i - \mu_0}{\sigma/\sqrt{n}} \right| \geq t_{\alpha/2}(n-1) \right\} = \alpha$$

若  $|\bar{\epsilon}_i - \mu_0| \geq t_{\alpha/2}(n-1)\sigma/\sqrt{n}$ ,则拒绝假设  $H_0$ ;若  $|\bar{\epsilon}_i - \mu_0| < t_{\alpha/2}(n-1)\sigma/\sqrt{n}$ ,则接受假设  $H_0$ 。本文中  $\mu_0 = 0, \sigma$  是样本标准差,  $t_{\alpha/2}(n-1)$  是自由度为  $n-1$  的  $t$  分布在显著性水平为  $\alpha$  时的上侧分位数。

对回归误差进行高斯性检验的具体方法如下:

(1) 对  $n \times p$  维数据矩阵  $\mathbf{X}$ ,标准化各维变量  $x_i(1 \leq i \leq p)$ ,使其均值为 0 且标准差为 1,即  $\bar{x}_i^{(0)} \leftarrow 0, \sigma_i^{(0)} = 1$ 。

(2) 基于高斯噪声模型进行单变量回归,即变量  $x_i$  作为响应变量,变量  $x_j$  作为预测变量,回归拟合得到系数  $\hat{\beta}_{ij} = x_i^T x_j (i \neq j, 1 \leq i, j \leq p)$ 。

(3) 计算  $x_i$  的估计值  $\hat{x}_i = \hat{\beta}_{ij} x_j$ ,对回归误差  $x_i - \hat{x}_i$  的均值  $\bar{\epsilon}_i$  进行  $t$  检验。若  $P$  值大于  $\alpha$ ,则接受假设  $H_0$ ,说明误差是由高斯噪声引起,可以判定变量  $x_i$  和  $x_j$  没有关联,马尔可夫网络相应节点间没有边;若  $P$  值小于  $\alpha$ ,则拒绝假设  $H_0$ ,表明误差由非高斯性噪声引起,变量  $x_i$  和  $x_j$  有相关关系,马尔可夫网络相应节点间存在边。

(4) 通过改变  $\alpha$  值,可以控制检验的效果,即学习的错误率和马尔可夫网络的稀疏度。放大  $\alpha$ ,则更多的边被包含到马尔可夫网络中,错误率增加;缩小  $\alpha$ ,则网络变得稀疏,发现的边具有更高的置信度。

### 2.3 基于 boosting 方法的自回归更新

本文算法利用变量集合  $\mathbf{S}$  更新变量  $x_i$  的值及样本标准差  $\sigma_i$ ,逐步消除非高斯噪声的影响,增强变量噪声的高斯性。这个过程也是变量选择过程,逐步从集合  $\mathbf{S}$  筛选出相

关变量并保留无关变量。具体的迭代过程基于 boosting 方法的自回归更新策略,即通过上次  $x_i$  更新下次  $x_i$  值,逐步增强变量选择能力,得到更多相关变量。算法结束时,集合  $S$  剩余变量与  $x_i$  无关,而从集合  $S$  删除的变量则与  $x_i$  有关。对所有变量进行迭代更新,可求得邻接矩阵并构建高维稀疏网络。

本文提出式(5)和式(6)作为迭代更新公式,重估  $x_i$  及样本标准差  $\sigma_i$ 。

$$x_i^{(m+1)} \leftarrow \left( x_i^{(m)} - \frac{\sum_{j \in S_i^{(m)}} \hat{\epsilon}_{ij}}{1 + |S_i^{(m)}|} \right) \quad (5)$$

$$\sigma_i^{(m+1)} = \sqrt{\frac{\sum (x_i^{(m+1)} - \bar{x}_i^{(m+1)})^2}{n-1}} \quad (6)$$

式中,  $m$  是迭代步数;集合  $S$  包含未检验变量或与  $x_i$  无关

的变量;  $\hat{\epsilon}_{ij} = x_i - \hat{\beta}_{ij}x_j$ ,  $\bar{x}_i^{(m+1)} = \frac{1}{n} \sum_{k=1}^n x_{ki}^{(m+1)}$ 。

