

Bayes 网络学习的 MCMC 方法

岳 博, 焦李成

(西安电子科技大学 雷达信号处理国家重点实验室, 陕西 西安 710071)

摘要: 基于 Bayes 统计理论, 提出了一种从数据样本中学习 Bayes 网络的 Markov 链 Monte Carlo (MCMC) 方法. 首先通过先验概率和数据样本的结合得到未归一化的后验概率, 然后使用此概率指导随机搜索算法寻找“好”的网络结构模型. 通过对 Alarm 网络的学习表明了本算法具有较好的性能.

关键词: Bayes 网络; Markov 链 Monte Carlo 方法; 模型选择; 随机搜索

中图分类号: TP301 **文献标识码:** A

MCMC approach to Bayesian networks learning

YUE Bo, JIAO Li-cheng

(Key Lab for Radar Signal Processing, Xidian University, Shanxi Xi'an 710071, China)

Abstract: A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. In many cases, the authors hoped to learn Bayesian networks from data. Using the Markov chain Monte Carlo (MCMC) approach, this paper proposed a Bayesian statistical method for learning Bayesian networks from data, in terms of network structures and parameters. Prior specification and stochastic search were two important components of this approach. The combination of prior probability and data samples induced a posterior distribution that would guide the stochastic search towards the network structures having the maximal posterior probability. The performance of this approach is illustrated by the learning of the Alarm network from data.

Key words: Bayesian networks; Markov chain Monte Carlo; model selection; stochastic search

1 引言 (Introduction)

作为一种表示概率空间的模型, Bayes 网络利用一组随机变量之间的条件独立关系减少了对这组变量的联合分布进行表示时所需要的参数的个数, 同时也借助于反映这些条件独立关系的网络结构得到了有效的概率推理算法, 这些都使得 Bayes 网络模型在众多的领域中得到了越来越广泛的应用, 例如医疗诊断和故障诊断系统、信息检索和时间序列分析等.

一般地, 人们希望通过数据样本来学习 Bayes 网络模型, 其中对网络结构的学习是主要的, 这是因为网络结构模型空间的规模是随着网络结点个数的增加而呈超指数增长的, 对此进行搜索需要有效的算法来完成, 而在已知网络结构以后, 如何得到网络参数只是一个相对简单的参数估计问题. 已有的 Bayes 网络学习算法主要有 K_2 算法^[1] 和 B-搜索 (B-search) 算法^[2] 等. 这两种算法都首先定义一种评价网络结构模型优劣的准则, 例如后验概率、条件熵

等, 然后使用贪婪算法搜索好的网络结构模型.

本文使用 Bayes 统计和 Markov 链 Monte Carlo 方法来搜索后验概率最大的网络结构模型. 将网络结构的先验分布和数据样本相结合, 可以得到尚未归一化的网络结构模型的后验概率, 由此指导随机搜索算法寻找“好”的网络结构模型. 与贪婪算法相比较, 本文中的方法能够对更大的网络结构模型空间进行搜索^[3]. 同时, 在初始条件相同的情况下, 贪婪算法的多次运行都将得到相同的网络结构序列和“最优”的网络, 而多次执行本文中的随机搜索算法将搜索到更大部分的模型空间, 因此将得到更好的结果. 通过对 Alarm 网络的学习验证了本算法具有较好的性能.

2 Bayes 网络模型及学习 (Learning Bayesian networks)

在本文中, 只考虑以一组离散型随机变量 $U = \{X_1, X_2, \dots, X_N\}$, $N \geq 1$ 作为结点的 Bayes 网络^[4,5] 的学习问题, 对相关内容的讨论, 也以离散变量的情况为例.

2.1 Bayes 网络模型(Bayesian networks)

一个 Bayes 网络是一个二元组 $B = \langle G, \Theta \rangle$, 其中

1) $G = \{\Pi_1, \Pi_2, \dots, \Pi_N\}$ 是一个有向无环图(DAG), 其结点为

$$U = \{X_1, X_2, \dots, X_N\}, N \geq 1,$$

$\Pi_i \subset 2^U, i = 1, \dots, N$ 是结点 X_i 的父结点集合;

2) $\Theta = \{P(X_i | \Pi_i) | X_i \in U, i = 1, \dots, N\}$ 是一组条件概率的集合, 称为网络参数.

