一种基于随机抽样的贝叶斯网络结构学习算法

胡春玲 胡学钢
(合肥工业大学计算机与信息学院 合肥 230009)

摘 要 针对贝叶斯网络的结构学习问题,基于并行随机抽样的思想提出了结构学习算法PCMHS,构建多条并行的收敛于Boltzmann分布的马尔可夫链。首先基于节点之间的互信息,进行所有马尔可夫链的初始化,在其迭代过程中,基于并行的MHS抽样总体得到产生下一代个体的建议分布,并通过对网络中弧和子结构的抽样产生下一代个体。算法PCMHS收敛于平稳分布,具有良好的学习精度,而该算法又通过使其初始分布和建议分布近似于其平稳分布,有效提高了马尔可夫链的收敛速度。在标准数据集上的实验结果验证了算法PCMHS的学习效率和学习精度明显优于经典算法MHS和PopMCMC。

关键词 贝叶斯网络,结构学习,随机抽样,马尔可夫链,建议分布

Stochastic Sample Based Algorithm for Learning Bayesian Networks

HU Chun-ling HU Xue-gang
(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract Based on the ideas of parallel stochastic sampling, this paper put forward an algorithm PCMHS for learning Bayesian networks. The PCMHS algorithm runs multi parallel Markov chains converging to Boltzmann distributions. The algorithm PCMHS, based on the mutual information between nodes, initializes all Markov chains. In the process of iteration, the algorithm, based on the population from parallel Metropolis-Hasting samplers, generates the proposal distribution for the next generation, and uses arc sample and sub-structure sample to produce the individuals of the next generation. The algorithm PCMHS converges to stationary distribution and its learning accuracy is high. In addition, it designs initial sample and proposal distribution as close as possible to the stationary distribution, and greatly improves the convergence rate. The experimental results on standard data set prove that the learning accuracy and efficiency of the algorithm PCMHS greatly outperform the classical algorithms MHS and PopMCMC.

Keywords Bayesian networks, Structure learning, Stochastic sampling, Markov chain, Proposal distribution

1 引言

贝叶斯网络将概率理论和图形理论相结合,是随机变量之间定性与定量依赖关系的图形表示,它具有形象直观的知识表示形式,以及接近人类思维特征的推理方式,被广泛用于专家系统、模式识别、机器学习和数据挖掘等领域。研究如何根据数据和专家知识高效、准确地建立贝叶斯网络,是十多年来研究热点之一,是贝叶斯网络更加广泛、有效地用于实际问题领域的关键引人注目的焦点之一[1,2]。目前,对于这一类学习问题,主要有基于打分-搜索的学习方法和基于依赖分析的学习方法[3,4]。基于打分-搜索的学习方法过程简单、规范,但存在搜索空间巨大,可能收敛于局部最优解等问题;基于依赖分析的学习方法学习效率较高,而且能够获得全局最优解,但存在节点之间的独立性或条件独立性判断困难,高阶的条件独立性检验的结果不够可靠等问题。

MCMC(Markov Chain Monte Carlo)方法[5,6]是源于统计物理学和生物学的一类重要的随机抽样方法,该方法广泛应用于机器学习、统计和决策分析等领域的高维问题的推理和求积运算。MHS (Metropolis-Hasting Sampler)抽样算法[7]作为MCMC方法中常用的抽样方法之一,通过构建一条马尔可夫链,模拟一个收敛于Boltzmann分布的系统。收敛之后的样本为来自于这一平稳分布的抽样,因而能够较好地保证样本的多样性,所得样本可以直接用来对平稳分布进行矩估计,避免了高维积分的计算。该算法被评为20世纪对科学和工程领域产生重大影响的十大算法之一[8]。Madigan首次将MHS抽样算法引入贝叶斯网络结构学习[9],该算法采用局部的弧增加、删除和反向的均匀分布作为抽样过程的建议分布,利用抽样过程产生的来自目标平稳分布的网络结构样本来估计贝叶斯网络的结构特征,因此,MHS抽样算法在收敛之后具有良好的学习精度,但MHS抽样过程的融合性差,收敛速度慢。此后,针对MHS抽样算法的存在的问题,出现了不同的改进算法[10,11],这些算法都着力于改善MHS抽样算法的收敛速度,其中具有代表性的是由Myers提出的PopMCMC算法[12]。总之,将MHS抽样算法引入贝叶斯网络结构学习