**定理 1** 迭代更新定理。设  $S_i^{(0)} \leftarrow \{1, \dots, p\} \setminus \{i\}$ , 利用集合  $S$  中的变量更新  $x_i$ , 则  $x_i^{(m+1)} \leftarrow \left( x_i^{(m)} - \frac{\sum_{j \in S_i^{(m)}} \hat{\epsilon}_{ij}}{1 + |S_i^{(m)}|} \right)$ 。

证明

$$\begin{aligned} x_i^{(m+1)} &\leftarrow \frac{x_i^{(m)} + \sum_{j \in S_i^{(m)}} (\hat{\beta}_{ij}x_j)}{1 + |S_i^{(m)}|} = \\ &\frac{x_i^{(m)} + \sum_{j \in S_i^{(m)}} (x_i^{(m)} - \hat{\epsilon}_{ij})}{1 + |S_i^{(m)}|} = \\ &\frac{x_i^{(m)} + \sum_{j \in S_i^{(m)}} x_i^{(m)} - \sum_{j \in S_i^{(m)}} \hat{\epsilon}_{ij}}{1 + |S_i^{(m)}|} = \\ &\frac{(1 + |S_i^{(m)}|)x_i^{(m)} - \sum_{j \in S_i^{(m)}} \hat{\epsilon}_{ij}}{1 + |S_i^{(m)}|} = \\ &x_i^{(m)} - \frac{\sum_{j \in S_i^{(m)}} \hat{\epsilon}_{ij}}{1 + |S_i^{(m)}|} \end{aligned}$$

证毕

### 2.4 算法结束条件

为避免迭代步数过多可能出现的过拟合,即筛选出过多变量及增加错误率,本文使用 AIC 约束及时停止迭代。通过逐步筛选  $x_i$  的相关变量,建立线性回归模型并计算相应 AIC 值,直到 AIC 不减小或达到最大迭代步数为止。

AIC 是一种常用的模型评价指标,用来判别和推理模型适应性。它提供一种对信息损失的测度,在构建模型时实现模型预测精度与自身复杂度间的平衡。AIC 比贝叶斯信息准则(Bayesian information criterion, BIC)具有更好的理论优势及较佳的实践性能,而且比交叉验证的时间复杂度低<sup>[18]</sup>。

AIC 运用式(7)评价模型的好坏,其中  $L$  是似然函数值,  $k$  是模型中参数的个数:

$$AIC = 2k - 2\ln(L) \quad (7)$$

AIC 准则要求其值越小越好,  $k$  值越小,  $L$  值越大,则

AIC 值越小。  $k$  小意味着模型简洁,  $L$  大意味着模型精确。因此, AIC 在评价模型时兼顾了简洁性和精确性<sup>[19]</sup>。

对有限样本集或高维数据,当样本数  $n$  小于参数  $k$ , AIC 可简化为

$$AIC = \ln \frac{RSS}{n} + \frac{n+k}{n-k-2} \quad (8)$$

式中, RSS 是模型的预测值与真实值的残差平方和,  $RSS_i = \sum_{k=1}^n \hat{\epsilon}_{ki}^2$ 。在高斯噪声模型假设下,本文采用文献[20]中提出的式(9)计算 AIC。

$$AIC = 2k + n \ln(RSS) \quad (9)$$

### 2.5 算法流程

为计算邻接矩阵  $A$  并生成高维稀疏马尔可夫网络,本文所提算法的实现步骤如下:

**步骤 1** 初始化回归系数  $\hat{\beta}_{ij} = 0$ , 回归残差  $\hat{\epsilon}_{ij} = 0$ , 显著性水平值  $\alpha = 0.05$ , 迭代步数  $m = 0$ , 最大迭代步数  $T = 2n$ 。

**步骤 2** 逐个计算  $x_j (j \neq i, j \in S_i^{(m)})$  对  $x_i (1 \leq i \leq p)$  的回归系数  $\hat{\beta}_{ij}$ , 并对回归误差  $x_i - \hat{x}_i$  的均值  $\bar{\epsilon}_i$  进行  $t$  检验。假若  $P_{ij} < \alpha$ , 则  $S_i^{(m)} \leftarrow S_i^{(m)} \setminus \{j\}$ 。