一个 Bayes 网络 $B = \langle G, \Theta \rangle$ 确定了一个概率空间 $\langle \Omega, F, P \rangle$, 其中

$$P(U) = \prod_{i=1}^N P(X_i | \Pi_i). \quad (1)$$

根据 Bayes 网络的定义, 从数据样本中学习 Bayes 网络, 就是要学习网络模型的结构 G 和参数 Θ , 而学习网络结构就是寻找每一个节点的父结点.

下面给出一个关于 Bayes 网络结构的命题.

命题 1 如果 G 是一个有向无环图(DAG), V_1, V_2 是 G 中任意两个结点, 那么总能够在 V_1 和 V_2 之间添加一条弧, 使得所得的图仍然是一个 DAG.

证 需要说明的是在 V_1 和 V_2 之间添加一条任意方向的弧之后都构成有向环的情况是不可能发生的. 假设在 G 中添加由 V_1 指向 V_2 的弧 $V_1 \rightarrow V_2$ 后形成有向环, 作者把此环路记作 V_1, V_2, R, V_1 , 其中 R 表示构成此环路的所有其它结点; 又如果在添加由 V_2 指向 V_1 的弧 $V_2 \rightarrow V_1$ 之后也形成有向环 V_2, V_1, T, V_2 , 那么在原来的图 G 中就会存在一个有向环 V_1, T, V_2, R, V_1 , 这就与 G 是一个 DAG 相矛盾, 由此命题得证.

2.2 Bayes 网络学习的 Bayes 方法(Bayesian method for learning Bayesian networks)

由 Bayes 公式, 在已知数据样本 $D = \{D_1, D_2, \dots, D_M\}$, 其中 $D_i = \{X_1^i, X_2^i, \dots, X_N^i\}, i = 1, \dots, M$ 时, 网络结构 S 的后验概率 $P(S | D)$ 为

$$P(S | D) = \frac{P(D | S)P(S)}{P(D)} \propto P(D | S)P(S). \quad (2)$$

上式中 $P(S)$ 是先验概率, $P(D | S)$ 是数据样本 D 在网络结构 S 下的似然函数. 在本文的计算中, 使用 $P(D | \hat{\Theta}, S)$ 来近似 $P(D | S)$, 这里的 $\hat{\Theta}$ 是网络参数 Θ 在数据 D 下的极大似然估计(MLE), 并且由式(1), 有

$$P(D | S) = \prod_{m=1}^M \prod_{i=1}^N P(X_i^m | \Pi_i^m, \hat{\Theta}_i, S). \quad (3)$$

由于 Bayes 网络结构模型空间的规模 $|S|$ 是随着网络结点数的增加呈超指数增长的, 在式(2)中

无法得到计算 $P(S | D)$ 所需的归一化常数, 因此不能得到真正的 $P(S | D)$, 在这种情况下可以使用 Markov 链 Monte Carlo 方法(MCMC)按照未归一化的概率分布 $P(D | S)P(S)$ 对网络结构 S 进行抽样.

3 Markov 链 Monte Carlo 方法(Markov chain Monte Carlo method)

虽然 $P(S | D)$ 是无法直接计算得到的, 但是可以首先构造一个 Markov 链, 使得其极限分布收敛于网络结构的后验分布 $P(S | D)$, 然后使用 Monte Carlo 方法对此 Markov 链进行抽样, 那么所得到的网络结构的样本序列

$$S^0, S^1, S^2, \dots \quad (4)$$

的概率分布收敛于 $P(S | D)$. 即便如此, 因为网络结构的模型空间是非常大的, 利用抽样计数的方法得到 $P(S | D)$ 也是不可能的, 但是可以在此序列中挑出具有最大后验概率的网络结构.

本文中使用的 MCMC 方法是 Metropolis-Hastings 算法, 作者用它来对网络结构进行抽样和搜索. Metropolis-Hastings 算法的基本思想是, 首先给定一个网络结构的初值 S^0 , 然后循环执行下述两步:

- 1) 按照产生概率 $G(S^i, S^*)$ 生成一个新的网络结构 S^* ;
- 2) 以接受概率

$$A(S^i, S^*) = \min \left\{ 1, \frac{G(S^*, S^i)P(D | S^*)P(S^*)}{G(S^i, S^*)P(D | S^i)P(S^i)} \right\} \quad (5)$$

使得 $S^{i+1} = S^*$, 否则 $S^{i+1} = S^i$.

