不完备证据条件下的 Bayesian 网络参数学习*)

刘 震 周明天
(电子科技大学计算机科学与工程学院 成都 610054)

摘 要 在 Bayesian 网络推理中,对节点做参数学习是必不可少的。但在学习过程中,常常会出现证据丢失,导致参数收敛速度减慢,同时影响参数学习的精确度,甚至给参数收敛带来困难。针对这样的问题,本文提出一种证据丢失参数模型,并推导出包含学习率的 EM 更新算法。收敛性能的理论分析和仿真试验结果两方面均表明,新算法与传统处理算法相比,在不降低参数估计精度的前提下,具有更快的收敛速度,为保证不完备证据条件下可信高效的 Bayesian 网络参数学习提供了一条可行的解决途径。

关键词 Bayesian 网络,证据丢失,EM(η)算法,学习率

Parameter Learning in Bayesian Network under Incomplete Evidence Input

LIU Zhen ZHOU Ming-Tian
(College of Computer Science and Engineering, UESTC, Chengdu 610054)

Abstract To Infer in a Bayesian network, parameter learning for a given network node is obviously necessary. But during the course of parameter learning, evidence loss would happen from time to time and therefore slow down the parameter convergence, influence the accuracy of parameter learning, and even cause no parameter convergence. Aiming at this question, this paper proposes a parameter model under evidence loss and deduce an EM updating algorithm which contains learning rate. Compared with the traditional algorithms, both of the converging performance analysis and simulation testing results show that new algorithm has much quicker convergence rate without degrading the accuracy of parameter estimation. New algorithm provides a feasible way to ensure a trusted and efficient Bayesian network parameter learning under the situation of evidence loss.

Keywords Bayesian network, Evidence loss, EM(η) algorithm, Learning rate

1 引言

Bayesian 网络也称为信念网络,最早由 Judea Pearl 于 1988 年提出。Bayesian 网络本质上是一种基于概率的不确定性推理网络[1]。Lauritzen 和 Spiegelhalter 基于影响图理论(Theory of Influence Diagram)对 Bayesian 网络加以了深化与发展[2]。经过十多年的发展,Bayesian 网络理论在不确定推理学习领域已确立了不可动摇的地位,并得到了众多研究者的接受和认同[3]。Bayesian 网络是无环的 DAG 图,节点之间存在着因果的逻辑联系[4]。通过学习与推理,Bayesian 网络可以利用一些先验的知识对一些现象进行识别、分类和预测。作为一种基于概率的不确定性推理方法,贝叶斯网络在处理不确定信息的智能化系统中得到了重要的应用,已成功地应用于医疗诊断、统计决策、专家系统等领域[5~7]。

对于无证据丢失的 Bayesian 网络,Bayesian 参数估计是比较通用的方法。然而,在实际的节点参数学习过程中,常常发生证据丢失,似然分布会发生偏移、畸变,导致 Bayesian 参数估计出现严重误差[8]。对于这类具有复杂边界的极大似然问题,基于混合高斯模型的 EM(Expectation Maximized)算法是一种常采用的计算工具[9~11]。然而,由于这种 EM 算法收敛速度慢,计算效率低,很难满足实际决策支持系统的性能要求。针对传统 EM 算法所存在的计算效率问题,有些研究者将 BPCA 算法、基于最小二乘方的回归分析(LSR)等方法[12,13]也引入到该类问题的研究中。采用这类算法虽然可以显著提高参数估计收敛的速度,但由于潜在的高偏依(Bias)缺陷,存在较严重的估计误差问题;对于某些实际应用,例如基因的微序列分析问题,难以保证可信的 Bayesian 网络参数估计精度[14]。所以,解决这类问题的关键是如何平衡估计误差和收敛速度之间的矛盾,使算法模型具有更高的实用性。

本文基于标准似然函数构建证据丢失的计算模型,利用 X^2 距离近似估计证据丢失导致的误差距离;推导出包含学习率的 EM 算法。实验结果表明,新算法与传统处理算法相比,在不降低估计精度的前提下具有更快的收敛速度,能够较好地保证不完备证据条件下可信高效的 Bayesian 网络参数估计。

