

# 基于双尺度约束模型的 BN 结构自适应学习算法

戴晶帼<sup>1,2</sup> 任佳<sup>1</sup> 董超<sup>2</sup> 杜文才<sup>1,3</sup>

**摘要** 在无先验信息的情况下, 贝叶斯网络 (Bayesian network, BN) 结构搜索空间的规模随节点数目增加呈指数级增长, 造成 BN 结构学习难度急剧增加. 针对该问题, 提出基于双尺度约束模型的 BN 结构自适应学习算法. 该算法利用最大互信息和条件独立性测试构建大尺度约束模型, 完成 BN 结构搜索空间的初始化. 在此基础上设计改进遗传算法, 在结构迭代优化过程中引入小尺度约束模型, 实现结构搜索空间小尺度动态缩放. 同时, 在改进遗传算法中构建变异概率自适应调节函数, 以降低结构学习过程陷入局部最优解的概率. 仿真结果表明, 提出的基于双尺度约束模型的 BN 结构自适应学习算法能够在无先验信息的情况下保证 BN 结构学习的精度和迭代寻优的收敛速度.

**关键词** 贝叶斯网络, 结构学习, 约束模型, 遗传算法

**引用格式** 戴晶帼, 任佳, 董超, 杜文才. 基于双尺度约束模型的 BN 结构自适应学习算法. 自动化学报, 2021, 47(8): 1988–2001

**DOI** 10.16383/j.aas.c180226

## BN Structure Adaptive Learning Algorithm Based on Dual-scale Constraint Model

DAI Jing-Guo<sup>1,2</sup> REN Jia<sup>1</sup> DONG Chao<sup>2</sup> DU Wen-Cai<sup>1,3</sup>

**Abstract** In the absence of prior information, the size of the search space for Bayesian network (BN) structures grows exponentially with the increasing number of nodes, resulting in the great difficulties of BN structure learning. To solve this problem, BN structure adaptive learning algorithm based on dual-scale constraint model is proposed. The maximum mutual information and conditional independence tests are first used in the proposed algorithm to build a large-scale constraint model, completing the initialization of the search space for BN structures. Based on this, an improved genetic algorithm is presented, introducing a small-scale constraint model during the iteration to realize the dynamic adjustment of the search space. At the same time, an adaptive control function of mutation probability is constructed in order to reduce the probability of getting trapped in a local optimum. The simulation results show that BN structure adaptive learning algorithm based on dual-scale constraint model can guarantee the accuracy of learning BN structure and the convergence speed of the iteration without prior information.

**Key words** Bayesian network, structure learning, constraint model, genetic algorithm

**Citation** Dai Jing-Guo, Ren Jia, Dong Chao, Du Wen-Cai. BN structure adaptive learning algorithm based on dual-scale constraint model. *Acta Automatica Sinica*, 2021, 47(8): 1988–2001

贝叶斯网络 (Bayesian network, BN) 理论为不

确定性知识的表示, 节点间复杂关系的表达和信息融合推理提供了有效的模型框架<sup>[1–3]</sup>. 然而, 与其他图形化模型学习方法类似, BN 也存在研究对象节点增加导致候选结构数量呈指数级增长的问题. 因此, 从大规模候选结构集合中快速、准确地获得一个与数据样本匹配程度最优的结构模型是 BN 结构学习亟待解决的问题之一.

研究人员尝试利用条件独立性 (Conditional independence, CI) 测试度量节点间的依赖关系, 然后构造满足这些依赖关系的模型结构<sup>[4–6]</sup>. 但该类型学习方法运算次数一般为节点个数的指数次幂, 当面对节点数量较多的复杂网络时, CI 测试效率较低. 与此同时, 研究人员将模型先验信息作为约束条件, 并与结构评分函数和搜索算法组合, 共同完成 BN 结构学习<sup>[7–10]</sup>. 但该类型方法在模型先验信息不准确或发生错误的情况下, 将导致结构学习结果精度低或陷入局部最优. 为此, 文献 [11] 通过构建不

收稿日期 2018-04-17 录用日期 2019-02-25

Manuscript received April 17, 2018; accepted February 25, 2019  
国家国际科技合作专项 (2015DFR10510), 国家自然科学基金 (61562018), 国家海洋局南海维权技术与重点实验室开放基金 (1704), 海口市重点科技计划项目 (2017041) 资助

Supported by International Science and Technology Cooperation Projects of China (2015DFR10510), National Natural Science Foundation of China (61562018), Open Foundation of Key Laboratory of Technology and Application for Safeguarding of Marine Rights and Interests (1704), Key Science and Technology Projects of Haikou, Hainan Province (2017041)

本文责任编辑 朱军

Recommended by Associate Editor ZHU Jun

1. 海南大学信息与通信工程学院 海口 570228 中国 2. 自然资源部海洋环境探测技术与应用重点实验室 广州 510310 中国 3. 澳门城市大学数据科学研究院 澳门 999078 中国

1. School of Information and Communication Engineering, Hainan University, Haikou 570228, China 2. Key Laboratory of Marine Environmental Survey Technology and Application, Ministry of Natural Resources, Guangzhou 510310, China 3. Institute of Data Science, City University of Macau, Macau 999078, China

确定先验信息的评分函数并设计先验搜索算子, 提高对错误先验信息的适应能力. 但该方法并不能完全消除错误先验信息对 BN 结构学习产生的负面影响. 此外, 研究人员将节点间依赖关系检验与结构优化学习方法相结合, 进行无先验信息情况下的 BN 结构学习<sup>[12-13]</sup>. 在该思路下, 文献 [14] 提出了一种混合式 BN 结构学习方法, 该方法通过计算互信息量 (Mutual information, MI) 构造基于支撑树权重矩阵的节点序适应度函数, 利用改进遗传算法 (Genetic algorithm, GA) 获得节点排序, 最后将节点排序作为 K2 算法的输入进行结构学习; Wong 等和 Ji 等分别将低阶 CI 测试引入进化算法和蚁群优化算法, 利用 CI 测试辨识部分节点对的独立关系以缩小结构搜索空间, 在此基础上利用评分函数和启发式搜索以获得 BN 结构<sup>[15-16]</sup>; Li 等则提出 I-GREEDY-E 算法<sup>[17]</sup>, 该算法首先在空图上计算节点间的 MI 以确定无向边, 然后利用 CI 测试确定边的指向, 最后在等价类模型空间中利用贪婪搜索算法获得 BN 结构. 上述学习方法在无先验信息的情况下实现了对结构搜索空间的约束, 并在约束空间内利用搜索方法获得了模型结构, 一定程度上提高了多节点复杂网络结构的学习效率. 然而, 在无先验信息的情况下, 以 CI 测试或 MI 检验为约束条件构造的结构搜索空间尺度固定, 无法在结构学习过程中动态调整结构搜索空间的规模大小, 易导致潜在最优解丢失.

针对上述问题, 提出基于双尺度约束模型的 BN 结构自适应学习算法 (BN structure adaptive learning algorithm based on dual-scale constraint model, DSC-AL). 该算法将最大互信息 (Maximum mutual information, MMI) 和 CI 测试结合建立结构搜索空间的大尺度约束模型, 限制 BN 结构搜索空间的规模, 完成结构搜索空间的初始化. 借鉴 GA 迭代寻优过程完成 BN 结构学习. 在迭代寻优过程中建立小尺度约束模型完成搜索空间动态调整, 构建变异概率的自适应调节函数, 提高 GA 全局搜索能力. 在实验仿真中, 选择了不同节点规模的标准 BN 模型和不同样本量对 DSC-AL 算法的有效性进

行测试, 并通过与 Lee 等提出的基于双重 GA 编码的 BN 结构学习算法<sup>[18]</sup> 和 Gheisari 等构建的基于粒子群优化的 BN 结构学习方法<sup>[19]</sup>, 以及经典的 K2 算法<sup>[20]</sup> 进行性能对比, 验证 DSC-AL 算法的性能.

## 1 问题描述

为达到大尺度约束模型完成结构搜索空间的初始化, 小尺度约束模型实现结构搜索空间动态缩放的目的, 在观测样本驱动下, DSC-AL 算法将 MI 和 CI 测试相结合辨识节点间的依赖或独立关系, 形成限制结构搜索空间的大尺度约束模型. 然而受 MI 和 CI 测试方法辨识精度的影响, 可能导致结构搜索空间约束不合理, 将会出现以下两种情况, 即约束尺度过大将最优结构排除在结构搜索空间之外, 或约束尺度较小导致结构搜索空间规模依然较大. 因此, 还需要建立小尺度约束模型完成结构搜索空间动态缩放. DSC-AL 算法框架如图 1 所示.

根据图 1 给出的算法框架, 大尺度约束模型作用于算法的初始阶段, 通过剔除不满足约束条件的 BN 模型限制结构搜索空间, 而小尺度约束模型针对 GA 种群中的每一个个体, 在迭代寻优过程中, 通过删除不满足约束条件的连接边调整每个个体表示的 BN 结构. DSC-AL 算法为实现结构搜索空间小尺度动态缩放, 降低 GA 结构寻优过程发生早熟收敛的概率, 需要解决以下三个关键问题:

- 1) 小尺度约束模型需要随 GA 学习进程不断调整, 达到动态缩放结构搜索空间的目的;
- 2) GA 学习进程中若产生新个体, 需要反复检测无环性, 并使用修复算子对无效结构进行无环修复, 这将导致算法的时间成本增加, 学习效率下降;
- 3) GA 学习进程需要保持种群多样性以及减少对优势基因的破坏, 以降低学习方法陷入局部最优的可能性.