这样得到的抽样序列的极限分布就是网络结构的后验分布 $P(S | D)$.

为了使用 Metropolis-Hastings 算法对网络结构进行抽样, 必须给出一个恰当的网络结构的产生概率 $G(S^i, S^*)$, 本文通过等概率随机选择下述3种操作中的一个, 由网络结构 S^i 生成一个新的网络结构 S^* 来定义产生概率 $G(S^i, S^*)$:

- 1) 删除一条弧: 随机选择一条弧进行删除;
- 2) 弧的反向: 随机选择一条弧, 在满足 DAG 的要求下, 将它的弧头结点和弧尾结点进行交换;
- 3) 添加一条弧: 随机选择两个相互之间不存在弧的结点, 由命题 1, 我们总能够在这两个结点之间添加一条弧.

如果由 S^i 生成的新的网络结构 S^* 具有较大的后验概率, 由式(2), 即

$$P(D | S^*)P(S^*) \geq P(D | S^i)P(S^i),$$

那么根据式(5), 接受概率 $A(S^i, S^*) = 1$, 因为

$G(S^i, S^*) = G(S^*, S^i)$. 这样抽样序列中的下一个样本点一定是具有较大后验概率的 S^* , 也就是说, 式(4)所示的抽样序列有着向网络结构模型空间中具有较大后验概率的模型靠近的趋势, 因此可以将 MCMC 方法作为一种搜索算法来使用.

4 网络结构先验概率 $P(S)$ 和接受概率 $A(S^i, S^*)$ 的计算 (Computation of prior $P(S)$ and acceptance probability $A(S^i, S^*)$)

在 Metropolis-Hastings 算法中, 需要给出网络结构的先验概率 $P(S)$ 和计算两个网络结构模型下样本似然的比值 $P(D | S^*) / P(D | S^i)$.

4.1 网络结构的先验概率 $P(S)$ (The prior probability $P(S)$ for the network structure)

利用不同的先验概率分布, 能够灵活地控制网络结构模型的某些特性, 例如网络中弧的个数、网络中自由参数的个数等. 事实上, 在以某种准则作为模型评价标准的模型选择方法中, 本文的先验概率相当于那里的惩罚项. 本文使用 Bayes 信息准则 (BIC) 中的惩罚函数作为网络结构的先验概率

$$\log P(S) = -\frac{1}{2} \log M \cdot \sum_{i=1}^n q_i (r_i - 1) + c, \tag{6}$$

其中 c 是归一化常数, $\sum_{i=1}^n q_i (r_i - 1)$ 是网络中自由参数的个数, r_i 是节点 X_i 的状态数, q_i 是 Π_i 的状态数. 显然, 网络结构越复杂, 式(6)所赋予的先验概率越小. 虽然式(6)给出的先验概率不是归一化的, 但对于式(5)已经足够了. 同时式(6)的另一个特性是它是可分解的, 这使得接受概率的计算非常方便和简单.

4.2 接受概率 $A(S^i, S^*)$ 的计算 (Computation acceptance probability $A(S^i, S^*)$)

在给出产生概率 $G(S^i, S^*)$ 时所定义的 3 种操作——弧的删除、反向和添加——对于网络结构而言都是一种局部性的操作, 它最多只改变两个结点的父结点集合, 这大大简化了接受概率 $A(S^i, S^*)$ 的计算.