2 新网络计算模型

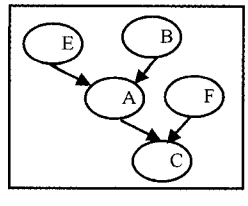
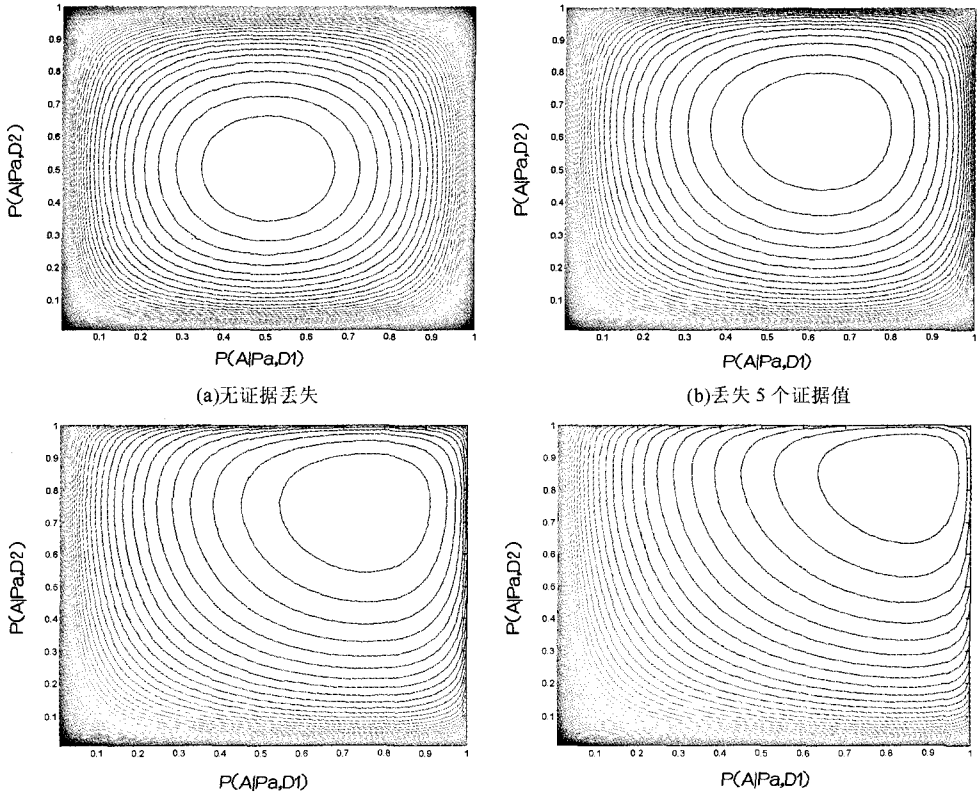


图 1 五节点 Bayesian 网络

*)电子科学基金(No. 51415010101DZ02)。刘 震 博士研究生,主要研究方向为智能安全、不确定推理;周明天 教授,博士生导师,主要研究方向为网络计算、网络安全技术。

在输入证据发生丢失的情况下，经典 Bayesian 参数估计理论通常会失效。现以图 1 这样的一个简单 Bayesian 网络为例加以说明。在图 2 中，显示了当节点 A 发生不同数量的输入证据丢失时所得到的满足多项式分布的对数似然分布图。从图中可以看出，随着证据丢失数的增加，节点 A 的对数似然分布在发生明显的偏畸。如果仍然采用经典 Bayesian 参数估计方法，必然会导致参数估计的严重误差。所以，在这种情况下，已不能再使用完备证据条件下的网络参数估计方法。



(a)无证据丢失　(b)丢失 5 个证据值

(c)丢失 12 个证据值　(d)丢失 20 个证据值

图 2　不同证据丢失数条件下的节点对数似然分布图

2.1　基本符号定义

Bayesian 网络是通过条件概率表(CPT)实现对节点的参数化。一般地，给定一个网络结构 S，如果令 X_i 表示网络的某一个节点，令 pa_i 表示 X_i 的一组父节点。那么令 $x_i^k(k=1,\cdots,r_i)$ 代表 X_i 的 r_i 个可能取值，同时令 $pa_i^j(j=1,\cdots,q_i)$ 代表 pa_i 的 q_i 个可能取值，本文定义参数 Θ_{ijk} 代表 CPT 表中的节点参数概率 $p(X_i=x_i^k \mid pq_i=pq_i^j)$，定义 Θ 代表整个节点参数向量。

2.2　参数模型建模

给定证据样本集 $D=\{y_1,\cdots,y_n\}$，假设存在一个有证据丢失参数向量 $\tilde{\Theta}$ 和一个标准无证据丢失的模型参数向量 $\bar{\Theta}$，本文考虑构建 $\tilde{\Theta}$ 与 $\bar{\Theta}$ 之间的估计误差距离在 D 样本集上的证据丢失参数模型 $F(\tilde{\Theta})$。为了平衡误差距离，其表达式是完备证据条件下 Bayesian 参数估计式的修正形式：

$$F(\tilde{\Theta})=\eta L_D(\tilde{\Theta})-d(\tilde{\Theta},\bar{\Theta}) \tag{4}$$

这里 $L_D(\tilde{\Theta})$ 是模型参数向量 $\tilde{\Theta}$ 在 D 上的归一化对数似然函数(选择对数似然的目的是在不改变标准似然函数的单调性的条件下，方便微分运算)，其中 $L_D(\tilde{\Theta})=\frac{1}{N}\sum_{l=1}^{N}\log p_{\tilde{\Theta}}(y_l)$。式(4)中的项 $d(\tilde{\Theta},\bar{\Theta})$ 是丢失参数模型与标准参数模型之间由于证据丢失导致的误差距离估计。类似于最小二乘回归分析的罚函数思想，令该距离保持 $\tilde{\Theta}$ 足够接近 $\bar{\Theta}$。参数 η 是引入的权重学习率，总大于 0。