到稿日期:2008-05-21 本文受安徽省自然科学基金课题(编号050420207)资助。
胡春玲(1970-),女,博士研究生,主要研究方向为数据挖掘与贝叶斯网络;胡学钢(1961-),男,教授,硕士生导师,CCF会员,主要研究方向为知识工程与数据挖掘。

能够较好地解决进化学习方法中由于个体趋同而产生的早熟问题，保证算法的学习精度，但该算法存在的收敛速度慢和收敛性判断困难等问题仍未能得到有效解决。如何更有效地将MHS算法用于贝叶斯网络的结构学习成为近年来引人注目的研究方向之一。

初始值和建议分布是影响MHS算法收敛速度的重要因素，若建议分布等于目标平稳分布，此时抽样成为来自于目标平稳分布的独立抽样[6]，因而具有较快的收敛速度。本文提出的算法PCMHS(Parallel Crossover Metropolis-Hasting Sampler)采用并行迭代抽样方式，通过对初始样本和建议分布的设计来改善抽样过程的收敛速度，并通过对网络中子结构的抽样来改善抽样过程的融合性，提高算法的学习效率。从单个MHS抽样的角度来看，算法PCMHS是自适应的，具有较快的收敛速度；从总体的角度来看，该算法具有固定的转移概率，且转移概率均不为零，因而保证了抽样过程的遍历性和收敛性。通过与同类算法的实验比较，也验证了算法PCMHS明显改善了抽样算法迭代过程的收敛速度，并具有较好的学习精度。

为了叙述方便，本文采用 $X=\{X_1,X_2,\cdots,X_n\}$ 表示领域变量，$r_1,r_2,\cdots,r_n$ 为相应变量可能的取值情况，$D=\{D_1,D_2,\cdots,D_m\}$ 为训练数据集，Π_{X_i} 为网络结构 s 中节点 X_i 的父节点集，$\theta_{ijk}=P(X_i=x_{ik}\mid\Pi_{X_i}=\Pi_{X_i})$ 表示 X_i 取其第 k 个取值，而 Π_{X_i} 取其第 j 个取值时的条件概率。$\theta_{ij}=(\theta_{ij1},\theta_{ij2},\cdots,\theta_{ijr_i})$，显然有：$0\leqslant\theta_{ijk}\leqslant1$，$\sum_k\theta_{ijk}=1$。

2 背景知识

2.1 BDe评分

BDe(Bayesian Dirichlet equivalent)评分又称贝叶斯评分，是贝叶斯网络结构的一种评分方法，该方法源于贝叶斯公式。BDe评分能够充分结合关于网络结构的先验知识，首先将网络结构 s 的先验知识表示为先验概率 $P(s)$，利用贝叶斯公式，当给定数据集 D 时，网络结构 s 的后验概率计算如下：

$$P(s|D)=\frac{P(s)P(D|s)}{P(D)} \tag{1}$$

其中：

$$P(D\mid s)=\int P(D\mid s,\theta_s)P(\theta_s\mid s)\mathrm{d}\theta_s \tag{2}$$

因为 $P(D)$ 与网络拓扑结构无关，从公式(1)可以看出，BDe评分可以定义如下：

$$Score_{\mathrm{BDe}}(s|D)=\ln(P(s)P(D|s))=\ln P(s)+\ln P(D|s) \tag{3}$$

如果所有网络拓扑结构 s 的先验概率 $P(s)$ 都相等时，则可取 $\ln P(D|s)$ 作为BDe评分。在样本集 D 完备的前提下，假设参数 θ 满足局部独立性和全局独立性假设，并且其先验概率服从Dirichlet分布，即：

$$P(D|s)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})} \tag{4}$$

则：

$$P(\theta_{ij}\mid s)=Dir(\alpha_{ij1},\alpha_{ij2},\cdots,\alpha_{ijr_i})=\frac{\Gamma(\alpha_{ij})}{\prod_k\Gamma(\alpha_{ijk})}\prod_k(\theta_{ijk})^{\alpha_{ijk}}$$

