

基于 MMPC-FPSO 贝叶斯网络混合结构学习方法

董文佳¹, 方洋旺^{1,2}, 彭维仕³, 闫晓斌¹

(1. 西安邮电大学计算机学院, 西安, 710121; 2. 西北工业大学无人系统技术研究院, 西安, 710072;

3. 武警工程大学装备管理与保障学院, 西安, 710086)

摘要 针对贝叶斯网络结构学习的过程中网络结构规模随节点数增加呈指数增长, 导致网络结构搜索空间增大, 进而导致网络结构学习算法效率低下的问题, 提出一种基于最大最小父子集合约束与萤火虫粒子群搜索算法的贝叶斯网络混合结构学习方法。首先, 针对粒子群算法在解决贝叶斯网络结构学习过程中, 随机初始化网络结构种群导致算法搜索效率低下, 网络结构准确性低的问题提出一种基于改进的最大最小父子集合算法的种群约束方法。其次, 针对传统的基于粒子群评分搜索方法速度慢, 精度低, 易陷入局部最优的问题, 提出一种基于萤火虫算子的粒子寻优策略。最后, 为了验证所提方法的正确性和优越性, 将上述方法用于3种标准网络的结构学习。仿真结果表明: 所提算法与传统的基于粒子群的结构学习方法相比, 所得的贝叶斯信息准则评分与标准网络评分的差距分别缩小了68.7%、65.5%、34.1%。

关键词 结构学习; 贝叶斯网络; 粒子群算法; MMPC 算法

DOI 10.3969/j.issn.2097-1915.2024.02.009

中图分类号 TP18 **文献标志码** A **文章编号** 2097-1915(2024)02-0076-09

A Hybrid Structure Learning Method Based on MMPC-FPSO for Bayesian Networks

DONG Wenjia¹, FANG Yangwang^{1,2}, PENG Weishi³, YAN Xiaobin¹

(1. School of Computer Science and Technology, Xi'an University of Posts & Telecommunications, Xi'an 710121, China;

2. Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China;

3. School of Equipment Management and Support, People Armed Police Engineering University, Xi'an 710086, China)

Abstract Aimed at the problems that in the process of bayesian network structure learning, the network structure size increases in exponential with the number of nodes, in leading to the expansion of the network structure search space, and, in turn, hampering the efficiency of network structure learning algorithms, a bayesian network hybrid structure learning method, MMPC-FPSO, is introduced in combination with maximum-minimum parent-child set constraints (MMPC) and firefly particle swarm optimization (FPSO). Firstly, in view of addressing the issues of low algorithm efficiency and inaccurate network structure due to random initialization of the network structure population in the process of bayesian network structure learning using particle swarm algorithms, a population constraint method is proposed based on the improved MMPC algorithm. Secondly, in view of tackling the problems of slow speed, low accuracy, and susceptibility to local optima in traditional particle swarm-based scoring search methods, a particle optimization strategy based on the firefly algorithm is presented. Finally, in order to validate the correct-

收稿日期: 2023-07-01

基金项目: 国家自然科学基金(61973253)

作者简介: 董文佳(1997-), 男, 河北衡水人, 硕士生, 研究方向为贝叶斯网络。E-mail: 18730864465@163.com

引用格式: 董文佳, 方洋旺, 彭维仕, 等. 基于 MMPC-FPSO 贝叶斯网络混合结构学习方法[J]. 空军工程大学学报, 2024, 25(2): 76-84.
DONG Wenjia, FANG Yangwang, PENG Weishi, et al. A Hybrid Structure Learning Method Based on MMPC-FPSO for Bayesian Networks [J]. Journal of Air Force Engineering University, 2024, 25(2): 76-84.

ness and superiority of the proposed method, the three standard networks are applied to the structure learning. The simulation results demonstrate that the gap between the obtained BIC scores and the scores of standard networks is reduced by 68.7%, 65.5%, 34.1%, respectively by the proposed algorithm, compared to the traditional particle swarm-based structure learning methods.