**步骤 3** 根据式(5)和式(6)更新  $x_i^{(m+1)}$  及标准差  $\sigma_i^{(m+1)}$ 。

**步骤 4** 根据筛选出的  $x_i$  相关变量构建多元线性回归模型并通过式(9)计算 AIC 值。

**步骤 5** 更新  $S_i^{(m+1)} \leftarrow S_i^{(m)}$ , 增量迭代步数  $m = m + 1$ 。如果  $m$  达到最大迭代步数  $T$  或 AIC 值不再减小,则执行步骤 6; 否则回到步骤 2。

**步骤 6** 根据式(2)和式(3),由最小 AIC 值对应模型的回归系数  $\beta_i$  计算偏相关系数  $\rho_i$  和邻接矩阵  $A$ , 并生成稀疏马尔可夫网络。

### 2.6 算法复杂度分析

本文算法的时间复杂度与迭代步数  $m$  及变量个数  $p$  紧密相关。其中,步骤 2 的时间复杂度为  $O(p(p-1))$ , 但由于马尔可夫网邻接矩阵的对称性,时间复杂度缩减为  $O(\frac{1}{2}p(p-1))$ ; 从步骤 3~步骤 6 进行迭代更新并构建多元线性回归模型的时间复杂度为  $O(m(p-1))$ , 由于最大迭代步数设为  $2n$ , 因此 INR 算法步骤 3~步骤 6 的时间复杂度是  $O(2n(p-1))$ 。因而, INR 算法在最差情况下的总的复杂度为  $O((2n+p/2)(p-1))$ 。在高维稀疏马尔可夫网中往往  $n \ll p$ 。所以,本文算法的时间复杂度得到有效的改善。

## 3 实验结果

为测试算法的性能,采用仿真及真实数据分别进行实验,并将测试结果与最新的一些网络构建算法进行比较。实验运行环境为: Windows XP 操作系统, Pentium T3400 2.16 GHz, 1 GB 内存, Matlab7.11 平台上编程实现。

### 3.1 仿真数据实验

为评价算法在高维稀疏数据集上的性能,本文采用

文献[21]提出的模型产生仿真数据。利用精度矩阵  $\Sigma^{-1}$  表示稀疏网络原始结构,在稀疏情形下,令  $(\Sigma^{-1})_{ii} = 1$ ,  $(\Sigma^{-1})_{i,i-1} = (\Sigma^{-1})_{i-1,i} = 0.5$ ,其余矩阵值全部为 0。从  $\mathbf{X} \sim N(0, \Sigma + \sigma^2 \mathbf{I})$  生成数据( $\mathbf{I}$  表示单位矩阵),并在产生的数据集中添加高斯噪声,噪声  $\sigma \sim U(0.3, 0.5)$ 。共产生 100 个数据集,每个数据集  $n = 100, p = 200$ ,生成的网络大约有 500 条边。

图 1 显示 100 个仿真数据集上 INR 算法与 glasso、BoostiGraph 算法的受试者工作特征曲线(receiver operating characteristic, ROC),图 2 显示 3 种算法的错误发现率(false discovery rate, FDR)曲线。从图 1 和图 2 看出,本文算法与 glasso 和 BoostiGraph 方法相比,提升了学习精度并使错误发现率下降,性能具有明显优势。

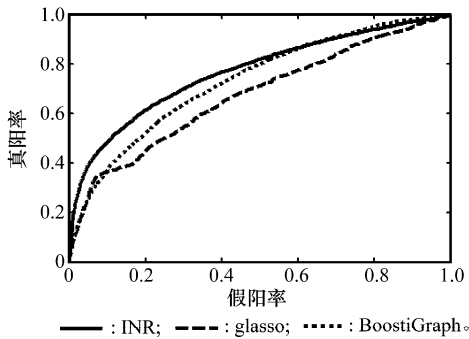


图 1 3 种算法的受试者工作特征曲线比较

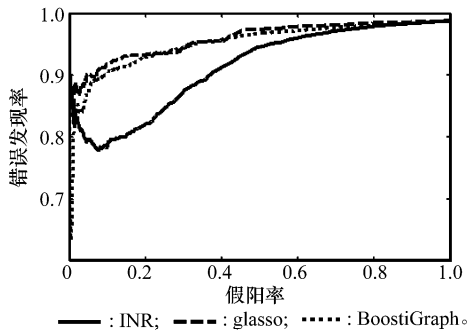


图 2 3 种算法的错误发现率比较

表 1 显示样本数 100 节点数不同时,分别独立生成 100 个数据集上 3 种算法的平均运行时间(以 s 为单位)。从表中可以看出,INR 算法在计算速度方面远高于 glasso,稍优于 BoostiGraph,具有较高的效率优势。