通过寻找式(4)的极值点，可以最小化误差距离，使证据丢失的参数估计值最大限度趋近甚至等于标准参数值，从而达到在证据丢失条件下估计可信参数的目的。由于存在证据丢失，直接解析地求解式(4)的极值点在计算上是有困难的，因此只能选择一种近似算法。首先考虑采用 Taylor 一阶近似作为对数似然的近似表达式：

$$L_D(\tilde{\Theta})\approx L_D(\bar{\Theta})+\nabla L_D(\bar{\Theta})\cdot(\tilde{\Theta}-\bar{\Theta}) \tag{5}$$

代入替换(4)式中的项 $L_D(\tilde{\Theta})$，则原式修改为以下等式：

$$\hat{F}(\tilde{\Theta})=\eta[L_D(\bar{\Theta})+\nabla L_D(\bar{\Theta})\cdot(\tilde{\Theta}-\bar{\Theta})]-d(\tilde{\Theta},\bar{\Theta}) \tag{6}$$

做 $L_D(\bar{\Theta})$ 的梯度分解，可以得到一个梯度向量表达式：

$$\nabla_{ijk}L_D(\bar{\Theta})=\frac{1}{\Theta_{ijk}}\frac{\sum_{l=1}^{N}p\Theta(x_i^k,pa_i^j\mid y_l)}{N}=\frac{\xi_{\Theta}(x_i^k,pa_i^j\mid D)}{\Theta_{ijk}} \tag{7}$$

其中 $\xi_{\Theta}(x_i^k,pa_i^j\mid D)$ 代表 $\bar{\Theta}$ 上的基于样本的平均。因为对每个 i,j，必须保证条件 $\sum_k\Theta_{ijk}=1$，所以(6)式是一个有限制条件的极值问题。根据这个限制条件，由 *Lagrange* 乘子法，可以总结得出：对于每个 $i,j,\tilde{\Theta}$，必须满足 $\frac{\partial}{\partial\tilde{\Theta}_{ijk}}(\hat{F}(\tilde{\Theta})+\sum_{i'j'}r_{ij'}(\sum\Theta_{i'j'k'}-1))=0$，即是说，

$$\eta\nabla_{ijk}L_D(\tilde{\Theta})-\frac{\partial}{\partial\tilde{\Theta}_{ijk}}d(\tilde{\Theta},\bar{\Theta})+\gamma_{ij}=0 \tag{8}$$

其中 γ_{ij} 是待定的 Lagrange 系数。下面分析距离 d 的表达式。

2.3　距离

两个分布 p 和 q 之间的 χ^2 距离可以定义为

$$\chi^2(p\|q)=\frac{1}{2}\sum_x(p(x)-q(x))^2/q(x) \tag{9}$$

利用 χ^2 距离近似每一个分布 $p_{\bar{\Theta}}(\chi_i^k|P_{a_i}=pa_i^j)$。即是说，使用以下距离函数

$$\widehat{\chi^2}(\tilde{\Theta}\|\bar{\Theta})=\sum_i\sum_j p_{\bar{\Theta}}(Pa_i=pa_i^j)\chi^2(\tilde{\Theta}_{ij}\|\bar{\Theta}_{ij}) \tag{10}$$

令 $\widehat{p}(pa_i^j)$ 作为先验分布近似替换 $p_{\bar{\Theta}}(pa_i^j)$，与前面方法类似，可以推导求得 $\frac{\partial}{\partial\tilde{\Theta}_{ijk}}\chi^2(\tilde{\Theta}_{ij}\|\bar{\Theta}_{ij})=\widehat{p}(pa_i^j)(\tilde{\Theta}_{ijk}/\bar{\Theta}_{ijk}-1)$。代入式(8)，就可以得到 $\eta\nabla_{ijk}L_D(\bar{\Theta})-\hat{p}(pa_i^j)(\tilde{\Theta}_{ijk}/\bar{\Theta}_{ijk}-1)+\gamma_{ij}=0$。整理得

$$\tilde{\Theta}_{ijk}=\frac{\eta}{\widehat{p}(pa_i^j)}\nabla_{ijk}L_D(\bar{\Theta})\bar{\Theta}_{ijk}+\frac{\gamma_{ij}}{\widehat{p}(pa_i^j)}\bar{\Theta}_{ijk}+\bar{\Theta}_{ijk} \tag{11}$$

根据式(7)替换 $\nabla_{ijk}L_D(\bar{\Theta})$ 为它的表达式，累加 k 求和，可以得到

$$1=\frac{\eta}{\widehat{p}(pa_i^j)}\sum_k\xi_{\bar{\Theta}}(x_i^k,pa_i^j|D)+\frac{\gamma_{ij}}{\widehat{p}(pa_i^j)}+1 \tag{12}$$