## 2 双尺度约束模型

### 2.1 大尺度约束模型

包含  $n$  个节点的 BN 可生成的结构个数如式

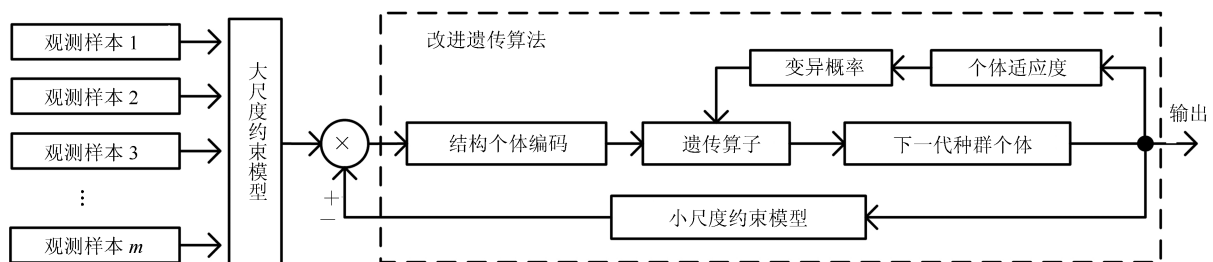


图 1 DSC-AL 算法框架示意图

Fig. 1 The framework of DSC-AL algorithm

(1) 所示<sup>[21]</sup>:

$$f(n) = \begin{cases} 1, & n = 0 \\ \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i), & n > 0 \end{cases} \quad (1)$$

由上式可知, BN 结构搜索空间规模随着节点个数的增加呈指数级增长. 为限制 BN 结构搜索空间, 利用节点连接边存在性约束与节点连接边缺失性约束<sup>[22]</sup> 构建大尺度约束模型.

1) 节点连接边存在性约束

BN 中的节点  $x$  与节点  $y$  之间的 MI 可以采用式 (2) 进行计算:

$$I(x; y) = \sum_{i=1}^r \sum_{j=1}^s p(a_i, b_j) \log_2 \frac{p(a_i, b_j)}{p(a_i)p(b_j)} \quad (2)$$

式中:  $r$  和  $s$  分别表示节点  $x$  与节点  $y$  的状态个数,  $a_i$  是  $x$  的第  $i$  个可能状态值,  $b_j$  是  $y$  的第  $j$  个可能状态值. 两个节点之间的 MI 越大, 两者之间的依赖性越强, 即两个节点间可能存在一条连接边.

2) 节点连接边缺失性约束

CI 测试在给定观测样本和约束集  $Z$  的基础上, 采用卡方检验判定原假设  $H_0: I(x, y | Z)$  是否成立. 式 (3) 中统计量  $\chi^2$  服从自由度为  $(r-1) \times (s-1) \times t$  的分布, 其中  $t$  表示约束集  $Z$  的状态个数, 即

$$\chi^2 = 2 \sum_{a,b,c} m_{xyZ}^{abc} \log_2 \frac{m_{xyZ}^{abc} m_Z^c}{m_{xZ}^{ac} m_{yZ}^{bc}} \quad (3)$$

式中  $m_{xyZ}^{abc}$  表示观测样本中满足节点  $x$  与  $y$  的状态值分别为  $a$  和  $b$ , 且约束集  $Z$  的状态值为  $c$  的观测样本个数. 给定显著性水平  $\alpha$  ( $0 < \alpha < 1$ ), 当  $\chi^2 < \chi_\alpha^2$ , 接受原假设, 即  $I(x, y | Z)$  成立, 这就表示在给定约束集  $Z$  的条件下, 节点  $x$  和  $y$  独立. 此时, 对应的两节点之间不存在连接边.

根据节点连接边存在性约束与节点连接边缺失性约束, 主要通过以下 6 个步骤构建大尺度约束模型:

a) 利用 MI 计算每个节点与其他节点之间的依赖程度大小;

b) 将 MI 的计算结果排序, 获取相应的 MMI, 找出与各节点依赖程度最高的相关节点. 由于 MI 的对称性, 即  $I(x; y) = I(y; x)$ , 每个节点的 MMI 所对应的连接边是无向的;

c) 随机给定连接边指向, 确定 BN 结构的节点顺序. 根据该节点序列, 剔除搜索空间中不满足该约束条件的候选结构, 从而缩小结构搜索空间;

d) 根据 c) 获得的节点序列建立相应的邻接矩阵, 该矩阵的行序对应节点序列. 由于节点序列中需要满足父节点排在子节点的前面, 因此只需考虑上三角邻接矩阵. 将矩阵的上三角部分中的所有元素设置为 1. 然后根据以下步骤中 CI 测试的结果, 修改满足条件独立的节点对所对应的矩阵元素;

e) 考虑到高阶 CI 测试的不可靠性且较高的计算复杂度<sup>[23]</sup>, 采用一阶 CI 测试, 即  $|Z| = 1$ , 利用式 (3) 计算邻接矩阵中所有节点的卡方统计量, 储存每一个节点与其他节点之间的 CI 测试结果中对应的最大  $p$ - 值;

f) 将显著性水平  $\alpha$  (初始值随机给定) 作为判断条件独立性的阈值. 若最大  $p$ - 值大于  $\alpha$ , 则两节点之间无连接边存在, 因此将上三角邻接矩阵中的元素 1 修改为 0. 在此基础上可对 c) 中获得的结构搜索空间完成进一步缩放.

2.2 小尺度约束模型

根据构建大尺度约束模型的步骤 f), 发现利用 CI 测试辨识节点独立性关系的精度与其显著性水平  $\alpha$  取值相关. 若  $\alpha$  取值过小, 易出现网络的部分连接边被永久删除, 极有可能丢失潜在最优解; 若  $\alpha$  取值过大, 结构搜索空间规模依然巨大, 且候选结构中可能存在较多错误连接边, 导致结构学习的效率和精度下降. 因此, 需要小尺度约束模型控制显著性水平  $\alpha$  的更新, 实现对结构搜索空间的动态缩放.

显著性水平  $\alpha$  的更新主要依赖结构复杂度评估函数对 GA 产生的每一代新的 BN 结构的复杂度评估. 根据评估结果产生更新后的  $\alpha_{new}$  完成结构搜索空间的进一步缩放, 更新后的结构搜索空间以及更新后的  $\alpha_{new}$  将作用于 GA 迭代搜索过程 (相关内容将在第 3 节进行介绍), 小尺度约束模型工作原理如图 2 所示.

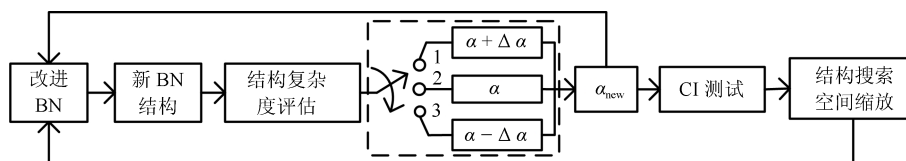


图 2 小尺度约束模型工作原理

Fig. 2 The working principle of small-scale constraint model

图 2 中的结构复杂度评估函数是在 BIC 评分函数的基础上获得的. 考虑到 BIC 评分函数是一种带惩罚函数的极大似然函数评分方法<sup>[23]</sup>, 评分函数中的第二项能够对 BN 结构复杂度进行评估, 因此可利用 BIC 评分函数中的第二项对  $\alpha$  进行动态更新. BIC 评分函数如式 (4) 所示, 即

$$\begin{aligned} score_{BIC} = & \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log_2 \frac{m_{ijk}}{m_{ij*}} - \\ & \sum_{i=1}^n \frac{q_i(r_i - 1)}{2} \log_2 m \end{aligned} \quad (4)$$

式中:  $n$  表示结构中的节点个数,  $q_i$  表示节点  $x_i$  的父节点集合的状态数,  $r_i$  表示节点  $x_i$  的状态数,  $m_{ijk}$  是观测数据中满足节点  $x_i$  的父节点集合为第  $j$  个状态且节点  $x_i$  为第  $k$  个状态的样本个数,  $m_{ij*} = \sum_{k=1}^{r_i} m_{ijk}$ ,  $m$  是观测数据样本总数. BIC 评分的第二项是描述 BN 结构复杂度的惩罚函数, 结构复杂度越高, 相应的惩罚函数值越大, 如式 (5) 所示:

$$penalty = \sum_{i=1}^n \frac{q_i(r_i - 1)}{2} \log_2 m \quad (5)$$

根据小尺度约束模型的工作原理, 构建小尺度约束模型主要通过以下 3 个步骤实现:

a) 利用式 (5) 计算当代种群中每个个体的结构复杂度  $penalty_k$ ;

b) 将上一步骤中计算得到的  $penalty_k$  与历史最优个体的结构复杂度  $penalty_{best}$  比较; 若  $penalty_k < penalty_{best}$ , 则选择输入 1; 若  $penalty_k > penalty_{best}$ , 则选择输入 3; 若  $penalty_k = penalty_{best}$ , 则选择输入 2. 将更新后的  $\alpha_{new}$  作为种群个体 CI 测试的新的显著性水平;

c) 利用更新后的显著性水平  $\alpha_{new}$  调整结构搜索空间.

需要说明的是, 步骤 b) 中根据复杂度评估结果更新  $\alpha$ , 即当种群个体的结构复杂度小于历史最优个体的结构复杂度时, 需要增加该个体结构的连接边, 减少 CI 测试时满足条件独立性关系的节点对的数量, 即提高显著性水平  $\alpha$ , 因此在  $\alpha$  更新过程中选择输入 1, 此时  $\alpha_{new} = \alpha + \Delta\alpha$ ; 当种群个体的结构复杂度大于历史最优个体的结构复杂度时, 应该减少结构的连接边数量, 即降低显著性水平  $\alpha$ , 因此在  $\alpha$  更新过程中选择输入 3, 此时  $\alpha_{new} = \alpha - \Delta\alpha$ ; 若种群个体和历史最优个体有相同的结构复杂度, 此时,  $\alpha$  的取值保持不变, 因此在  $\alpha$  更新过程中选择输入 2.