假设在进行弧的删除、反向或添加时所对应的两个结点是 X_j 和 X_k , 同时在网络结构为 S^i 时它们的父结点集合分别为 Π_j 和 Π_k , 在 S^* 下的父结点集合为 Π_j^* 和 Π_k^* , 由式(3), 样本似然的比值 $P(D | S^*) / P(D | S^i)$ 为

$$\frac{P(D | S^*)}{P(D | S^i)} =$$

$$\begin{aligned} & \frac{\prod_{m=1}^M \prod_{l=1}^N P(X_l^m | \Pi_l^{*m}, \hat{\Theta}_l^*, S^*)}{\prod_{m=1}^M \prod_{l=1}^N P(X_l^m | \Pi_l^m, \hat{\Theta}_l^i, S^i)} = \\ & \frac{\prod_{m=1}^M \prod_{l=1}^N \frac{P(X_l^m | \Pi_l^{*m}, \hat{\Theta}_l^*, S^*)}{P(X_l^m | \Pi_l^m, \hat{\Theta}_l^i, S^i)}}{\prod_{m=1}^M \frac{P(X_j^m | \Pi_j^m, \hat{\Theta}_j^i, S^i) P(X_k^m | \Pi_k^m, \hat{\Theta}_k^i, S^i)}{P(X_j^m | \Pi_j^{*m}, \hat{\Theta}_j^*, S^*) P(X_k^m | \Pi_k^{*m}, \hat{\Theta}_k^*, S^*)}} \end{aligned} \tag{7}$$

上式表明, 在计算 $P(D | S^*) / P(D | S^i)$ 时, 只需要两个结点 X_j, X_k 在两种网络结构 S^i, S^* 下的 4 组网络参数的 MLE 估计 $\hat{\Theta}_j^*, \hat{\Theta}_k^*, \hat{\Theta}_j^i$ 和 $\hat{\Theta}_k^i$ 就足够了, 与其它节点是无关的.

5 算法性能与实例 (Function and example of algorithm)

以 Alarm 网络^[6,7]为例来说明本文所提出的算法的性能. 首先使用此网络抽样出 10,000 个数据样本, 然后使用本文的 MCMC 方法在不同样本容量的数据下抽样得到一个网络结构的序列, 在本例中抽出 5000 个网络结构样本. 其中在网络结构的生成中, 弧的添加、删除、反向操作以及保持原网络结构不变这 4 种情况是等可能发生的, 因此在各种情况下, 产生概率 $G(S^i, S^*)$ 都为 1/4. 数据样本的容量由 1000 变化到 10000, 每增加 1000 个数据进行一次计算. 接下来在这 5000 个网络结构样本中找出具有最大后验概率的网络结构, 以此计算数据样本的似然函数值, 结果如图 1 所示. 图中的横轴为样本容量, 纵轴为数据样本的似然值. 图 1 的另一条曲线是在相同的数据样本上使用 B-搜索算法所得的结果. 相对于模型空间的维数而言, MCMC 方法在进行了非常少的 5000 次抽样中, 就得到了与 B-搜索算法相当的结果, 而后者仅仅在搜索网络的第一条弧时就需要对 $37 \times 36 = 1332$ 种可能的情况进行计算.

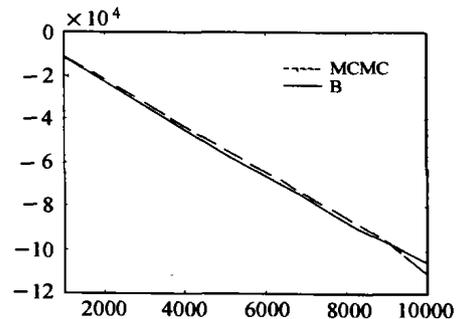


图 1 样本似然

Fig. 1 The likelihood of the data

(下转第 588 页)

象,所以常规 Smith 补偿的方法受到很大的限制.本文采用 Elman 网络补偿常规模型的建模误差,减少了精确建模的要求.

1) 如果纯滞后环节建模精确,即 $\tau = \tau_p$, 只要 Elman 网络通过训练满足精度要求就可以认为完全抵消了纯滞后环节对控制品质及系统稳定性的不利影响.

2) 如果纯滞后环节建模不精确,即 $\tau \neq \tau_p$, 则纯滞后环节不能被完全抵消,但其不利影响可以大幅削弱,削弱的程度取决于其模型精确度.

通过仿真及实验结果可以看出,基于 Elman 网络补偿的 Smith 预测控制利用了神经网络的非线性逼近特性及神经元自学习、自组织的能力,弥补了常规 Smith 预测控制在解决模型不确定系统中的不足之处,使得其广泛应用成为可能.