整理得到系数 $\gamma_{ij}=-\eta\xi_{\bar{\Theta}}(pa_i^j|D)$。再代回到式(11)，容易得到

$$\tilde{\Theta}_{ijk}=\frac{\eta}{\widehat{p}(pa_i^j)}\xi_{\bar{\Theta}}(x_i^k,pa_i^j|D)-\frac{\eta}{\widehat{p}(pa)}\xi_{\bar{\Theta}}(pa_i^j|D)\bar{\Theta}_{ijk}+\bar{\Theta}_{ijk} \tag{13}$$

2.4 更新规则

下面需要引入 $\hat{p}(pa_i^j)$ 的估计作为 $p_{\bar{\Theta}}(Pa_i=pa_i^j)$ 的近似值。在统计计算中，使用基于已知样本的期望作为近似是一种合理的解决途径，所以可以得到

$$\widehat{p}(pa_i^j)=\frac{1}{N}\sum_{l=1}^{N}p_{\bar{\Theta}}(Pa_i=pa_i^j|y_l)=\xi_{\bar{\Theta}}(pa_i^j|D) \tag{14}$$

下面开始考虑证据更新规则。将(14)式代入到(13)式，可以得到

$$\begin{aligned}\tilde{\Theta}_{ijk}&=\frac{\eta\xi_{\bar{\Theta}}(x_i^k,pa_i^j|D)}{\xi_{\bar{\Theta}}(pa_i^j|D)}-\frac{\eta\xi_{\bar{\Theta}}(pa_i^j|D)}{\xi_{\bar{\Theta}}(pa_i^j|D)}\cdot\bar{\Theta}_{ijk}+\bar{\Theta}_{ijk}\\&=\eta\frac{\xi_{\bar{\Theta}}(x_i^k,pa_i^j|D)}{\xi_{\bar{\Theta}}(pa_i^j|D)}+(1-\eta)\bar{\Theta}_{ijk}\end{aligned} \tag{15}$$

这个等式描述了参数计算的权重平均，其权重平均是通过 pa_i^j 和当前参数 $\bar{\Theta}_{ijk}$ 基于学习率的权重划分实现样本平均。当 $\eta=1$ 时，这个更新规则就退化为经典的 EM 算法。因此，我们把这种参数化的更新规则称为 $EM(\eta)$。下面研究 η 的取值。

1)当 $\eta<1$，在每一步 EM 更新和当前参数向量之间，EM(η)实例化新的参数值到一个权重组合。因此，新的参数一般会介于旧的参数和新引入的证据之间，其收敛速度会慢于经典 EM 算法。

2)当 $1<\eta<2$，和经典 EM 算法相比，EM(η)以更快的步长做加权平均。通过本文第 4 节的仿真结果验证表明，它的性能优于经典 EM 算法，收敛速度最好情况下可以加快 1 倍左右。

3)当 $\eta>2$，本文在第 3 节算法收敛性能分析中会证明其在某些条件不满足的情况下将无法保证算法收敛。

3 收敛性能分析

本节采用矩阵分析的方法研究 EM(η)算法的收敛性能。为了方便讨论，现将 EM(η)看作参数向量 $\bar{\Theta}$ 上的一个算子 Φ，并在其充分小的一个邻域内，采用其 Taylor 展式的线性部分作为这个算子的近似。先设

$$\Phi(\bar{\Theta})_{ijk}=\eta\frac{\sum_l p_{\bar{\Theta}}(x_i^k,Pa_i^j+y_l)}{\sum_{k',l}p_{\bar{\Theta}}(x_i^{k'},Pa_i^j|y_l)}(1+\eta)\bar{\Theta}_{ijk} \tag{16}$$

引理 1 如果 θ^* 是似然函数的一个局部极大值，那么对于任何 η 值，均有 $\Phi(\theta^*)=\theta^*$ 成立。

证明：由于满足条件 $\sum_k\theta_{ijk}=1$，可以得到以下等式对于任何局部极大均成立(通过引入 Lagrange 乘子法)：

$$\frac{\partial}{\partial\theta_{ijk}}[L_D(\theta)+\gamma(\sum_{k'}\theta_{ijk'}-1)]_{\theta=\theta^*}=0 \tag{17}$$

因此

$$\frac{1}{N\theta_{ijk}^*}=\sum_l^N P_{\theta^*}(x_i^k,Pa_i^j|y_l)+\gamma=0 \tag{18}$$

(18)式两边乘上 θ_{ijk}，再对所有的 k 累加求和，可得系数 $\gamma=-\frac{1}{N}\sum_{k',l}P_{\theta^*}(x_i^{k'},Pa_i^j|y_l)$。将 γ 的值再代回(18)式，化简后可得

$$\theta_{ijk}^*=\frac{\sum_l P_{\theta^*}(x_i^k,Pa_i^j|y_l)}{\sum_{k',l}P_{\theta^*}(x_i^{k'},Pa_i^j|y_l)} \tag{19}$$