其中，$\alpha_{ij}=\sum_k\alpha_{ijk}$，$N_{ij}=\sum_{k=1}^{r_i}N_{ijk}$ (5)

2.2 MHS抽样算法

MHS(Metropolis-Hasting Sampler)抽样算法作为MC-MC方法中常用的抽样方法之一，通过构建一条收敛于平稳分布的马尔可夫链，模拟一个收敛于Boltzmann分布的系统。若系统温度为 T，$E(s)$ 表示状态空间 S 上任一状态 s 的能量，则状态空间上的平稳Boltzmann分布为：

$$P(s)=\frac{1}{Z}\exp(\frac{-E(s)}{T}),\forall s\in S$$

其中，Z 为正则化因子。

MHS算法的抽样过程如下：

(1)记当前状态为 $s^{(c)}$，根据建议分布 $R(s^{(n)}\mid s^{(c)})$，生成下一个状态 $s^{(n)}$。

(2)根据概率规则，决定是否接受生成的新状态。新状态的接受概率计算公式为：

$$A(s^{(n)}\mid s^{(c)})=\min\left[1,\frac{p(s^{(n)})R(s^{(c)}\mid s^{(n)})}{p(s^{(c)})R(s^{(n)}\mid s^{(c)})}\right] \tag{6}$$

则转移概率为：

$$T(s^{(n)}\mid s^{(c)})=\begin{cases}R(s^{(n)}\mid s^{(c)})A(s^{(n)}\mid s^{(c)}) & s^{(n)}\neq s^{(c)}\\ 1-\sum_{s^{(n')}\neq s^{(c)}}R(s^{(n')}\mid s^{(c)})A(s^{(n')}\mid s^{(c)}) & s^{(n)}=s^{(c)}\end{cases} \tag{7}$$

(3)若新状态被接受，则取代原有状态；若被拒绝，则保留原有状态。

3 贝叶斯网络的结构学习算法PCMHS

针对贝叶斯网络结构学习问题，本文提出的PCMHS抽样学习算法的状态空间为 $S=\{s\}$，其中 s 为网络结构，对任意的 $s\in S$，其能量函数定义为 $E(s)=-\Omega(s|D)$，其中 $\Omega(s|D)$ 是关于网络结构 s 的BDe评分，一般假设系统温度 T 为1。PCMHS抽样算法是一种由若干个并行MHS抽样构成的元MHS抽样，该抽样算法的目标平稳分布为网络结构的后验概率分布 $P(s|D)$，即有：$p(s)=\frac{1}{Z}(\exp(\frac{-E(s)}{T})=\frac{1}{Z}\{\exp(\sigma(s|D))\}=p(s|D)$，但初始值和建议分布是影响抽样算法收敛速度的重要因素。算法PCMHS通过对初始值和建议分布的设计来改善算法的学习效率。该算法首先基于节点之间的互信息，在缩减的状态空间中进行网络结构的初始化，然后进入并行的迭代抽样过程，每一次迭代都基于来自当前样本总体的建议分布 $R(s^{(n)}\mid s^{(c)},\psi)$，从当前样本的每一个个体出发，对网络中的弧和子结构进行MHS抽样，构建多条并行的收敛于 $p(s|D)$ 分布的马尔可夫链。

3.1 算法PCMHS的初始化

互信息[13]是节点之间直接或间接的信息流量，度量了节点之间依赖程度。如果节点之间的互信息或条件互信息小于某个预定的阈值 ε(取值一般介于0.001～0.05)，那么在网络中这两个节点之间就不存在边。互信息的计算公式如下：

$$I(X_i,X_j)=\sum_{X_i,X_j}P(X_i,X_j)\log\frac{P(X_i,X_j)}{P(X_i)P(X_j)} \tag{8}$$