参考文献(References):

- [1] 田杰,龚至豪,任雪梅.纯滞后系统的单神经元 Smith 预测控制[J].计算机仿真,2001,18(1):40-42
(TIAN Jie, GONG Zhihao, REN Xuemei. Single neuron Smith predictive control of time delay systems [J]. *Computer Simulation*, 2001, 18(1):40-42.)
- [2] 王建晖,齐昕,顾树生.一类纯滞后系统模糊 Smith 控制策略的

研究[J].控制与决策,1998,13(2):141-145.

(WANG Jianhui, QI Xin, GU Shusheng. Research on fuzzy Smith control strategy of time delay systems [J]. *Control and Decision*, 1998, 13(2):141-145.)

- [3] 孙增圻,张再兴,邓志东.智能控制理论与技术[M].北京:清华大学出版社,1997.
(SUN Zengqi, ZHANG Zaixing, DENG Zhidong. *Theory and Technology of Intelligent Control* [M]. Beijing: Tsinghua University Press, 1997.)
- [4] 田杰,陈杰,张宇河.基于 Elman 网络补偿模型的 Smith 预测控制[J].北京理工大学学报,2002,22(3):343-346.
(TIAN Jie, CHEN Jie, ZHANG Yuhe. Smith predictive control based on Elman network compensatory model [J]. *J of Beijing Institute of Technology*, 2002, 22(3):343-346.)
- [5] 陈晓东,马广富,王子才.改进的 Elman 网络与机理模型的互补建模方法[J].系统仿真学报,1999,11(2):97-100.
(CHEN Xiaodong, MA Guangfu, WANG Zicai. A mutually compensatory modeling method based on mechanism model and improved Elman network [J]. *J of System Simulation*, 1999, 11(2):97-100.)

作者简介:

田杰 (1972—),男,工学博士.主要研究领域为信息融合,智能控制等.E-mail:jjeti@263.net;

陈杰 (1965—),男,工学博士,教授,博士生导师.主要研究领域和方向为:复杂系统的多指标优化与决策,智能控制,非线性控制,生物信息处理等理论与方法研究.Email:chenjie@bit.edu.cn.

(上接第 584 页)

6 结论(Conclusion)

基于 Bayes 统计理论,本文提出了一种从数据样本中学习 Bayes 网络的 Markov 链 Monte Carlo 方法.首先使用先验概率和数据样本相结合得到未归一化的后验概率,然后以此指导随机搜索算法寻找“好”的网络结构模型.对 Alarm 网络的学习表明本方法具有较好的性能.

参考文献(References):

- [1] COOPER G, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data [J]. *Machine Learning*, 1992, 9(3):309-374.
- [2] BUNTINE W. Theory refinement on Bayesian networks [A]. *Proc of 7th Conf Uncertainty Artificial Intelligence* [C]. Los Angeles, CA, 1991:652-660.
- [3] CHIPMAN H, GEORGE E, McCULLOCH R. Bayesian CART model search [J]. *J of the American Statistical Association*, 1998,

93(4):935-948.

- [4] HECKERMAN D, GEIGER D, CHICKERING D. Learning Bayesian networks: The combination of knowledge and statistical data [J]. *Machine Learning*, 1995, 20(2):197-243.
- [5] KHALFALLAH F, MELLOULI K. Optimized algorithm for learning Bayesian network from data [A]. *Proc of the European Conf on Symbolic and Quantitative Approach to Reasoning and Uncertainty* [C]. London: Cambridge Press, 1999.
- [6] LIU J, CHANG K, ZHOU J. Learning Bayesian networks with a hybrid convergent method [J]. *IEEE Trans Systems, Man, Cybernetics*, 1999, 29(2):436-449.
- [7] [http://www.cs.huji.ac.il/labs/complbio/Repository/Datasets/ alarm/alarm.htm](http://www.cs.huji.ac.il/labs/complbio/Repository/Datasets/alarm/alarm.htm).

作者简介:

岳博 (1970—),男,博士研究生,主要研究方向为概率模型, Bayes 统计方法等.E-mail: yuebo@rsp.xidian.edu.cn;

焦李成 (1959—),男,教授,博士生导师,主要研究领域为非线性科学,智能信息处理等.