这样就立即得到结论 $\Phi(\theta^*)=\theta^*$，引理得证。

令 $\tilde{\theta}$ 表示 $\Phi(\theta)$，有

$$\tilde{\theta}-\theta^*=\Phi(\theta)-\Phi(\theta^*)=\nabla\Phi(\theta^*)(\theta-\theta^*)+o(\theta-\theta^*) \tag{20}$$

$\nabla\Phi(\theta^*)$ 是一个 $m\times m$ 矩阵，m 是 θ 的维数。显然，在 θ^* 的某个邻域内，$\nabla\Phi(\theta^*)(\theta-\theta^*)$ 形成了 Φ 的一个线性近似。如果令 α_{ijk}^l 代表 $p_{\theta^*}(x_i^k,pa_i^j|y_l)$，当 $k=k'$ 时，有 $\delta_{kk'}=1$，否则 $\delta_{kk'}=0$。那么对(20)式导数部分有

$$\begin{aligned}\nabla\Phi_{kk'}=\frac{\partial\Phi(\theta)_{ijk}}{\partial\theta_{ijk'}}&=\delta_{kk'}(1-\eta)+\frac{\eta}{\theta_{ijk'}\sum_{k''l}\alpha_{ijk''}^l}(\delta_{kk'}\sum_l\alpha_{ijk}^l-\\&\sum_l\alpha_{ijk}^l\alpha_{ijk'}^l)\end{aligned} \tag{21}$$

由式(19)知，$\theta_{ijk}^*=\sum_l\alpha_{ijk}^l/\sum_{k''l}\alpha_{ijk''}^l$，那么

$$\nabla\Phi_{kk'}=\delta_{kk'}-\eta\frac{\sum_l\alpha_{ijk}^l\alpha_{ijk'}^l}{\theta_{ijk'}\sum_{k''l}\alpha_{ijk''}^l} \tag{22}$$

换成矩阵形式：

$$\nabla\Phi(\theta^*)=I-\eta M$$

其中 $M_{kk'}=\frac{\sum_l\alpha_{ijk}^l\alpha_{ijk'}^l}{\theta_{ijk'}\sum_{k''l}\alpha_{ijk''}^l}$ (23)

矩阵 M 可以分解成两个矩阵的乘积 $M=QR$。其中，Q 是一个对角阵($Q_{kk'}=1/\theta_{ijk'}=1/\theta_{ijk'}$)，$R$ 是一个 $m\times m$ 矩阵($R_{kk'}=\frac{\sum_l\alpha_{ijk}^l\alpha_{ijk'}^l}{\sum_{k''l}\alpha_{ijk''}^l}$)。令 $\vec{\alpha}_l=(\alpha_{ij1}^l,\alpha_{ij2}^l,\cdots,\alpha_{ijm}^l)$，$\beta=\sum_{kl}\alpha_{ijk}^l$，那么矩阵 $R=\frac{1}{\beta}\sum_l\vec{\alpha}_l^T\vec{\alpha}_l$。显然，如果 $\theta_{ijk}>0$，矩阵 Q 是对称和正定的。而 R 是对称和半正定的，因为对于 $\varepsilon\in\mathbb{R}^m$ 有(24)式成立：

$$\varepsilon R\varepsilon^T=\frac{1}{\beta}\sum_l\varepsilon(\vec{\alpha}_l^T\vec{\alpha}_l)\varepsilon^T=\frac{1}{\beta}\sum_l(\vec{\alpha}_l\varepsilon^T)^T\vec{\alpha}_l\varepsilon^T=\frac{1}{\beta}(\varepsilon\cdot\vec{\alpha}_l)^2\geqslant 0 \tag{24}$$

下面给出本节需要引用的两条引理(不再给出证明过程)。

引理 2 令 A 和 B 分别是两个维数为 $n\times m$ 和 $m\times n$ 的矩阵($m\geqslant n$)，那么 AB 的特征值也是 BA 的特征值。

引理 3 令 $\lambda_{\max}(A)$ 代表矩阵 A 的最大特征值，$\{A_i\}_{i=1}^N$ 是 N 个具有相同维数矩阵的集合。那么，$\lambda_{\max}(\sum_{i=1}^N A_i)\leqslant\sum_{i=1}^N\lambda_{\max}(A_i)$。

根据以上引理，可以得到

$$\lambda_{\max}(R)=\frac{1}{\beta}\lambda_{\max}(\sum_{l=1}^N\vec{\alpha}_l^T\vec{\alpha}_l)$$

$$\leqslant\frac{1}{\beta}\sum_{l=1}^{N}\lambda_{\max}(\vec{\alpha}_l^T\vec{\alpha}_l)=\frac{1}{\beta}\sum_{l=1}^{N}\lambda_{\max}(\vec{\alpha}_l\vec{\alpha}_l^T)$$

$$=\frac{1}{\beta}\sum_{l=1}^{N}\|\vec{\alpha}_l\|_2^2\leqslant\frac{1}{\beta}\sum_{l=1}^{N}\|\vec{\alpha}_l\|_1$$