在 GA 迭代寻优过程中存在新个体适应度不及前代个体的情况. DSC-AL 算法通过联赛选择算子和精英保留机制将前代最优个体直接选入下一代中进行迭代寻优, 当代种群中的最优个体即为历史最优个体. 将当代种群中每一个个体与该最优个体的结构复杂度进行比较, 更新显著性水平. 此外, BIC 评分函数的罚项用于复杂度评估具有节点可分解性<sup>[8]</sup>, 以节点为单位将每个节点的结构复杂度用向量的方式保存. 在迭代寻优过程中通常只有少数节点的结构产生变化, 因此每次迭代只需更新个体表示的 BN 结构中少数节点的结构复杂度. 在动态调整显著性水平时直接利用结构复杂度向量中对应元素进行计算, 可以降低计算复杂性.

### 3 基于双尺度约束模型的 BN 结构自适应学习算法

#### 3.1 BN 结构自适应学习过程

基于双尺度约束模型的 BN 结构自适应学习算法设计思路如下:

1) 设计种群个体的编码方案. 将节点顺序和 CI 测试的显著性水平  $\alpha$  引入种群编码. 其目的主要包含两个方面: 避免迭代寻优过程中对新个体无环性的反复检验或修复无效结构; 利用双尺度约束模型限制结构搜索空间, 提高多节点复杂结构的学习精度和迭代寻优的收敛速度.

2) 动态调节变异概率. 在改进 GA 的进化过程中, 为保持种群多样性与结构继承性之间的平衡, 通过动态调节变异概率以提高结构学习的收敛速度和全局寻优能力.

基于双尺度约束模型的 BN 结构自适应学习算法流程如图 3 所示.

#### 3.2 编码生成

BN 结构  $G = (X, E)$  是一个有向无环图 (Directed acyclic graph, DAG). 令节点集合  $X = \{x_i\}_{i=1,2,\dots,n}$  代表领域变量, 节点间的有向边  $E = \{e_{ij}\}$  代表节点间的直接依赖关系,  $pa(x_i)$  表示节点  $x_i$  的父节点集合. 使用 GA 完成 BN 结构学习, 可使用邻接矩阵  $A = (e_{ij})$  对 BN 结构进行编码表达<sup>[24]</sup>. 然而随机生成的邻接矩阵, 其对应的结构无法保证 BN 无环要求. 为此设计一种新的编码方案以解决该问题.

新的个体编码方案的表达式如式 (6) 所示:

$$\begin{cases} C_i = O_i M_i S_i \\ O_i = x_1 x_2 \cdots x_n, x_k \in X, k = 1, 2, \dots, n \\ M_i = e_{12} e_{13} e_{23} \cdots e_{1n} e_{2n} \cdots e_{(n-1)n} \\ S_i = \alpha \end{cases} \quad (6)$$

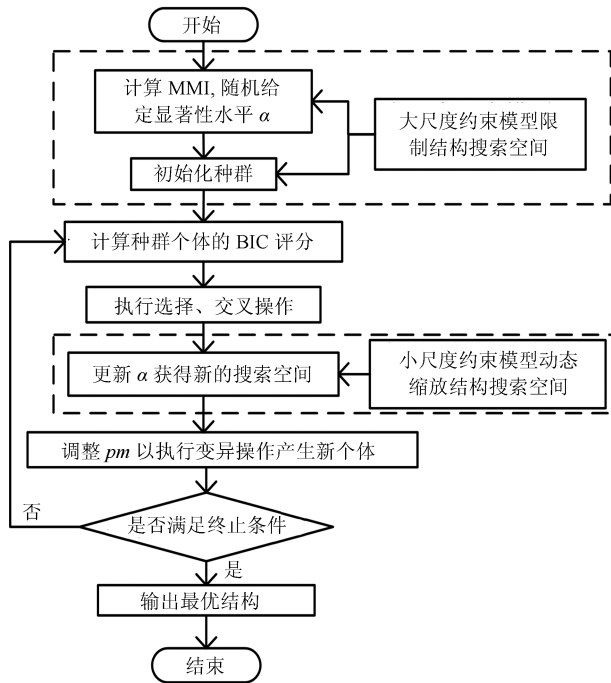


图 3 DSC-AL 算法流程图

Fig. 3 The flowchart of DSC-AL algorithm

$$e_{jk} = \begin{cases} 1, & \text{若 } x_j \text{ 是 } x_k \text{ 的父节点} \\ 0, & \text{其他} \end{cases} \quad (7)$$

式 (6) 中:  $C_i$  表示第  $i$  个种群个体,  $O_i$  代表 BN 结构的节点顺序,  $S_i$  代表 CI 测试的显著性水平,  $M_i$  表示根据第  $i$  个种群个体的节点序列构建的上三角邻接矩阵, 其中元素取值如式 (7) 所示. 在节点顺序中, 每一个节点的父节点都必须排在该节点的前面. 这一原则使得遗传算子是封闭的<sup>[25]</sup>, 即在进化过程中无需检测或修复种群个体, 能够保证产生的新个体是无环的. 此外, 将 CI 测试的显著性水平  $\alpha$  加入个体编码中, 使得显著性水平  $\alpha$  随迭代过程自适应更新.

为了说明搜索空间中所有的 BN 结构均能采用上述个体编码方案进行表达, 给出相关定理, 并进行证明.

**引理 1**<sup>[18]</sup>. 任何一个具有  $n + 1$  个节点的 DAG 都能够被分解为一个根节点 (或叶节点) 和一个具有  $n$  个节点的 DAG 的连接.

**定理 1.** 任何一个具有  $n$  个节点的 DAG 都能够使用一个确定的节点排序和与该序列对应的上三角邻接矩阵以及 CI 测试的显著性水平进行表示.

**证明.** 由 CI 测试的定义可知, 显著性水平大小决定两节点之间是否存在连接边, 并且该连接边是无向边, 因此并不影响节点排序. 当节点排序确定后, CI 测试的显著性水平只决定该序列对应的上三角邻接矩阵中的元素, 即条件独立的两节点对应的

矩阵元素必为 0. 因此, 只需选择适当的显著性水平保证条件独立的节点对集合为 DAG 中元素为 0 对应的节点对集合的子集.

下面用数学归纳法证明.

当  $n = 2$  时, 命题显然成立. 假设  $x_1$  是根节点,  $x_2$  是叶节点, 则存在节点排序为  $O = 12$ , 对应的上三角邻接矩阵  $M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ , 此时显著性水平  $\alpha$  的取值保证节点  $x_1$  与  $x_2$  不是条件独立;

假设命题对具有  $n$  个节点的 DAG 成立. 考虑  $G$  为任一具有  $n + 1$  个节点的 DAG. 根据引理 1,  $G$  可分解为一个根节点 (或叶节点) 和一个具有  $n$  个节点的 DAG 的连接. 若  $G$  分解为一个根节点和一个具有  $n$  个节点的 DAG  $G'$  的连接, 由假设可知,  $G'$  可由一个节点排序  $O$  及其对应的上三角邻接矩阵  $M$  和显著性水平  $\alpha$  表示, 其中

$$O_i = x_1 x_2 \cdots x_n, x_i \in \{1, 2, \dots, n + 1\} \quad (8)$$

$$M = \begin{bmatrix} 0 & e_{12} & e_{13} & \cdots & e_{1(n-1)} & e_{1n} \\ 0 & 0 & e_{23} & \cdots & e_{2(n-1)} & e_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & e_{(n-1)n} \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad (9)$$

根据显著性水平  $\alpha$  的取值, 条件独立的节点对所对应的  $e_{ij} = 0$ , 即在  $G'$  中该节点对之间必无连接边. 若  $G$  的根节点为  $x_{n+1}$ , 则  $G$  的节点排序可表示为  $O' = x_{n+1} x_1 x_2 \cdots x_n$ , 故其对应的上三角邻接矩阵为

$$M' = \begin{bmatrix} 0 & e_{(n+1)1} & e_{(n+1)2} & e_{(n+1)3} & \cdots & e_{(n+1)(n-1)} & e_{(n+1)n} \\ 0 & 0 & e_{12} & e_{13} & \cdots & e_{1(n-1)} & e_{1n} \\ 0 & 0 & 0 & e_{23} & \cdots & e_{2(n-1)} & e_{2n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & e_{(n-1)n} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & M'_{n+1} \\ 0 & M \end{bmatrix} \quad (10)$$

其中  $M'_{n+1}$  为  $1 \times n$  的矩阵, 在加入根节点后, 调整显著性水平  $\alpha$  的取值, 使得满足 CI 测试的条件独立关系的节点对集合为  $G$  中无连接边的节点对集合的子集.