当数据集完备时，可以通过扫描一遍数据库，得到互信息计算所需的概率参数。

算法PCMHS通过互信息 $I(X_i,X_j)$ 来缩减搜索空间，生成关于网络结构的初始样本。该算法首先通过最大生成树算法，生成尽可能多的互异的样本个体，再基于节点之间的互信息，根据预定的阈值 ε，去掉一些在网络结构中不可能存在的边，此后在缩减的状态空间中，采用随机的方法生成互异的样本个体，这样生成的初始样本比完全随机生成的样本更加接

近于该抽样过程的平稳分布 $P(s|D)$，则基于该样本总体计算的建议分布 $R(s^{(n)}|s^{(c)},\psi)$ 也更近似于网络结构的平稳分布 $P(s|D)$。

3.2 算法 PCMHS 的迭代抽样过程

从接受概率的计算公式(6)可以发现：若建议分布等于目标平稳分布，则接受概率等于 1，此时抽样成为来自于目标平稳分布的独立抽样，抽样过程具有较快的收敛速度。然而，在实际抽样问题中，因为目标分布事先并不知道，选择等于目标平稳分布的建议分布是不可能的。事实上，只要使建议分布近似于目标分布，也可以有效地提高收敛速度。算法 PCMHS 运用所有并行 MHS 上一次的抽样总体去估计目标平稳分布的特征，试图选择与目标平稳分布尽可能接近的建议分布。

假设并行 PCMHS 抽样算法的建议分布簇为 $R(s^{(n)}|s^{(c)},\psi)$，其中 ψ 为建议分布参数，若该并行抽样的总体长度为 τ，建议分布的参数函数为 $\hat{\psi}(s_1,s_2,\cdots,s_\tau)$，则在其状态元组 $(s_1,s_2,\cdots,s_\tau)$ 的当前取值为 $(s_1^{(c)},s_2^{(c)},\cdots,s_\tau^{(c)})$ 的条件下，$s_i^{(n)}$ 的建议分布是 $R(s_i^{(n)}|s_i^{(c)},\hat{\psi}(s_1^{(c)},s_2^{(c)},\cdots,s_\tau^{(c)}))$，即：基于抽样总体选择尽可能符合目标平稳分布的特征的参数估计来定义建议分布，在算法的迭代过程中，随着样本质量的不断优化，可以使得建议分布簇 $R(s^{(n)}|s^{(c)},\psi)$ 近似独立于 $s^{(c)}$ 而无限接近于 $P(s|D)$，这时抽样过程的收敛速度接近于独立抽样的收敛速度。

在进化学习设计思想中，总是希望好的基因块能够在下一代的个体中保留下来，进而减少学习过程的迭代次数，提高学习效率。这一思想同样可以用于并行抽样，来自并行抽样的不同个体之间的信息交换有助于在下一代的个体中保留好的子结构[7]，进而提高算法的收敛速度。在算法 PCMHS 迭代过程中，当前样本中的一部分个体基于来自样本总体的弧信息进行抽样，另一部分个体基于当前样本总体的子结构信息进行抽样，即：选择一定比例的个体进行交叉操作，在进行交叉操作的两个网络中交换部分节点的父节点集，交叉产生的新个体仍按照式(6)计算的接受概率决定是否被接受，其建议分布的计算也同样基于上一代抽样总体，对网络子结构的抽样可提高抽样过程的融合性，使得好的子结构在下一代的个体中保留下来。具体来说，并行抽样算法 PCMHS 基于总体中弧的频度来设计单链抽样的建议分布，基于局部的子结构的频度来设计双链交叉的建议分布。