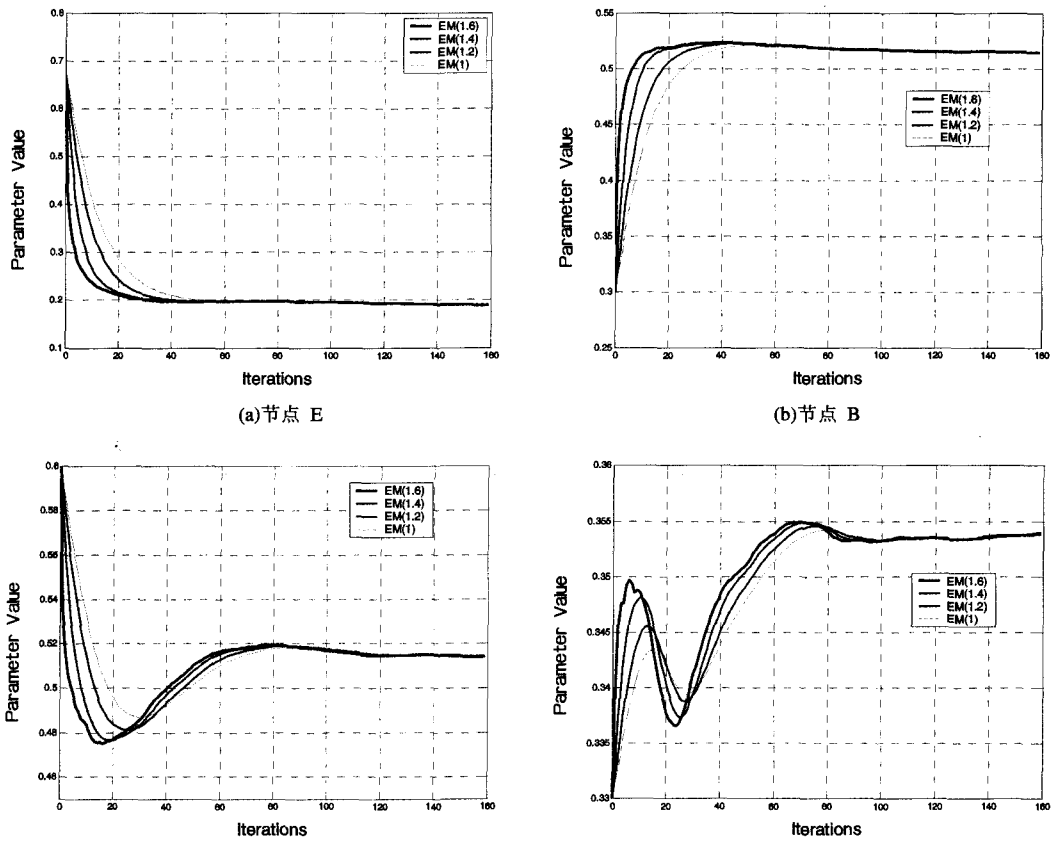
$$=\frac{1}{\beta}\sum_{l=1}^{N}\sum_k\alpha_{ijk}^l=\frac{\beta}{\beta}=1 \qquad (25)$$

令 $\lambda_{\min}$ 和 $\lambda_{\max}$ 分别是 R 的最大和最小特征值。$\nabla\Phi(\theta^*)=I-\eta QR$ 在某一个 Ω 空间的 1 范数 $\rho_\eta=\max\{|1-\eta\lambda_{\max}|,|1-\eta\lambda_{\max}|\}$，当 $\eta\in(0,2)$，均可以保证 $\rho_\eta<1$，满足以下不等式成立：

$$\|\nabla\Phi(\theta^*)(\theta-\theta^*)\|\leqslant\rho_n\|\theta-\theta^*\|<\|\theta-\theta^*\| \qquad (26)$$

由压缩映射定理，当满足 $\eta\in(0,2)$，一定能够确保迭代算子 $EM(\eta)$ 收敛到 θ^*。当 $\eta>2$，由于有可能使 $\rho_n>1$，所以不能保证 $EM(\eta)$ 一定收敛到 θ^*。