表 1 稀疏仿真数据集上平均运行时间比较 s

节点数/个	算法		
	INR	BoostiGraph	glasso
100	0.56	0.679	0.04
300	1.916	2.251	2.237
500	7.273	9.343	18.310
1 000	19.523	21.856	112.87
5 000	96.439	101.457	670.93

### 3.2 真实数据实验

针对小样本数据在不同维数和低假阳时算法性能评价问题,比较了 INR、glasso 和 BoostiGraph 算法在细胞信号网络数据<sup>[22-23]</sup>,大肠杆菌基因网络数据<sup>[24]</sup>和前节所述仿真数据集的 ROC 曲线下面积(area under curve, AUC)与假阳数为 5 时真阳计数。其中,细胞信号网络数据包括 100 个样本和 11 个变量;仿真数据包括 100 个样本和 200 个变量;大肠杆菌数据包括 445 个样本和 1 211 个变量。

图 3 比较了算法在不同数据集的 AUC 值,从图中可以看出,INR 算法在小样本集不同维数条件下 AUC 值均比 BoostiGraph 及 glasso 更高,INR 最小的 AUC 值也高于 0.7。这表明本文算法在低维和高维数据上都可以有效推理网络结构并且具有较高的性能优势。

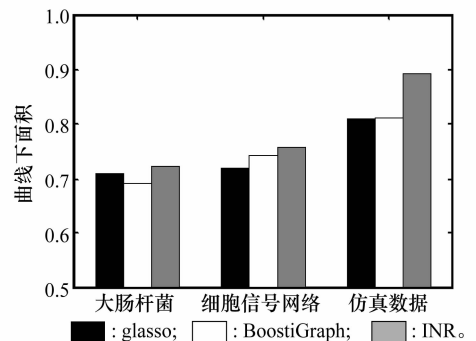


图 3 3 种算法在不同数据集上的性能比较

采用假阳数为 5 时的真阳计数来评价 3 种算法的推理性能,如图 4 所示。从图中可以看出,本文算法在不同数据集上,当假阳数较低时真阳数超过其他两种算法,表明 INR 能推理更多真实边,具有更优的推理性能。

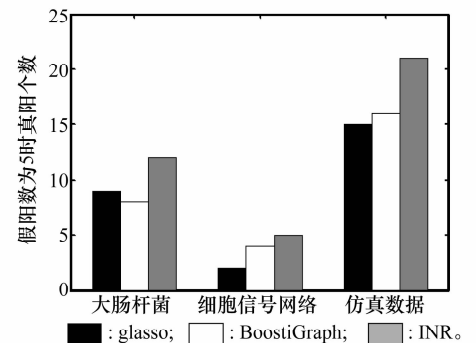


图 4 3 种算法在假阳数为 5 时的性能比较

## 4 结论

本文提出一种新的马尔可夫网络构建算法,将回归误差统计检验与自回归更新策略相结合,并在迭代过程中引入 AIC 约束方法来提高算法的效率。实验表明,新算法在精度和效率方面具有更优的性能。本文认为该优势主要基于以下两个原因:①迭代过程引入基于 boosting 方法的更新策

略,改善了学习精度,减小了错误率;②通过单变量回归及AIC约束有效提高求解速度,减少迭代步数。本文算法的计算复杂度是  $O((2n+p/2)(p-1))$ ,而同类算法往往是  $O(p^2)$ 甚至更高。因此,INR算法在构建高维稀疏马尔可夫网络时,有效降低了计算复杂度,明显改善了时间性能。