同理, 若  $G$  分解为一个叶节点和一个具有  $n$  个节点的 DAG  $G''$  的连接, 则  $G''$  可由一个节点排序  $O$  及其对应的上三角邻接矩阵  $M$  和显著性水平  $\alpha$  表示, 如式 (8) 和式 (9) 所示. 若  $G$  的叶节点为  $x_{n+1}$ , 则  $G$  的节点排序可表示为  $O' = x_1 x_2 \cdots x_n x_{n+1}$ , 故其对应的上三角邻接矩阵

为

$$M' = \begin{bmatrix} 0 & e_{12} & e_{13} & e_{14} & \cdots & e_{1n} & e_{1(n+1)} \\ 0 & 0 & e_{23} & e_{24} & \cdots & e_{2n} & e_{2(n+1)} \\ 0 & 0 & 0 & e_{34} & \cdots & e_{3n} & e_{3(n+1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & e_{n(n+1)} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} = \begin{bmatrix} M & M''_{n+1} \\ 0 & 0 \end{bmatrix} \quad (11)$$

其中  $M''_{n+1}$  为  $n \times 1$  的矩阵, 在加入叶节点后, 调整显著性水平  $\alpha$  的取值, 使得满足 CI 测试的条件独立关系的节点对集合为  $G$  中无连接边的节点对集合的子集. 故命题对任一具有  $n+1$  个节点的 DAG 成立.  $\square$

### 3.3 自适应变异算子

GA 中变异操作的作用是保持种群的多样性. 传统的 GA 变异算子所使用的变异概率是一个固定值, 即 GA 每一代种群中的所有个体都使用相同的变异概率, 无法随 GA 迭代过程发生改变. 变异概率取值固定时, 若变异概率取值太小, 在 GA 迭代的中后期, 种群多样性将会急剧下降, GA 不易跳出局部极值点; 若变异概率取值过大, 极易破坏种群个体的优势基因, 从而导致 GA 的收敛速度缓慢. 为此, 设计一种基于种群个体适应度的自适应变异算子, 在 GA 迭代过程中为不同阶段的种群个体以及同一种群中的不同个体提供变异概率, 确保种群的多样性, 克服 GA 的早熟收敛.

种群个体的适应度能够反应变异概率变化情况. 当种群个体适应度逐渐趋于一致时, 变异概率应该适度增大, 以防止 GA 迭代陷入局部最优; 当种群个体的适应度较分散时, 变异概率应该适度减小, 使 GA 迭代尽快收敛到全局最优解<sup>[26]</sup>. 因此, 可通过种群的最大适应度  $fit_{max}$  和平均适应度  $fit_{avg}$  来描述种群适应度的一致性, 构建的变异概率  $p_m$  的自适应调节函数如式 (12) 所示.

$$p_m^k(g) = \begin{cases} p_m^k(g-1) \frac{fit_{max} - fit_k}{fit_{max} - fit_{avg}}, & fit_k \geq fit_{avg} \\ p'_m, & fit_k < fit_{avg} \end{cases} \quad (12)$$

式中:  $p_m^k(g)$  表示第  $g$  代种群中第  $k$  个个体的变异概率,  $p'_m$  为一常数,  $0 < p'_m < 1$ ,  $fit_k$  表示第  $k$  个个体的适应度值,  $fit_{max} - fit_{avg}$  的差值反映了种群适应度的集中情况,  $fit_k$  与  $fit_{avg}$  的比较则描述了第  $k$  个个体的优异程度.  $fit_k$  的计算公式如式 (13)

所示:

$$fit_k = \frac{score_k - score_{min}}{score_{max} - score_{min}} \quad (13)$$

式中:  $score_k$  是第  $k$  个个体所表示的 BN 结构的 BIC 评分,  $score_{min}$  是当前种群中的最低 BIC 评分,  $score_{max}$  是当前种群中的最高 BIC 评分.

在获得变异概率  $p_m$  的基础上, 若变异位置出现在表示上三角邻接矩阵的编码部分, 则该位置上的基因由 0 变为 1 或者由 1 变为 0; 若变异位置出现在节点顺序的编码部分, 则随机选择另一个节点与该位置上的节点互换. 需要注意, 随机选定的个体变异位置将不包含显著性水平  $\alpha$  的部分, 即个体中的显著性水平  $\alpha$  不参与变异.

### 3.4 其他算子

DSC-AL 算法的选择算子采用联赛选择方法从当前种群中挑选出优良个体作为父代执行交叉, 变异操作. 此外, 算法使用单点交叉算子, 具体实现过程如算法 1 所示, 其中关于交叉点位于种群个体编码的第一部分, 即节点顺序的情况, 图 4 给出了一个实例.

#### 算法 1. 交叉算子

输入. 种群个体  $ind$ , 交叉概率  $p_c$

输出. 交叉后的新个体  $new\_ind$

- 1) 交叉概率  $p_c \rightarrow$  交叉位置  $pos$ ;
- 2)  $pos$  出现在  $ind$  编码的第三部分, 即显著性水平  $\rightarrow$  不交叉;
- 3)  $pos$  出现在  $ind$  编码的第二部分, 即连接边  $\rightarrow$  从  $pos$  开始的后半段交叉;
- 4)  $pos$  出现在  $ind$  编码的第一部分, 即节点顺序  $\rightarrow$  二, 三部分直接互换, 第一部分按照图 4 交叉.

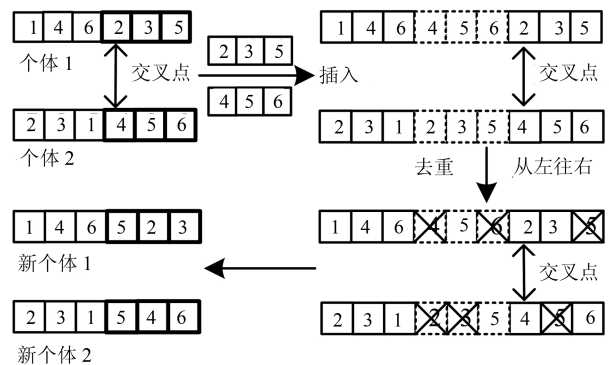


图 4 节点顺序交叉方法

Fig. 4 The crossover of node order

## 4 仿真实验

### 4.1 实验方案设计

#### 4.1.1 对比实验

1) 为验证 DSC-AL 算法中各功能模块的有效性, 选择的对比算法如下: 利用随机初始种群替代

大尺度约束模型的算法 (DSC-AL + RdInit), 采用固定显著性水平替代小尺度约束模型的算法 (DSC-AL + FixAlp), 使用固定变异概率的算法 (DSC-AL + FixP) 以及在进化过程中随机改变显著性水平的算法 (DSC-AL + RdAlp);

2) 为验证 DSC-AL 算法学习 BN 结构的性能, 采用下列算法进行对比实验: Lee 利用节点顺序和上三角邻接矩阵提出的基于双重 GA 编码的 BN 结构学习算法 (Dual genetic algorithm, DGA) 和 Gheisari 等构建的基于粒子群优化的 BN 结构学习方法 (BN construction algorithm using particle swarm optimization, BNC-PSO) 以及 K2 算法.

4.1.2 仿真模型及实验平台

1) 选用三种不同规模的标准 BN 模型 ASIA, CAR\_DIAGNOSIS2 和 ALARM 对结构学习方法进行测试, 验证 DSC-AL 算法对不同规模的 BN 结构进行学习的性能, 其中 ASIA 模型为 8 个节点的小规模网络, CAR\_DIAGNOSIS2 模型包含 18 个节点, ALARM 模型包含 37 个节点. 三种模型的拓扑结构如图 5 所示;

2) 实验平台为 Intel Core i5-5300U, 2.30 GHz

CPU, 32 位的微机, 操作系统为 Windows 7, 算法程序均在 MATLAB R2014a 上编译.

4.1.3 仿真实验参数设置

1) 由于 BN 中父节点集合的状态个数随父节点数量的增加呈指数级增长, 为简化算法实现过程, 在仿真实验中令每个节点的父节点个数不超过 5 个;

2) 由于 GA 和粒子群优化方法的迭代寻优过程具有很强的随机性, 为保证对比实验能够充分反映每种算法的性能, 除 K2 算法外, 其他迭代优化算法均独立仿真运行 30 次完成测试. K2 算法是一种确定性算法, 其学习结果与输入的节点顺序有关. 为保证公平性, 在不引入先验知识的情况下, 随机给出不同的节点序列作为 K2 算法的输入, 同样执行 30 次仿真实验;

3) DSC-AL 算法的参数设置如下: 种群规模为 200, 联赛选择算子中联赛规模取 4, 交叉概率为 0.9, 变异算子中常数  $p'_m = 0.1$ , 小尺度约束模型中  $\Delta\alpha = 0.02$ ;

4) DGA 的参数设置如下: 种群规模为 500, 交叉概率为 0.65, 变异概率为 0.05;

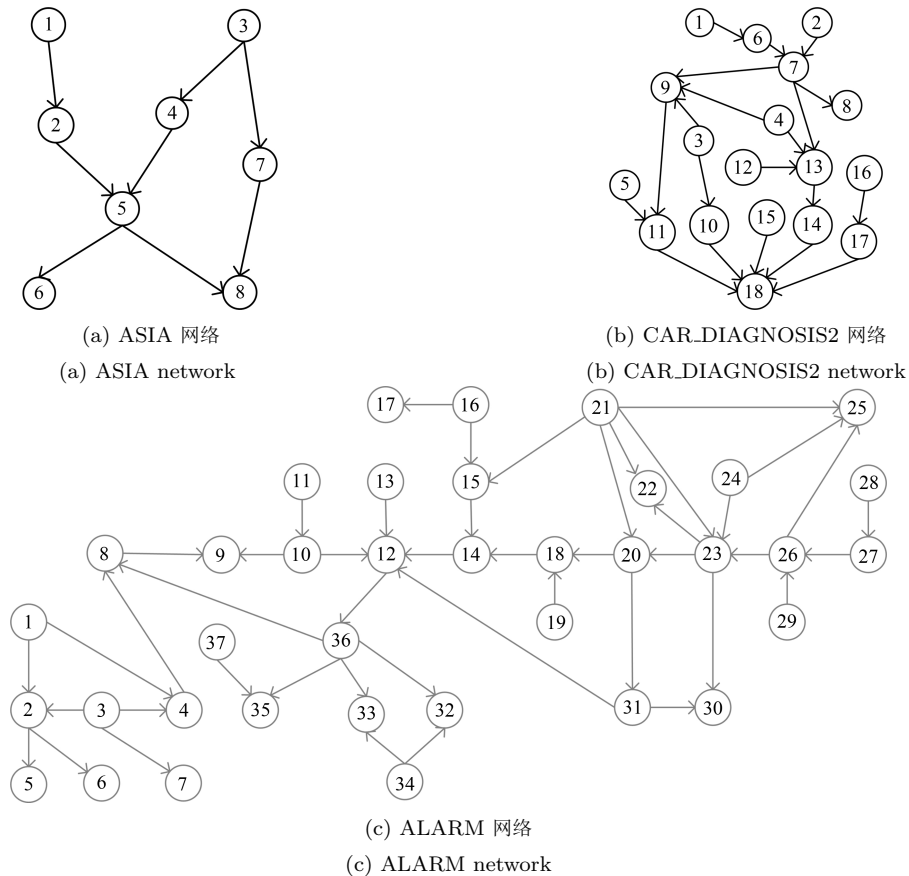


图 5 三种标准 BN 结构示意图

Fig. 5 Three benchmark BNs

5) BNC-PSO 的参数设置如下: 种群规模为 200, 其他参数与文献 [19] 取值相同;