(1)记 π_{ij} 为分布 $P(s|D)$ 中存在从节点 X_i 到 X_j 的有向弧的概率，A_{ij} 为当前总体中存在从节点 X_i 到 X_j 的有向弧的个体数，则 $(A_{ij},A_{ji},\tau-A_{ij}-A_{ji})$ 服从多项式分布 $(\pi_{ij},\pi_{ji},1-\pi_{ij}-\pi_{ji})$，假设概率向量 $(\pi_{ij},\pi_{ji},1-\pi_{ij}-\pi_{ji})$ 服从超系数均为 1 的 Dirichlet 先验分布，则其后验概率服从超系数为 $(1+A_{ij},1+A_{ji},1+\tau-A_{ij}-A_{ji})$ 的 Dirichlet 分布，此时 π_{ij},π_{ji} 的后验估计分别为：

$$\hat{\pi}_{ij}=E(\pi_{ij}|A_{ij})=\frac{A_{ij}+1}{\tau+3}$$
$$\hat{\pi}_{ji}=E(\pi_{ji}|A_{ji})=\frac{A_{ji}+1}{\tau+3} \tag{9}$$

以上弧概率的期望值将取代 MHS 抽样中弧增加、弧删除和反向的均匀分布，作为 PCMHS 算法中基于弧抽样的建议分布。

(2)记 φ_Y 为分布 $P(s|D)$ 中节点子集 Y 的某种父子结构的概率，B_Y 为当前抽样总体中存在此种父子结构的的个体数，则 $(B_Y,\tau-B_Y)$ 服从多项式分布 $(\varphi_Y,1-\varphi_Y)$。假设概率向量 $(\varphi_Y,1-\varphi_Y)$ 服从超系数均为 1 的 Dirichlet 先验分布，则其后验概率服从超系数为 $(1+B_Y,1+\tau-B_Y)$ 的 Dirichlet 分布，此时 φ_Y 的后验估计为：

$$\hat{\psi}_Y=E(\varphi_Y|B_Y)=\frac{B_Y+1}{\tau+2} \tag{10}$$

该后验估计作为 PCMHS 算法中双链交叉操作中基于网络中父子结构抽样的建议分布。

总之，算法 PCMHS 的主要步骤如下：

初始化：

(1)采用最大生成树算法生成初始样本的部分个体；

(2)基于节点之间的互信息，在缩减的状态空间中随机生成初始样本的其余个体。

迭代阶段：

(1)对上一代总体中的随机选择部分个体，按后验概率估计式(9)定义的建议分布进行弧抽样，新个体是否接受取决于按式(6)计算的接受概率；

(2)对上一代总体中的另一部分个体按后验概率估计式(10)定义的建议分布进行子结构的抽样，新个体是否接受取决于按式(6)计算的接受概率。

直到算法收敛或已达到预计的迭代次数。

3.3 算法 PCMHS 的性能分析

首先证明 PCMHS 抽样算法的收敛性，以下定理可以证明 PCMHS 抽样算法是收敛的，且收敛后 Boltzmann 平稳分布为关于网络结构的后验分布 $p(s|D)$。

定理 1 PCMHS 抽样算法的转移概率具有局部可逆性，即满足：

$$P(s^{(c)}|D)T(s^{(n)}|s^{(c)})=P(s^{(n)}|D)T(s^{(c)}|s^{(n)})$$

证明：

$$P(s^{(c)}|D)T(s^{(n)}|s^{(c)})=P(s^{(c)}|D)R(s^{(n)}|s^{(c)})\cdot \min\{1,\frac{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})}{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}\}$$
$$=P(s^{(c)}|D)R(s^{(n)}|s^{(c)})\frac{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})}{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}\cdot \min\{\frac{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})},1\}$$
$$=P(s^{(n)}|D)R(s^{(c)}|s^{(n)})\cdot\min\{1,\frac{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})}\}$$
$$=P(s^{(n)}|D)T(s^{(n)}|s^{(c)})$$