## 参考文献:

- [1] Jansen R, Yu H, Greenbaum D, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data[J]. *Science*, 2003, 302(5644):449-453.
- [2] Dobra A. Variable selection and dependency networks for genome-wide data[J]. *Biostatistics*, 2009, 10(4):621-639.
- [3] 邓海松, 马义中, 邵文泽. 基于多层稀疏先验的回归样条变量选择算法[J]. 系统工程与电子技术, 2010, 32(12):2725-2729. (Deng H S, Ma Y Z, Shao W Z. Variable selection for regression splines using hierarchical sparseness prior[J]. *Systems Engineering and Electronics*, 2010, 32(12):2725-2729.)
- [4] Tibshirani R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society*, 1996, 58(1):267-288.
- [5] Hoerl A E, Kennard R W. Ridge regression; biased estimation for nonorthogonal problems[J]. *Technometrics*, 1970, 12(1):55-67.
- [6] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso[J]. *Biostatistics*, 2008, 9(3):432-441.
- [7] Anjum S, Doucet A, Holmes C C. A boosting approach to structure learning of graphs with and without prior knowledge[J]. *Bioinformatics*, 2009, 25(22):2929-2936.
- [8] Bleakley K, Biau G, Vert J-P. Supervised reconstruction of biological networks with local models[J]. *Bioinformatics*, 2007, 23(13):i57-i65.
- [9] Yamanishi Y, Vert J-P, Kanehisa M. Supervised enzyme network inference from the integration of genomic data and chemical information[J]. *Bioinformatics*, 2005, 21(S1):i468-i477.
- [10] 朱明敏, 刘蔚, 杨有龙. 基于全条件独立的贝叶斯网络MPD-JT构造算法[J]. 系统工程与电子技术, 2010, 32(6):1325-1328. (Zhu M M, Liu W, Yang Y L. Construction algorithm of MPD-JT for Bayesian networks based on full conditional independence [J]. *Systems Engineering and Electronics*, 2010, 32(6):1325-1328.)
- [11] 高晓光, 赵欢欢, 任佳. 基于蚁群优化的贝叶斯网络学习[J]. 系统工程与电子技术, 2010, 32(7):1509-1512. (Gao X G, Zhao H H, Ren J. Bayesian network learning on algorithm based on ant colony optimization[J]. *Systems Engineering and Electronics*, 2010, 32(7):1509-1512.)
- [12] Dempster A P. Covariance selection[J]. *Biometrics*, 1972, 28(1):157-175.
- [13] Steffen L L. *Graphical models*[M]. Oxford: Oxford University Press, 1996.
- [14] Edwards D. *Introduction to graphical modelling*[M]. 2nd ed. New York: Springer, 1995.
- [15] Whittaker J. *Graphical models in applied multivariate statistics*[M]. New York: Wiley, 1990.
- [16] Soranzo N, Bianconi G, Altafini C. Comparing association network algorithms for reverse engineering of large-scale gene regulatory networks; synthetic versus real data[J]. *Bioinformatics*, 2007, 23(13):1640-1647.
- [17] Kramer N, Schafer J, Boulesteix A-L. Regularized estimation of large-scale gene association networks using graphical Gaussian models[J]. *BMC Bioinformatics*, 2009, 10(1):384-407.
- [18] Akaike H. A new look at the statistical model identification[J]. *IEEE Trans. on Automatic Control*, 1974, 19(6):716-723.
- [19] Burnham K P, Anderson D R. *Model selection and multimodel inference: a practical information-theoretic approach* [M]. 2nd ed. New York: Springer Verlag, 2002.
- [20] McQuarrie A D R, Tsai C L. *Regression and time series model selection*[M]. Singapore: World Scientific Publishing Corporation, 1998.
- [21] Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model[J]. *Biometrika*, 2007, 94(1):19-35.
- [22] Sachs K, Perez O, Péér D, et al. Causal protein-signaling networks derived from multiparameter single-cell data[J]. *Science*, 2005, 308(5721):523-529.
- [23] Werhli A V, Grzegorzcyk M, Husmeier D. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks[J]. *Bioinformatics*, 2006, 22(20):2523-2531.
- [24] Gama-Castro S, Jiménez-Jacinto V, Peralta-Gil M, et al. Regulondb (version 6.0): gene regulation model of escherichia coli k-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation[J]. *Nucleic Acids Research*, 2008, 36(S1):D120-D124.