6) DSC-AL + FixAlp 算法的参数设置如下: 显著性水平更新算子中  $\alpha = 0.05$ ;

7) 在 DSC-AL + FixP 算法中, 变异概率为 0.1;

8) 对比实验中, 每种算法学习中小型网络: ASIA 和 CAR\_DIAGNOSIS2 模型的最高迭代次数均为 300, 学习大规模网络: ALARM 模型的最高迭代次数均为 500; 均采用 BIC 评分准则对种群个体的适应度评估.

#### 4.1.4 算法性能评价指标

- 1) 初始种群中最优结构的 BIC 评分 (IBIC);
- 2) 算法获得的最终结构 BIC 评分 (BIC);
- 3) 算法获得的最终结构与标准模型结构之间的汉明距离 (SHD);
- 4) 算法运行时间 (RT);
- 5) 获得最优解的迭代次数 (BG).

#### 4.2 仿真实验与结果分析

实验 4.2.1 和 4.2.2 分别在 1000 组 ASIA 网络样本数据集 (ASIA-1000) 和 2000 组 CAR\_DIAGNOSIS2 网络样本数据集 (CAR\_DIAGNOSIS2-2000) 下对 DSC-AL 算法各功能模块的有效性进行仿真; 实验 4.2.3 分别在 2000 组和 5000 组 ALARM 模型样本数据集 (ALARM-2000 和 ALARM-5000) 下对 DSC-AL 算法的结构学习性能进行测试.

##### 4.2.1 ASIA 模型仿真实验

以 ASIA 模型为标准结构的仿真实验结果如表 1 所示. 表 1 给出了 ASIA-1000 样本数据集下, 7 种 BN 结构学习算法的 5 个性能指标平均结果, 即表中每种算法的性能指标值为 30 次重复实验的平均值, 括号中的数据为标准差. 在“数据集”列中, 括号中的数值为 ASIA 模型对应的 BIC 评分. 由于 K2 算法未使用 GA, 因此表 1 中 K2 对应的 IBIC, RT 和 BG 中无数据.

对表 1 仿真结果的分析如下:

1) 在 30 次的独立仿真实验中, DSC-AL 算法的平均结构汉明距离 (SHD) 为 1.3667, 在 7 种算法中最小. 结构汉明距离越小表示算法学习到的结构越接近标准网络结构, 即证明 DSC-AL 算法学习获得结构与标准 ASIA 模型结构相似度最高;

2) DSC-AL 算法的初始种群的平均 BIC 评分 (IBIC) 为 -2375.1, 比 DGA 算法 (-2406.9) 和 DSC-AL + RdInit 算法 (-2421.9) 的分值高, 与其他对比方法的 IBIC 分值接近. 证明利用 MMI 和 CI 测试构造的大尺度约束模型对 GA 种群进行初始化, 能够有效提高初始化种群的适应度;

3) 在相同的终止条件下, DSC-AL 算法的平均运行时间 (RT) 为 103.4270 秒, 比 DGA 算法 (173.9571 秒) 的用时短. 表明 DSC-AL 算法的学习效率优于 DGA 算法. 相较于基于 DSC-AL 算法的修改方法, DSC-AL 算法需要利用大尺度约束模型完成种群初始化操作, 且在迭代寻优过程中需要不断自适应调整 CI 测试的显著性水平和变异概率, 因此 DSC-AL 算法耗时较长;

4) DSC-AL 算法的平均结构汉明距离 (1.3667) 远远小于 K2 算法的平均结构汉明距离 (7.5667). 表明在无先验知识的情况下, DSC-AL 算法的学习精度远高于 K2 算法, 并且不需要依赖于专家经验. 为了清晰地对比 6 种结构学习算法 (除去 K2 算法) 的性能, 图 6 (a)~6 (c) 分别描述了在 ASIA-1000 样本数据下的平均汉明距离, 获得最优解的平均迭代次数和算法运行的平均消耗时间. 由图 6 可知, 与 DGA 算法相比, DSC-AL 算法能够使用较少的迭代次数获得最优的 BN 结构; 并且相较于 DSC-AL + FixP, DSC-AL + RdAlp 和 DSC-AL + RdInit 算法, 虽然 DSC-AL 算法在相同终止条件下所耗费的时间较多, 但是该算法学习到的最终结构明显优于其他算法, 获得最优解的平均迭代次数也少于其他算法. 但是在小规模 ASIA 模型的仿真结果中, 与 DSC-AL + FixAlp 算法比较, DSC-AL 算法的优势并不明显.

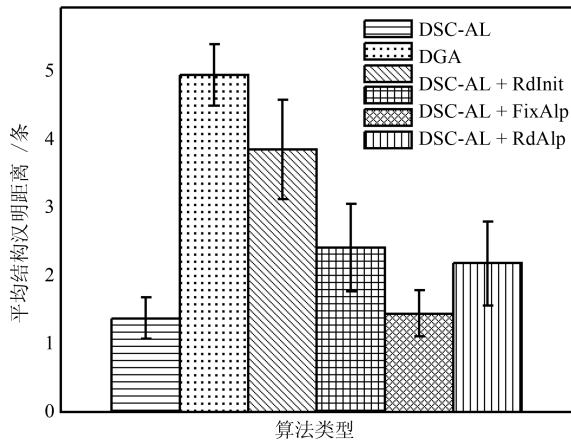
图 7 是在 ASIA-1000 样本数据下分别运行 30 次, 6 种算法 (除去 K2 算法) 获得最优结构的 BIC

表 1 ASIA 模型下不同算法结果对比

Table 1 Comparisons of different methods on ASIA network

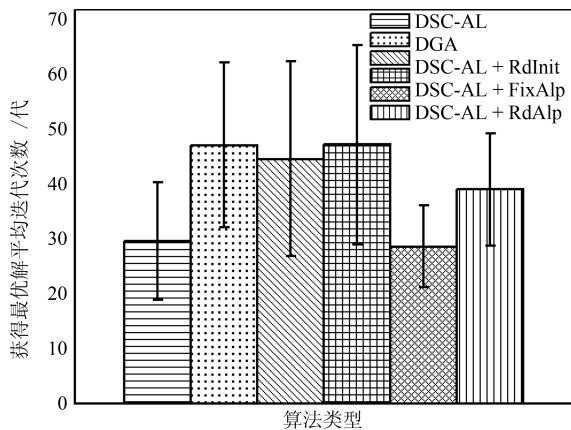
数据集	算法	IBIC	BIC	SHD	RT	BG
ASIA-1 000 (-2 325.3)	DSC-AL	-2 375.1 (3.6570)	-2 320.5 (2.1782)	1.3667 (0.7184)	103.4270 (17.5317)	29.6667 (25.1812)
	DGA	-2 406.9 (15.1353)	-2 329.5 (6.8571)	4.9333 (1.2576)	173.9571 (7.9109)	47.2333 (42.1775)
	K2	/	-2 342.1 (14.0940)	7.5667 (2.1284)	/	/
	DSC-AL + RdInit	-2 421.9 (19.5248)	-2 324.7 (4.7155)	3.8333 (2.0186)	104.3722 (23.3174)	44.7000 (49.7165)
	DSC-AL + FixAlp	-2 372.3 (0.2821)	-2 320.2 (1.7524)	1.4333 (0.9353)	62.6387 (9.6306)	28.6333 (20.8450)
	DSC-AL + RdAlp	-2 374.4 (2.7308)	-2 321.7 (3.3730)	2.1667 (1.7237)	85.2060 (7.6515)	39.1000 (28.7250)
	DSC-AL + FixP	-2 374.8 (2.9988)	-2 322.0 (3.2387)	2.4000 (1.7927)	68.6206 (12.4026)	47.2667 (51.1205)





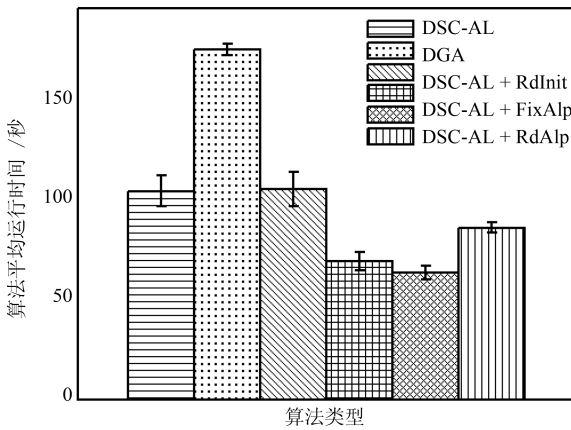
(a) 平均结构汉明距离

(a) Average structural Hamming distance



(b) 获得最优解平均迭代次数

(b) Average generation to obtain optimal solution



(c) 算法平均运行时间

(c) Average running time

图6 6种算法在ASIA-1000数据集下的3种性能指标的误差条形图

Fig.6 Error bar graph of 3 measures for 6 algorithms on ASIA-1000 data set

评分平均值随迭代次数变化的曲线. 由图7可知, DSC-AL算法初始种群中的最优结构评分相较于于

DGA算法高, 直至迭代终止, DSC-AL算法最终结构的BIC评分值始终高于DGA算法, 并且DSC-AL算法收敛速度比DGA算法快. 与其他修改算法对比, DSC-AL算法的BIC评分的平均值在迭代过程中始终高于DSC-AL + RdInit算法, 且DSC-AL算法的收敛速度高于DSC-AL + FixP算法和DSC-AL + RdAlp算法, 但是与DSC-AL + FixAlp算法的趋势及大小无明显差异.