Key words structure learning; Bayesian networks; particle swarm optimization; MMPC algorithm

贝叶斯网络^[1] (bayesian networks, BN) 具有强大的推理能力和可解释性, 被广泛应用于各个领域。而网络结构的构建是建立贝叶斯网络模型的关键环节。贝叶斯网络结构学习方法主要分为 3 类: ① 基于约束的结构学习方法^[2], 该类方法会因某个个体的检测错误, 导致整个网络结构构建出现错误。② 基于评分的结构学习方法^[3], 该类方法由于搜索空间会随着节点数增加呈指数上升, 因此存在收敛速度慢、精度低, 甚至无法得出有效解的问题。③ 基于混合的结构学习方法^[4], 该类方法虽然结合了约束和评分的优点, 但是仍然存在易陷入局部最优、搜索效率低的问题。

由于寻找最优网络结构是一个非确定多项式 (non-deterministic polynomial hard, NP-hard) 问题。因此, 群智能算法被广泛应用于贝叶斯网络结构学习中。出现了基于布谷鸟算法^[5]、粒子群算法^[6]、细菌觅食算法^[7]、鲸鱼算法^[8]、狼群算法^[9]、多头绒泡菌算法^[10]、蚁狮算法^[11] 等多种基于群智能的贝叶斯网络结构学习方法。其中粒子群算法因其参数少、结构简单、收敛速度快、全局寻优能力强等优点被广泛应用于贝叶斯网络结构学习中。

文献[12]设计了一种离散粒子群算法来解决 BN 结构的学习问题。利用变异算子和最近邻搜索算子克服算法早熟的缺点。文献[13]将遗传算法与粒子群算法相结合重新定义粒子更新规则, 在小数据集上取得了很好的效果。以上两种方法并没有对搜索空间进行约束, 导致初始化粒子的随机性较大, 算法效率较低。文献[14]通过计算节点之间的互信息对粒子种群初始化进行约束, 缩小寻优空间, 但是互信息需要对任意两个节点进行计算, 因此当数据维度高时, 模型构建效率低。文献[15]使用最大权重生成树限制初始空间, 并引入混沌搜索提高了算法的效率。因为初始种群构建过程中使用了树结构, 对根节点的依赖太大, 当根节点发生错误时, 会导致所有网络结构失效, 从而影响算法的准确性。文献[16]利用 PC (Peter and Clark) 算法对初始搜索空间进行约束, 虽然提高了收敛速度和精度, 但是交叉变异均采用随机操作, 导致收敛效率有待进一步提高。

为解决粒子群算法在学习贝叶斯网络结构的过程中, 候选网络个数随节点数增加呈指数增长, 导致

网络结构搜索空间增大, 进而导致算法搜索速度慢、学习精度低、易陷入局部最优的问题, 本文提出基于最大最小父子集合约束与萤火虫粒子群搜索算法 (max-min parent-child and firefly particle swarm optimization, MMPC-FPSO) 的贝叶斯网络混合结构学习方法。首先分析了粒子群算法在解决贝叶斯网络结构学习过程中的基本流程及优缺点, 然后提出了基于 MMPC-FPSO 的贝叶斯网络混合结构学习方法, 最后通过仿真实验验证算法的可行性和有效性。

1 贝叶斯网络结构学习方法分析

1.1 贝叶斯网络

贝叶斯网络使用二元组 (G, θ) 表示, 其中 G 是一个有向无环图 (DAG) 表示为 $G = (X, E)$ 。 X 表示网络中的节点集合, E 表示网络中有向边的集合, θ 表示网络中每个节点的条件概率分布集合, $\pi(X_i)$ 表示节点 X_i 的父节点集合, 根据贝叶斯网络条件独立的特性, 联合概率分布 $P(X_1, X_2, \dots, X_n)$ 可以写成每个随机变量条件概率的乘积:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \pi(X_i)) \quad (1)$$

若没有父节点, 则 $\pi(X_i) = \Phi$ 。

准确的网络结构决定了贝叶网络推断的有效性, 因此高效寻找准确的贝叶斯网络结构学习方法决定了贝叶斯网络模型构建的质量。

1.2 基于粒子群算法的结构学习方法分析

基于粒子群算法的贝叶斯网络结构学习算法主要包括种群初始化、计算粒子评分值、粒子位置更新、参数自适应调节 4 部分。

1.2.1 种群初始化

贝叶斯网络结构类似于数据结构中图的数据类型, 因此采用文献[12]的邻接矩阵编码方式对网络中节点和边进行储存。 n 个节点的贝叶斯网络需要定义一个 $n \times n$ 维的邻接矩阵, 表示形式如下:

$$\mathbf{X}_i = \begin{Bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{Bmatrix} \quad (2)$$

$$x_{ij} = \begin{cases} 1, & \text{节点 } i \text{ 是节点 } j \text{ 的父节点} \\ 0, & \text{其它} \end{cases} \quad (3)$$

式(2)中,邻接矩阵对角线的值均为0。为方便粒子进行交叉变异操作,将邻接矩阵重写为: $[x_{11}, x_{12}, \dots, x_{1n}, x_{21}, \dots, x_{2n}, \dots, x_{n1}, \dots, x_{nm}]$ 。图1表示候选网络结构转变为行向量的过程。