4 仿真试验结果



(a)节点 E　(b)节点 B

(c)节点 C　(d)节点 A

图 3　1000 个输入证据样本在随机 30%丢失率条件下的节点参数估计值曲线

表 1　不同学习率 η 条件下的平均收敛迭代次数比较表（证据丢失率 30%）

学习率 η	1	1.2	1.4	1.6
平均收敛迭代次数	60	47.5	37.25	32

表 2　不同证据丢失率条件下的平均收敛迭代次数比较表（学习率 η=1.6）

证据丢失率	5%	10%	15%	20%	25%	30%
平均收敛迭代次数	12	15.5	19	21.75	25	32.25

现将 EM(η)更新规则应用到图 1 所示的 Bayesian 网络中。为了测试算法的性能，本文从图 1 的网络中生成满足高斯分布 X～N(0.5,1.2)的随机训练样本，并且令样本集中的证据样本发生 30%规模的随机部分丢失。为了保证实验结果的可对比性，节点的参数学习曲线都满足相同的训练样本集输入和相同的证据丢失。在网络参数估计实验中，分别选择 1、1.2、1.4、1.6 作为学习率 η 的不同取值。选择图 1 中的节点 E,B,C 和 A 进行 Bayesian 参数估计实验。根据图 3(a)、(b)、(c)、(d)的实验结果，可以发现所有的实验数据均随着学习率 η 的取值增加，相应参数的收敛速度也随之加快。表 1 是 E,B,C,A 节点参数值在不同学习率 η 条件下参数收敛的平均迭代运算次数的比较。表 1 结果显示，当学习率=1.6 时，算法的收敛速度几乎比经典 EM 算法快了 1 倍。表 2 是选择学习率 η=1.6，不同证据丢失率条件下的平均收敛迭代次数对比。表 2 结果表明，随着证据丢失率的增加，EM(η)平均迭代收敛次数呈近似线性增长趋势。这样的算法复杂度确保了在较高证据丢失率情况下，算法仍然能维持相对较小的计算规模（一般当证据丢失率超过 40%，参数估计将基本丧失意义）。根据前一节收敛性能分析，不难理解，学习率 η 的取值越大，EM(η)在做加权平均的时候，迭代步长也会增大，所以参数的收敛速度会更快。但如果存在多个局部极大值，在某一个局部极大值的邻域内，η 取不同的值，将直接影响 EM(η)收敛的结果；如果 η 取值过大，在迭代过程中，可能会越过当前局部极大值点，使迭代的方向出现多次调整，导致计算量反而额外增加。在对节点 C 的仿真过程中就曾出现这样的情况：当 η=1.8 时，迭代收敛次数出现了异常的增加。通过对比实验，我们发现如果在(1,2)区间上选择适当的 η，在不降低参数估计精度的前提下，EM(η)可以大大改善参数

估计的收敛速度。

我们也尝试在 $\eta>2$ 条件下进行以上学习规则的实验。多次实验结果表明，当 $\eta>2$ 时，迭代过程非常不稳定，常常导致参数无法正常收敛。即使偶尔结果收敛，也常常伴随庞大的运算代价，这也验证了本文第 3 节所作的算法收敛性分析的结论是正确的。

讨论和未来的工作 众所周知，在 Bayesian 网络中做不完备证据的参数学习是非常困难的。本文通过构建含有学习率的 EM 算法模型，在可信的程度上解决了不完备证据条件下由于似然函数出现偏倚，导致 Bayesian 参数学习结果不可靠的问题。和标准 EM 算法相比，调整和选择适当的学习率，新算法可以明显加速节点参数学习的收敛速度。这在优化和改善有证据丢失 Bayesian 网络的参数学习和概率推理的运算效率方面，具有重要的现实意义。本文同时也有一些遗留的工作需要在未来的研究中解决。第一个问题就是在给定的 Bayesian 网络中，如何快速准确地选择适当学习率的问题。另一值得研究的问题是动态学习率的问题。本文在前面的试验中采用的都是固定学习率，但在很多实际情况中，如网络在线学习，其学习率应该是可以自适应调整的[15]。这两方面将是未来研究的两个重点课题。

参 考 文 献

1 Jensen F V. An Introduction to Bayesianian Networks. London: UCL Press, 1996. 61～66
2 Lauritzen S L, Spiegelhalter D J. Local computations with probabilities on graphical structures and their application to expert systems. J Roy Statist Soc Ser B, 1988, 50: 45～53
3 Lepar V, Shenoy P P. A Comparison of Lauritzen-Spiegelhalter, Hugin and Shenoy-Shafer Architectures for Computing Marginals of Probability Distributions. In: Cooper G, Moral S, eds. UAI, Morgan Kaufmann, 1998. 328～337
4 Dagum P, Luby M. An optimal approximation algorithm for Bayesianian inference[J]. Artificial Intelligence, 1997. 3～27
5 Androutsopoulos I, Koutsias J, Chandrinos V, et al. An Evaluation of Naive Bayesianian Anti-Spam Filtering. In: Workshop on Machine Learning in the New Information Age. C. 2000. 578～584
6 Murphy K. Active learning of causal Bayesian net structure: [Technical report]. Berkeley: Comp Sci Div, UC, 2001. 3～15
7 Kuo L, Lee J C. Bayes inference for S-shaped software-reliability growth models. IEEE Transactions on Reliability, 1997, 46(1)
8 Glymour C. Learning, prediction and casual Bayes nets. Review, Trends in Cognitive Sciences, 2003, 7(1)
9 Gopnik A, Glymour C. Casual maps and Bayes nets: a cognitive and computational account of theory - formation. In: Carruthers P, et al. eds. the Cognitive Basis of Science, Cambridge University Press, 2002
10 Ahn W, Kalish C. The role of mechanism beliefs in casual reasoning. In: Keil F, Wilson R, eds. Explanation of the cognition, MIT Pess, 2000
11 Spirites P. An anytime algorithm for casual inference. In: Proceedings of the conference on Artificial Intelligence and Statistics, Fort Lauderdale, 2001
12 Sehgal M S B. Collateral missing value imputation: a new robust missing value estimation algorithm for microarray data. Bioinformatics, 2005, 21: 2417～2423
13 Oba S, Sato Masa-aki. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics, 2003, 19: 2088～2096
14 Pe'er D, Regev A, Elidan G, et al, Inferring subnetworks from perturbed expression profiles[J]. Bioinformatics, 2001, 17 (suppl 1), S215
15 Glymour C. The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology, MIT Press, 2001