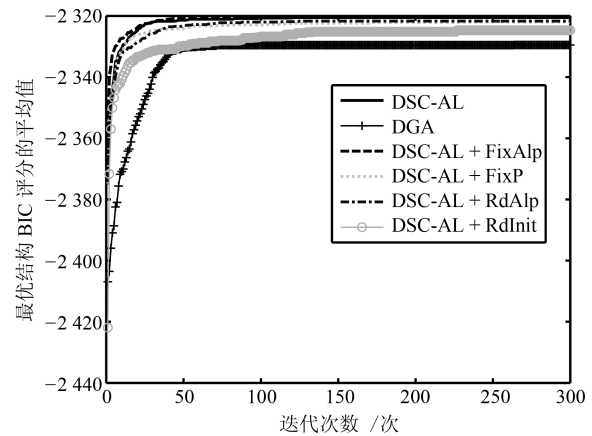


图7 ASIA-1000下最优结构BIC评分平均值变化曲线  
Fig.7 The curves of BIC scores for optimal structures on ASIA-1000 data set

图8是在ASIA-1000样本数据下分别运行30次, 6种算法(除去K2算法)获得的优于上一代种群的个体数平均值. 由图8可知, 每种算法在迭代寻优过程的后期所获得的比上一代更优的结构个数都趋于零. 然而即使将DGA算法的种群规模放宽至500(远远大于DSC-AL算法的种群规模), 该算法出现优于上一代种群的个体数平均值为零的情况的时间更早, 即更容易陷入局部最优, 发生早熟收敛.

#### 4.2.2 CAR\_DIAGNOSIS2模型仿真实验

以CAR\_DIAGNOSIS2模型为标准结构的仿真实验结果如表2所示. 表2给出了CAR\_DIAGNOSIS2-2000样本数据集下, 7种BN结构学习算法的5个性能指标平均结果, 即表中每种算法的性能指标值为30次重复实验的平均值, 括号中的数据为标准差. 在“数据集”列中, 括号中的数值为CAR\_DIAGNOSIS2模型对应的BIC评分. 由于K2算法未使用GA, 因此表2中K2对应的IBIC, RT和BG中无数据.

对表2仿真结果的分析如下:

1) 在30次的独立仿真实验中, DSC-AL算法的平均结构汉明距离(SHD)为6.8000, 该指标在7种算法中最小, 表明该算法学习获得的结构与标准CAR\_DIAGNOSIS2模型相似度最高;

表 2 CAR\_DIAGNOSIS2 模型下不同算法结果对比

Table 2 Comparisons of different methods on CAR\_DIAGNOSIS2 network

数据集	算法	IBIC	BIC	SHD	RT	BG
CAR_DIAGNOSIS2-2000 (-11 922)	DSC-AL	-13 865 (186.3612)	-11 774 (43.2254)	6.8000 (1.1861)	520.6599 (74.8401)	144.0667 (45.2601)
	DGA	-15 546 (271.5482)	-11 795 (51.1551)	13.2000 (1.7301)	856.7351 (85.2662)	222.7667 (21.2630)
	K2	/	-12 111 (198.0365)	23.5667 (5.4752)	/	/
	DSC-AL + RdInit	-15 661 (415.5809)	-12 034 (181.5865)	13.8333 (3.0181)	508.8949 (67.7425)	194.5000 (67.9111)
	DSC-AL + FixAlp	-13 557 (87.5065)	-11 745 (22.6139)	10.7000 (3.0867)	583.9935 (9.6306)	226.6667 (33.6988)
	DSC-AL + RdAlp	-13 883 (177.8057)	-11 820 (37.1534)	9.9000 (2.0060)	426.4885 (63.1594)	172.5667 (57.0485)
	DSC-AL + FixP	-13 860 (143.4086)	-11 825 (41.7158)	9.8667 (2.2242)	364.3424 (90.1956)	159.2667 (42.1303)

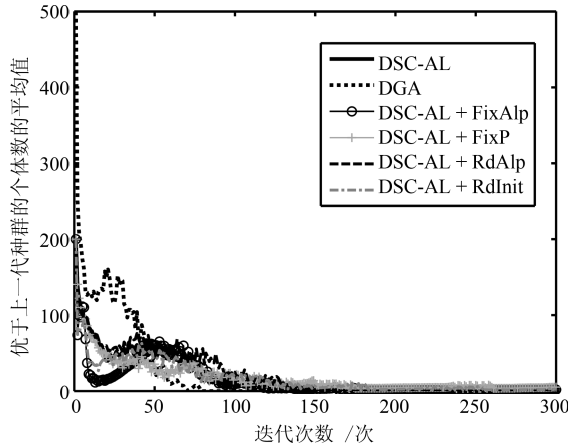


图 8 ASIA-1000 下优于上一代种群的个体数平均值变化曲线

Fig. 8 The curves of number of better individuals on ASIA-1000 data set

2) 在相同的终止条件下, DSC-AL 算法的平均运行时间 (RT) 为 520.6599 秒, 比 DGA 算法 (856.7351 秒) 的用时短, 表明 DSC-AL 算法的学习效率比 DGA 算法高。

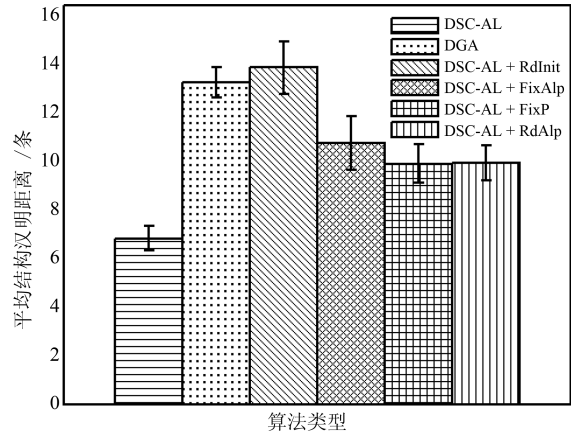
图 9 (a)~9 (c) 分别描述了上述 6 种算法 (除去 K2 算法) 在 CAR\_DIAGNOSIS2-2000 样本数据下的平均汉明距离, 获得最优解的平均迭代次数和算法运行的平均消耗时间。如图 9 所示, 可以获得与 ASIA 模型仿真结果所显示的相同结论。此外, 通过比较 CAR\_DIAGNOSIS2 模型和 ASIA 模型的仿真实验结果, 发现随着网络规模的增大, DSC-AL 算法的优势表现得更加明显。

图 10 是在 CAR\_DIAGNOSIS2-2000 样本数据下分别运行 30 次, 6 种算法 (除去 K2 算法) 获得最优结构的 BIC 评分平均值随迭代次数变化的曲线。由图 10 可知, 与 ASIA-1000 样本数据下的仿真结果相似, 与 DGA 算法相比, 从初始种群开始直至算法结束, DSC-AL 算法获得的每一代的最优结构性能更好, 并且 DSC-AL 算法收敛速度更快。此外, DSC-AL + RdInit 算法和 DGA 算法的初始种群中最优结构评分低于其他算法, 即上述两种算法的初始种群所表示的结构最差。

图 11 是在 CAR\_DIAGNOSIS2-2000 样本数据下分别运行 30 次, 6 种算法 (除去 K2 算法) 获得的优于上一代种群的个体数平均值。如图 11 所示, 由于 DGA 算法种群个体数为 500, 在迭代寻优过程的前期和中期, 该算法获得的优于上一代种群的个体数的平均值最高, 但是到迭代的后期, 与 ASIA 模型仿真结果相似, DGA 算法仍然最先出现比上一代更优结构个数为零的情况。

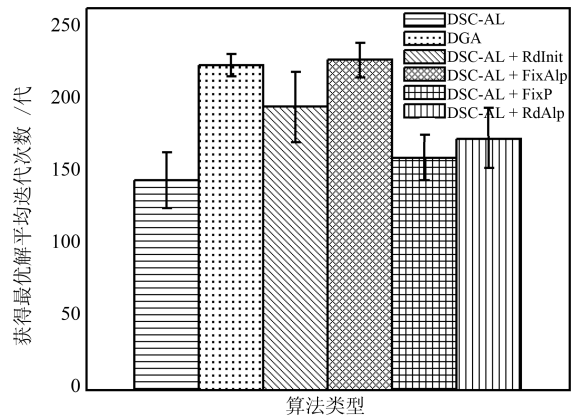
### 4.2.3 ALARM 模型仿真实验

前面两小节的实验结果已经验证 DSC-AL 算法各功能模块的有效性, 为了进一步探究 DSC-AL 算



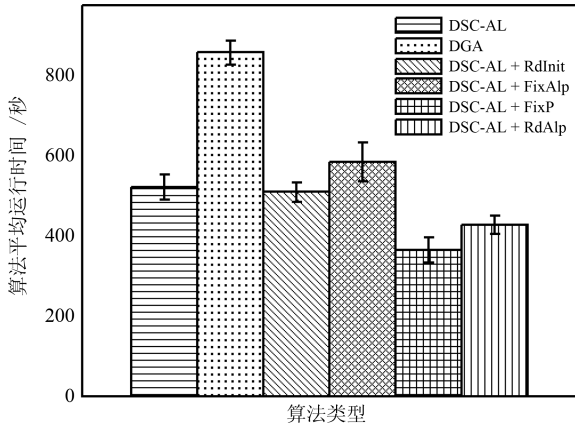
(a) 平均结构汉明距离

(a) Average structural Hamming distance



(b) 获得最优解平均迭代次数

(b) Average generation to obtain optimal solution



(c) 算法平均运行时间  
(c) Average running time

图 9 6 种算法在 CAR\_DIAGNOSIS2-2000 数据集下的 3 种性能指标的误差条形图

Fig. 9 Error bar graph of 3 measures for 6 algorithms on CAR\_DIAGNOSIS2-2000 data set