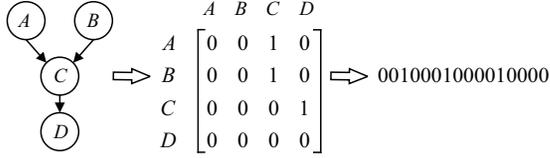


图1 网络拓扑结构的编码表示

通过随机初始化邻接矩阵,将其转变为行向量,由此得到初始化粒子种群。

1.2.2 计算粒子评分值

计算粒子评分值是指通过贝叶斯网络评分函数计算每个粒子的适应度值,粒子适应度值越大,粒子所代表的网络结构越准确。由于贝叶斯信息准则(bayesian information criterion, BIC)评分函数^[17]不需要参数先验知识,并且计算简单方便。因此采用 BIC 评分函数作为适应度值函数,即:

$$\log P(B_G | D) =$$

$$\sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \frac{m_{ijk}}{m_{ij*}} - \frac{1}{2} \sum_{i=1}^n q_i (r_i - 1) \log m \quad (4)$$

$$m_{ij*} = \sum_{k=1}^{r_i} m_{ijk} \quad (5)$$

$$\sum_{k=1}^{r_i} \frac{m_{ijk}}{m_{ij*}} = 1 \quad (6)$$

式中: B_G 代表与样本集合 D 具有一定关联的网络结构; n 为节点个数; q_i 为节点 X_i 的父节点集合取值的组合个数(若 X_i 无父节点,则 $q_i = 1$); r_i 表示节点 X_i 共有 r_i 个取值; m 为样本总数量。 m_{ijk} 是样本 D 中满足 $X_i = k, \pi(X_i) = j$ 的样本个数。

1.2.3 粒子位置更新

原有的粒子群位置更新公式仅适用于连续域空间,因此针对贝叶斯网络结构的离散空间重新定义粒子位置更新公式如下:

$$N_i^{t+1} = W_1(X_i^t, \omega) = \begin{cases} M(X_i^t), & r_1 < \omega \\ X_i^t, & \text{others} \end{cases} \quad (7)$$

$$S_i^{t+1} = W_2(N_i^{t+1}, c_1) = \begin{cases} C_p(N_i^{t+1}), & r_2 < c_1 \\ N_i^{t+1}, & \text{others} \end{cases} \quad (8)$$

$$X_i^{t+1} = W_3(S_i^{t+1}, c_2) = \begin{cases} C_g(S_i^{t+1}), & r_3 < c_2 \\ S_i^{t+1}, & \text{others} \end{cases} \quad (9)$$

式中: W_1 表示变异操作; ω 表示突变概率,变异过程中随机选择粒子的一条边进行删除操作; W_2 、 W_3 分别表示与个体最优值和全局最优值进行交叉操作;

c_1 、 c_2 表示交叉概率,交叉操作产生的新粒子由2个粒子之间的公共部分和2个粒子的随机部分组成。

由于贝叶斯网络是一个有向无环图,每次更改粒子位置后都需要保证它的无环性,本文中提到的所有去环操作均采用深度优先搜索遍历。首先,通过深度优先搜索方法获取邻接矩阵中环路上的节点。其次,随机删除环路上某个连接边,遍历所有节点,直到不存在环路为止得到有向无环图。

1.2.4 参数自适应调节

粒子在寻优的过程中,前期需要大范围的全局寻找较优位置,后期需要快速收敛到种群最优值位置。因此在前期需要较大的 ω 、 c_1 以及较小的 c_2 , 保证粒子的全局寻优能力,后期需要较小的 ω 、 c_1 以及较大的 c_2 , 保证粒子快速收敛。参数更新公式如下:

$$\omega = \omega_{\text{start}} - \frac{W_{\text{start}} - W_{\text{end}}}{\text{Max_It}} t \quad (10)$$

$$c_1 = c_{1_{\text{start}}} - \frac{c_{1_{\text{start}}} - c_{1_{\text{end}}}}{\text{Max_It}} t \quad (11)$$

$$c_2 = c_{2_{\text{start}}} - \frac{c_{2_{\text{start}}} - c_{2_{\text{end}}}}{\text{Max_It}} t \quad (12)$$