故得证。

定理 2 PCMHS 抽样算法收敛于平稳分布 $P(s|D)$。

证明：

$$\sum_c P(s^{(c)}|D)T(s^{(n)}|s^{(c)})=\sum_c P(s^{(c)}|D)R(s^{(n)}|s^{(c)})\cdot \min\{1,\frac{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})}{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}\}$$
$$=\sum_c P(s^{(c)}|D)R(s^{(n)}|s^{(c)})\frac{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})}{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}\cdot \min\{\frac{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})},1\}$$
$$=\sum_c P(s^{(n)}|D)A(s^{(c)}|s^{(n)})\cdot\min\{1,\frac{P(s^{(c)}|D)R(s^{(n)}|s^{(c)})}{P(s^{(n)}|D)R(s^{(c)}|s^{(n)})}\}$$
$$=\sum_c P(s^{(n)}|D)T(s^{(n)}|s^{(c)})=P(s^{(n)}|D)\sum_c T(s^{(n)}|s^{(c)})$$
$$=P(s^{(n)}|D)$$

故得证，即 $P(s|D)$ 是 PCMHS 抽样的平稳分布。

在保证收敛性的前提条件下，算法 PCMHS 从并行抽样的初始样本、建议分布和对网络子结构的抽样来提高迭代过程的收敛速度。在其迭代过程中，该算法将上一代抽样总体中弧和子结构的后验概率估计作为对下一代个体抽样的建议

分布。从总体的角度来看，算法 PCMHS 是一个具有固定转移概率的马尔可夫链，从而保证了其迭代抽样过程的遍历性和平稳性，从个体角度来看该建议分布是自适应的，因为该分布依赖于当前总体中弧和局部结构的全局信息，其自适应规则是基于弧和局部子结构的频度。相对于非自适应的 MHS 抽样，算法 PCMHS 性能的改进程度依赖于建议分布对 $P(s|D)$ 的近似程度，算法 PCMHS 基于节点之间的互信息生成的初始样本比完全随机生成的初始样本更加接近于网络结构的真实分布，在此基础上产生的建议分布一开始就能较好地近似于网络结构的真实分布，基于此建议分布产生的下一代总体的建议分布就能更好地拟合网络结构的真实分布，从而提高迭代过程的收敛速度。以下的实验结构也验证了算法 PCMHS 在保证学习精度的前提下，有效地提高了迭代过程的收敛速度。

4 算法 PCMHS 的实验结果分析

Asia 网络是一个用来验证贝叶斯网络学习算法的标准网络之一，该网络有 8 个节点，8 条边。根据网站 http://www.norsys.com 提供的 Asia 网概率分布表生成具有 10000 个实例的训练数据集和 1000 个实例的测试数据集，实验中抽样过程的迭代次数为 600 次，并行抽样产生的样本大小为 40，MHS 算法初始的 50 次迭代不计入迭代次数。

算法的收敛速度是衡量随机抽样算法性能的一个重要指标。PCMHS 算法的主要优点是有效提高了迭代过程的收敛速度。PopMCMC 算法是由 Myers 提出的一种在数据缺失条件下网络结构学习的并行抽样算法。该算法随机生成初始样本，基于网络结构中的弧进行抽样，但该算法在基于完备数据集的贝叶斯网络学习中，与同类算法相比，未能有效地改善收敛速度，原因在于，基于随机生成的初始样本的建议分布不能很好近似于目标平稳分布 $P(s|D)$，且弧抽样又不能有效地融合不同个体的优良子结构，导致迭代过程的收敛速度慢。另外，为了验证 PCMHS 算法中子结构抽样的有效性，设计算法 PMHS，该算法与 PCMHS 唯一不同的是，并行抽样的所有个体仅进行弧抽样。图 1 表示在有 10000 条记录的 Asia 数据集上，对算法 PCMHS，PMHS，PopMCMC 和 MHS 的收敛速度的比较。图中的水平线是真实 Asia 网络的 BDe 得分，算法 PCMHS 在迭代 150 次左右即收敛，PMHS 的迭代速度与 PCMHS 相对较为接近，但其收敛速度还是明显慢于 PCMHS，而 PopMCMC 和 MHS 算法 600 次迭代后仍远未收敛。