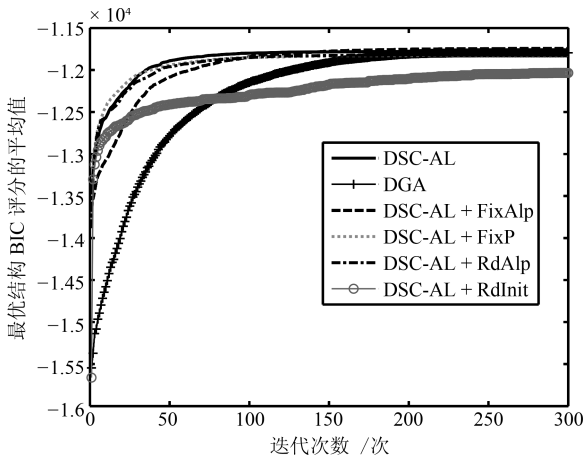


图 10 CAR\_DIAGNOSIS2-2000 下最优结构 BIC 评分平均值变化曲线

Fig. 10 The curves of BIC scores for optimal structures on CAR\_DIAGNOSIS2-2000 data set

法的结构学习性能,以 ALARM 模型为标准结构进行仿真实验,对比三种群智能优化算法,包括 DSC-

AL, BNC-PSO 和 DGA, 结果如表 3 所示. 表 3 分别给出了 ALARM-2000 和 ALARM-5000 样本数据集下, 3 种结构学习算法的 3 个性能指标平均结果, 即表中每种算法的性能指标值为 30 次重复实验的平均值, 括号中的数据为标准差. 在“数据集”列中, 括号中的数值为 ALARM 模型不同数据集对应的 BIC 评分.

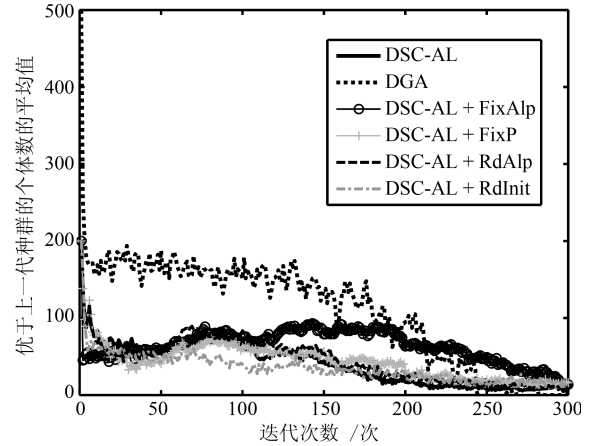


图 11 CAR\_DIAGNOSIS2-2000 下优于上一代种群的个体数平均值变化曲线

Fig. 11 The curves of number of better individuals on CAR\_DIAGNOSIS2-2000 data set

对表 3 仿真结果的分析如下:

1) 在 ALARM-2000 和 ALARM-5000 两个不同规模的数据集下, DSC-AL 算法的平均结构汉明距离 (SHD) 在 3 种基于群智能优化的 BN 结构学习算法中均最小, 表明 DSC-AL 算法学习获得结构与标准 AL-ARM 模型结构相似度最高, 并且受到数据规模的影响较小;

2) 在相同的终止条件下, 针对 ALARM-2000 和 ALARM-5000 两个不同数据集规模, 结合 3 种学习方法获得最优解的平均迭代次数 (GB):  $225.8000 < 267.7778 < 498.1667, 203.4000 <$

表 3 ALARM 模型下不同算法结果对比

Table 3 Comparisons of different methods on ALARM network

数据集	算法	SHD	RT	BG
ALARM-2000 (-20 294)	DSC-AL	15.1000 (2.7669)	2 898.8 (267.3125)	225.8000 (95.5671)
	DGA	33.5000 (3.5071)	2 910.5 (122.4261)	498.1667 (1.4720)
	BNC-PSO	25.3333 (5.5000)	2 689.1 (153.1974)	267.7778 (63.5227)
ALARM-5000 (-48 724)	DSC-AL	13.5000 (0.9718)	2 322.7 (106.2002)	203.4000 (85.6364)
	DGA	28.6667 (1.2111)	2 435.5 (239.3540)	498.3333 (3.1411)
	BNC-PSO	16.3000 (3.6833)	1 616.3 (473.0926)	315.9000 (98.0583)

度和迭代寻优的收敛速度提高. 与 DGA 和 BN315.9000 < 498.3333) 和平均结构汉明距离 (SHD) 的仿真结果可知, DSC-AL 算法能够使用最少的迭代次数获得与标准模型最相似的结构, 表明在不同规模数据集下, DSC-AL 算法的收敛速度和 BN 结构复原能力均优于 DGA 和 BNC-PSO 算法.

图 12 和图 13 是 3 种基于群智能方法的 BN 结构学习算法分别在 ALARM-2000 和 ALARM-5000 样本数据下运行 30 次获得最优结构的 BIC 评分平均值随迭代次数变化的曲线. 由图 12 和图 13 可知, 在不同数据样本规模下, 3 种算法评分平均值随迭代次数的变化趋势是相似的. DSC-AL 算法初始种群中的最优结构评分相较于 DGA 和 BNC-PSO 算法高, 直至迭代终止, DSC-AL 算法学习到的结构评分值始终明显高于 DGA 算法, 但与 BNC-PSO 算法获得的 BN 结构的评分值差异小; 此外, DSC-AL 算法收敛速度比 BNC-PSO 算法快, DGA 算法在迭代终止时并没有达到收敛状态.

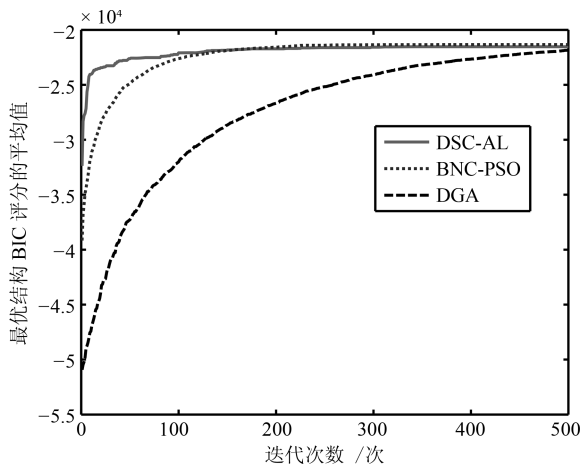


图 12 ALARM-2000 下最优结构 BIC 评分平均值变化曲线

Fig. 12 The curves of BIC scores for optimal structures on ALARM-2000 data set

仿真实验结果说明, 在无先验知识的情况下, 本文提出的 DSC-AL 算法通过采用大小尺度约束模型和变异概率自适应调节函数, 使得结构学习的精 C-PSO 两种基于评分搜索的算法不同, DSC-AL 算法是一种混合学习方法, 融合条件独立关系检验和智能优化方法. 为缩小结构搜索空间, DSC-AL 算法采用一阶 CI 测试和 MMI 计算结果结合, 假设  $n$  为变量数, 在最坏的情况下算法需要进行  $O(n^3)$  次 CI 测试和  $O(n^2)$  次 MI 计算. 虽然与 DGA 和 BNC-PSO 算法比较, DSC-AL 算法在进行迭代寻优之前增加了大尺度约束模型过滤无效结构的步骤, 但是在后续的评价搜索阶段, DSC-AL 算法利用节

点顺序设计编码方案, 使得该算法无需进行无环检验, 并且通过小尺度约束模型动态更新结构搜索空间, 进一步剔除无效结构, 提高迭代寻优的搜索效率. 相较于 DGA 和 BNC-PSO 算法, DSC-AL 算法完成 BN 结构学习任务的性能更优, 并且受到数据样本规模的影响小.