(上接第 153 页)

(3)此时 Agent 通信与协商语言的外部标准已经基本形成，通过遵循描述原语接口要求的 WSDL 文档构造有效的 Web 服务(包括管理 Agent 的服务和协商 Agent 的服务)，就可以建立起满足 Agent 之间的通信要求。

通过 Web 服务方式使用原语，管理 Agent 与协商 Agent 之间的通信与协商模型图 2 所示。

在图 2 所示是基于 SOA 的协商服务与协商服务管理平台(NSMP)组合后的运作模型图，图中中间部分为协商服务管理平台(NSMP)，上、下部分为协商服务端，它是通过接入 NSMP 而提供服务的，实际上是为参与协商的其它协商服务端、协商服务或协商管理平台提供协商服务，最后构建起多边多议题协商模型。

4 实验及原型评价

为了模拟多个协商 Agent 以协商服务的方式表现出来的协商效果，并验证面向服务架构的协商 Agent 管理平台的可行性、有效性以及可靠性，以 Java/J2EE 体系作为平台的工具，搭建了一个完整的实验平台，通过实验室的网络，共建立了多个个协商服务端，每协商服务端分别注册了 5 个以上以“茶叶”为协商商品的协商服务，同时分别注册了多种情况的协商：有以“价格”、“数量”、“茶叶等级”、“生产日期”共 4 个议题为一个提议的协商；有以“价格”、“数量”、“茶叶等级”共 3 个议题为一个提议的协商；还有其他 6 个，2 个，3 个议题为一个提议的协商等多类型多议题可变可调的协商。通过对该协商模型中协商 Agent 对模拟数据的计算结果的检测，验证了协商 Agent 实现了协商模型中的关键算法，并能为其他协商服务端提供有效的协商服务，同时也验证了该协商模型可以被应用于实际的电子商务的协商中，其协商的结果对用户的真正协商有较强的参考价值。

参 考 文 献

1 Chavez A, Maes P. Kasbah: An Agent Marketplace for Buying and Selling Goods[C]. In: Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology, 1996
2 Guttman R H, Moukas A G, Maes P. Agent-mediated Electronic Commerce: A Survey[J]. Knowledge Engineering Review, 1999, 13(3)
3 Sandholm T. eMediator: A Next Generation Electronic Commerce Server[C]. In: Proc. Nat'l Conf. Artificial Intelligence (AAAI-99), AAAI Press, Menlo Park, Calif., 1999. 923～924
4 Wurman P, Wellman M, Walsh W. The Michigan Internet AuctionBot: A Configurable Auction Server for Human and Software Agents[C]. In: Proceedings of the Second International Conference on Autonomous Agents (Agents-98), Minneapolis, MN, USA, ACM Press, New York, May 1998
5 Sycara K, Zeng Da-jun. Bayesian learning in negotiation. In: Working Notes of the AAAI 1996 Stanford Spring Symposium Series on Adaptation, Co-Evolution and Learning in Multi-Agent Systems. 1996. http://www.ri.cmu.edu/pubs/pub_2186.html
6 Jennings N R, Faratin D, Johnson M J, et al. Agent-Based business process management. Journal of Cooperative Information Systems, 1996, 5(2-3): 105～130
7 Sierra C, Faratin D, Jennings N R. A service-oriented negotiation model between autonomous agents[C]. In: Proceedings of the 8th European Workshop on Modeling Autonomous Agents in a Multi-agent World (MAAMAW'97), 1997. 17～35
8 王立春，陈世福. 多 Agent 多问题协商模型[J]. 软件学报，2002, 13(8): 1637～1643
9 梁茹冰. 基于资源的多 Agent 协商模型的研究[D]: [硕士研究生学位论文]. 2004
10 郭庆，陈纯. 基于整合效用的多议题协商优化[J]. 软件学报，2004, 15(5)
11 OWL Web Ontology Language Reference[Z]. Http://www.w3.org/TR/OWL-REF, 2004-02
12 W3C. Web Services Description Language (WSDL) 1.1[EB/OL]. http://www.w3.org/TR/wsdl 2001.05
13 W3C. SOAP Version 1.2 Part 1[EB/OL]. Messaging Framework——W3C Recommendation http://www.w3.org/TR/soap12-part1/ 2003.06