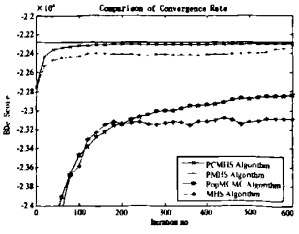

图 1 算法 PCMHS，PMHS，PopMCMC 和 MHS 的收敛速度比较

Log Loss 是衡量网络结构对测试数据集预测精度的一个重要指标，也是衡量算法学习精度的一个性能指标，其计算公式为

$$L=E(-\log(T|D))\approx-\frac{1}{m}\sum_{i=1}^{m}\log P(t_i|D)$$

其中，T 为测试数据集，其实例数为 m，t_i 为 T 中任一条实例，D 为训练数据集，L 的值越小，预测精度越高（若在 L 的计算公式去掉负号，则 L 值越大，预测精度越高）。以下实验给出了中采用有 1000 个实例的 Asia 数据集作为测试集，图 2 给出了算法 PCMHS，PMHS，PopMCMC 和 MHS 的预测精度的比较，显然 PCMHS 在测试数据集的 Log Loss 始终最小，预测精度最高。

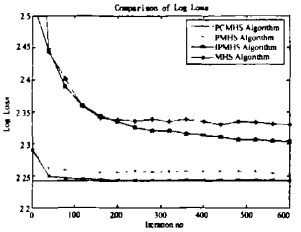

图 2 算法 PCMHS，PMHS，PopMCMC 和 MHS 的预测精度比较

结束语 收敛速度慢是随机抽样算法目前存在的主要问题。算法 PCMHS 从初始样本、建议分布和对网络中子结构的抽样 3 个方面对 MHS 抽样算法进行改进，分析和实验结果都证明了该算法能有效地提高迭代过程的收敛速度，同时，算法 PCMHS 在理论上能够保证收敛于平稳分布，因而具有良好的学习精度。下一步的工作将围绕如何有效地将该算法用于生物信息学领域的基因调控网络的构建。

参 考 文 献

[1] Chickering D, Herkerman D, Meek C. Large-sample Learning of Bayesian Networks is NP-Hard. Machine Learning Research [J], 2004, 5: 1287-1330

[2] Ellis B, Wong W. Sampling Bayesian Networks Quickly[C]// Interface. 2006

[3] Wong M L, Leung K S. An efficient data mining method for learning Bayesian Networks using an evolutionary algorithm-based hybrid approach[J]. IEEE Transactions on Evolutionary Computation, 2004, 8: 378-404

[4] Peng H, Ding C. Structure search and stability enhancement of Bayesian networks[C]// Proc. of the Third IEEE International Conference on Data Mining. 2003: 621-624

[5] Andrieu C, Freitas N D, Doucet A, et al. An introduction to MCMC for machine learning[J]. Machine Learning, 2003, 50: 5-43

[6] Robert C, Casella G. Monte Carlo statistical methods[M]. 2nd edition. Springer, 2004

[7] Gilks W, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo methods. Practice[M]. CRC Press, 1996

[8] Madigan D, York J. Bayesian graphical models for discrete data [J]. Intl. Statistical Review, 1995, 63: 215-232

[9] Beichl I, Sullivan F. The Metropolis algorithm[J]. Computing in Science & Engineering, 2000: 65-69

[10] Giudici P, Castelo R. Improving Markov Chain Monte Carlo Model Search for Data Mining[J]. Machine Learning, 2003, 50 (1/2): 127-158

[11] Friedman N, Koller D. Being Bayesian about network structure: A Bayesian Approach to Structure Discovery in Bayesian Networks[J]. Machine Learning, 2003, 50: 95-126

[12] Myers J W. Population Markov Chain Monte Carlo[J]. Machine Learning, 2003, 50: 175-196

[13] Cheng J, Greiner R, Kelly J. Learning Bayesian Networks from Data: An Efficient Information-theory Based Approach[J]. Artificial Intelligence, 2002, 137(1/2): 43-90