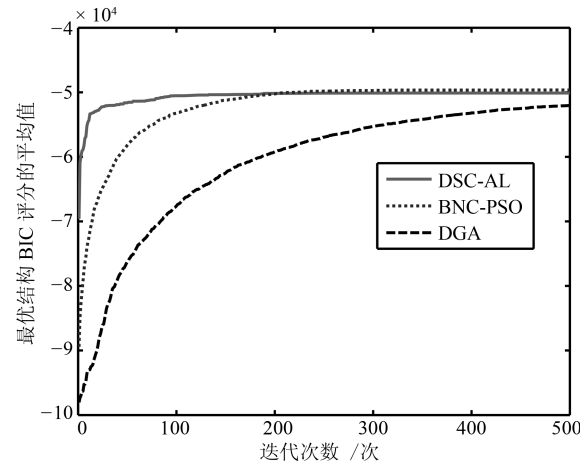


图 13 ALARM-5000 下最优结构 BIC 评分平均值变化曲线

Fig. 13 The curves of BIC scores for optimal structures on ALARM-5000 data set

## 5 结论

从大规模候选结构集合中快速、准确地获得一个与数据样本匹配程度最优的结构模型是 BN 结构学习亟待解决的问题之一. 为此, 提出了一种基于双尺度约束模型的 BN 结构自适应学习 (DSC-AL) 算法, 解决了在无先验知识的情况下, 通过数据样本信息动态限制搜索空间, 高效挖掘节点间相关性的 BN 结构学习问题. 该算法在 MMI 和 CI 测试的双重约束条件下产生 GA 的初始种群, 获得 BIC 评分高的种群个体, 加快收敛速度; 通过在个体编码中引入 CI 测试的显著性水平, DSC-AL 算法在进化过程中利用更新模型不断调整显著性水平的大小, 使得对应的搜索空间动态, 合理地变化, 提高算法的全局搜索能力; 同时充分利用种群个体的适应度分布自适应地调节变异概率, 克服算法的早熟收敛. 仿真结果验证了本文算法对解决 BN 结构学习问题的有效性, 该算法能够确保种群个体的多样性, 提高结构学习的精度和迭代寻优的收敛速度.

## References

- 1 Mohammadfam I, Ghasemi F, Kalatpour O, Moghimbeigi A. Constructing a Bayesian network model for improving safety behavior of employees at workplaces. *Applied Ergonomics*, 2017, **58**: 35–47

- 2 Zarei E, Azadeh A, Khakzad N, Aliabadi M M, Mohammadfam I. Dynamic safety assessment of natural gas stations using Bayesian network. *Journal of Hazardous Materials*, 2017, **321**: 830–840
- 3 Landis W G, Ayre K K, Johns A F, Summers H M, Stinson J, Harris M J, et al. The multiple stressor ecological risk assessment for the mercury-contaminated South River and upper Shenandoah River using the Bayesian network-relative risk model. *Integrated Environmental Assessment and Management*, 2017, **13**(1): 85–99
- 4 Wang Jing-Yun, Liu San-Yang, Zhu Ming-Min. Structure learning of chain graphs using the conditional independence tests. *Acta Electronica Sinica*, 2017, **45**(10): 2443–2448  
(王静云, 刘三阳, 朱明敏. 基于条件独立测试的链图结构学习算法. 电子学报, 2017, **45**(10): 2443–2448)
- 5 Madsen A L, Jensen F, Salmerón A, Langseth H, Nielsen T D. A parallel algorithm for Bayesian network structure learning from large data sets. *Knowledge-Based Systems*, 2017, **117**: 46–55
- 6 Villanueva E, Maciel C D. Efficient methods for learning Bayesian network super-structures. *Neurocomputing*, 2014, **123**: 3–12
- 7 Di Ruo-Hai, Gao Xiao-Guang, Guo Zhi-Gao. The modeling method with Bayesian networks and its application in the threat assessment under small data sets. *Acta Electronica Sinica*, 2016, **44**(6): 1504–1511  
(邸若海, 高晓光, 郭志高. 小数据集 BN 建模方法及其在威胁评估中的应用. 电子学报, 2016, **44**(6): 1504–1511)
- 8 Di Ruo-Hai, Gao Xiao-Guang, Guo Zhi-Gao. Bayesian networks structure learning based on improved BIC scoring. *System Engineering and Electronics*, 2017, **39**(2): 437–444  
(邸若海, 高晓光, 郭志高. 基于改进 BIC 评分的贝叶斯网络结构学习. 系统工程与电子技术, 2017, **39**(2): 437–444)
- 9 Adabor E S, Acquah-Mensah G K, Oduro F T. SAGA: a hybrid search algorithm for Bayesian network structure learning of transcriptional regulatory networks. *Journal of Biomedical Informatics*, 2015, **53**: 27–35
- 10 Masegosa A R, Moral S. An interactive approach for Bayesian network learning using domain/expert knowledge. *International Journal of Approximate Reasoning*, 2013, **54**(8): 1168–1181
- 11 Gao Xiao-Guang, Ye Si-Mao, Di Ruo-Hai, Kou Zhen-Chao. Bayesian network structures learning based on approach using incorporate priors method. *System Engineering and Electronics*, 2018, **40**(4): 790–796  
(高晓光, 叶思懋, 邸若海, 寇振超. 基于融合先验方法的贝叶斯网络结构学习. 系统工程与电子技术, 2018, **40**(4): 790–796)
- 12 Gasse M, Aussem A, Elghazel H. A hybrid algorithm for Bayesian network structure learning with application to multi-label learning. *Expert Systems with Applications*, 2014, **41**(15): 6755–6772
- 13 Li Ming, Zhang Ren, Hong Mei, Bai Cheng-Zu. Improved structure learning algorithm of Bayesian network based on information flow. *System Engineering and Electronics*, 2018, **40**(6): 1385–1390  
(李明, 张韧, 洪梅, 白成祖. 基于信息流改进的贝叶斯网络结构学习算法. 系统工程与电子技术, 2018, **40**(6): 1385–1390)
- 14 Liu Bin, Wang Hai-Yu, Sun Mei-Ting, Liu Hao-Ran, Liu Yong-Ji, Zhang Chun-Lan. Learning Bayesian network structure from node ordering searching optimal. *Journal of Electronics and Information Technology*, 2018, **40**(5): 1234–1241  
(刘彬, 王海羽, 孙美婷, 刘浩然, 刘永记, 张春兰. 一种通过节点序寻优进行贝叶斯网络结构学习的算法. 电子与信息学报, 2018, **40**(5): 1234–1241)
- 15 Wong M L, Leung K S. An efficient data mining method for learning Bayesian networks using an evolutionary algorithm-based hybrid approach. *IEEE Transactions on Evolutionary Computation*, 2004, **8**(4): 378–404
- 16 Ji Jun-Zhong, Zhang Hong-Xun, Hu Ren-Bing, Liu Chun-Nian. A Bayesian network learning algorithm based on independence test and ant colony optimization. *Acta Automatica Sinica*, 2009, **35**(3): 281–288  
(冀俊忠, 张鸿勋, 胡仁兵, 刘椿年. 一种基于独立性测试和蚁群优化的贝叶斯网学习算法. 自动化学报, 2009, **35**(3): 281–288)
- 17 Li B H, Liu S Y, Li Z G. Improved algorithm based on mutual information for learning Bayesian network structures in the space of equivalence classes. *Multimedia Tools and Applications*, 2012, **60**(1): 129–137
- 18 Lee J, Chung W, Kim E. Structure learning of Bayesian networks using dual genetic algorithm. *IEICE Transactions on Information and Systems*, 2008, **91**(1): 32–43
- 19 Gheisari S, Meybodi M R. BNC-PSO: Structure learning of Bayesian networks by particle swarm optimization. *Information Sciences*, 2016, **348**: 272–289
- 20 Cooper G F, Herskovits E. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 1992, **9**(4): 309–347
- 21 Robinson, R W. Counting unlabeled acyclic digraphs. In *Proceedings of the 5th Australian Conference on Combinatorial Mathematics*, Melbourne, Australia: Springer, 1976. 28–43
- 22 de Campos L M, Castellano J G. Bayesian network learning algorithms using structural restrictions. *International Journal of Approximate Reasoning*, 2007, **45**(2): 233–254
- 23 Liu Jian-Wei, Li Hai-En, Luo Xiong-Lin. Learning technique of probabilistic graphical models: a review. *Acta Automatica Sinica*, 2014, **40**(6): 1025–1044  
(刘建伟, 黎海恩, 罗雄麟. 概率图模型学习技术研究进展. 自动化学报, 2014, **40**(6): 1025–1044)
- 24 Wang Chun-Feng, Zhang Yong-Hong. Bayesian network structure learning based on unconstrained optimization and genetic algorithm. *Control and Decision*, 2013, **28**(4): 618–622  
(汪春峰, 张永红. 基于无约束优化和遗传算法的贝叶斯网络结构学习方法. 控制与决策, 2013, **28**(4): 618–622)
- 25 Larrañaga P, Karshenas H, Bielza C, et al. A review on evolutionary algorithms in Bayesian network learning and inference tasks. *Information Sciences*, 2013, **233**: 109–125
- 26 Omara F A, Arafa M M. Genetic algorithms for task scheduling problem. *Journal of Parallel and Distributed Computing*, 2010, **70**(1): 13–22



**戴晶帼** 海南大学信息科学技术学院博士研究生. 主要研究方向为贝叶斯网络, 智能优化.

E-mail: djgolivia\_edu@126.com

(**DAI Jing-Guo** Ph. D. candidate at the College of Infomation and Technology, Hainan University. Her research interest covers Bayesian network and

intelligent optimization.)



**任 佳** 海南大学信息科学技术学院教授. 主要研究方向为智能控制, 机器学习. 本文通信作者.

E-mail: renjia@hainu.edu.cn

(**REN Jia** Professor at the College of Infomation and Technology, Hainan University. His research interest covers intelligent control and machine learning. Corresponding author of this paper.)



**董 超** 国家海洋局南海调查技术中心副研究员. 主要研究方向为智能控制.

E-mail: dongchaoxj888@126.com

(**DONG Chao** Associate professor at South China Sea Marine Engineering surveying Center of State Oceanic Administration. His main research interest is intelligent control.)



**杜文才** 中国澳门城市大学数据科学研究院教授, 海南大学信息科学技术学院教授. 主要研究方向为数据挖掘, 物联网技术. E-mail: wencai@hainu.edu.cn

(**DU Wen-Cai** Professor at the Institute of Data Science, City University of Macau, China and the College of Infomation and Technology, Hainan University. His research interest covers data mining and internet of things.)