式中: t 表示当前迭代次数; Max_It 表示最大的迭代次数。

2 基于 MMPC-FPSO 的贝叶斯网络混合结构学习方法

首先,使用基于改进 MMPC 的种群约束方法构造骨干网络框架,在此基础上通过随机增边、反转生成初始种群,从而解决了基于评分搜索初始种群阶段随机性较大的问题。其次,在网络结构搜索阶段使用一种基于萤火虫算子的粒子寻优策略提高算法的收敛速度。

2.1 基于改进 MMPC 的种群约束方法

对于存在 n 个节点的贝叶斯网络而言,其所有可能的网络结构个数计算如下:

$$\begin{cases} f(1) = 1, & n = 1 \\ f(n) = \sum_{i=1}^n (-1)^{i+1} \frac{n!}{(n-1)! i!} f(n-i), & n > 1 \end{cases} \quad (13)$$

式中: n 为节点个数。可见,候选网络结构个数随着节点 n 的增长呈指数增加。因此,在整个指数级的搜索空间中随机初始化种群策略会导致空间覆盖率低,进而导致算法收敛速度慢、精度低的问题。

目前对搜索空间进行约束的方法有互信息^[18-19](mutual information, MI)、最大权生成树^[20]

(maximum weight spanning tree, MWST)、PC 算法^[21], MI 算法和 MWST 算法容易造成初始错误边的传播,影响网络结构精度,并且时间开销较大。PC 算法的准确性依赖于独立性检测的节点顺序。本文使用的 MMPC^[22]算法是一种局部检测独立性算法,大大降低了计算复杂度,并且该算法通过双层独立性检测确保了结果的精确性。由于原始的 MMPC 算法的最终结果是部分有向无环图,因此,本文在 MMPC 算法基础上进行改进,得到有向无环图,并将其作为骨干网络框架进行种群初始化。

使用本文提出的基于改进 MMPC 的种群约束方法,初始化种群主要分为以下步骤:

步骤 1 通过 MMPC 算法得到每个粒子的最大最小父子节点集合(CPC)。

步骤 2 计算每一个节点与其对应的 CPC 节点集合中每个元素的互信息(MI),确定每个节点与其对应的 CPC 节点集合中具有最大互信息的两条连接边。互信息计算公式如下:

$$I(X_i, X_j) = \sum_{x_i, x_j} p(x_i, x_j) \log \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \quad (14)$$

式中: $I(X_i, X_j)$ 表示节点之间的互信息,值越大,则表示节点间两个节点间的关联性越强。并且互信息值具有对称的性质, $I(X_i, X_j) = I(X_j, X_i)$,所以两个节点间的连接边是无向边,通过步骤 2 可以得到一个含有局部信息的无向图。

步骤 3 将无向图转换为有向图。利用条件相对平均熵计算不同方向的数值,通过比较大小确定无向边的方向。计算公式如下:

$$R(X_j \rightarrow X_i) = \frac{H(X_i | X_j)}{H(X_i) \times |X_i|} \quad (15)$$

$$H(X_i) = - \sum_{x_i=1}^n p(x_i) \log p(x_i) \quad (16)$$

$$H(X_i | X_j) = \sum_{i=1}^N \sum_{j=1}^M p(x_i, x_j) \log p(x_i | x_j) \quad (17)$$

式中: $|X_i|$ 是随机变量 X_i 的取值个数; $R(X_j \rightarrow X_i)$ 表示节点 X_j 指向节点 X_i 的可能性; $H(X_i)$ 表示随机变量 X_i 的熵; $H(X_i | X_j)$ 表示随机变量 X_i 的条件熵。如果 $R(X_i \rightarrow X_j) \geq R(X_j \rightarrow X_i)$,表示节点 X_i 指向节点 X_j 。

步骤 4 初始化种群。通过加边和反转两种随机操作,增加初始化种群的多样性。

加边:通过随机指定邻接矩阵 $x_{ij} = 0 (i \neq j)$ 的节点对,为其增加有向边,令 $x_{ij} = 1 (i \neq j)$ 。反转:随机选取已经含有连接边 $x_{ij} = 1 (i \neq j)$ 的节点对,

反转其方向,令 $x_{ji} = 1, x_{ij} = 0 (i \neq j)$ 。在初始化种群后,针对每个粒子检验其是否属于有向无环图,如果存在环路,采用深度遍历对其进行去环操作。

2.2 基于 FPSO 的搜索方法

为解决传统基于粒子群的贝叶斯网络结构学习算法收敛速度慢、易陷入局部最优的问题,提出基于萤火虫算子的粒子寻优策略。

2.2.1 汉明统一变异算子

本文引入一种基于汉明距离的变异算子,称为汉明统一变异算子(hamming uniform mutation by addition and deletion, H-UMAD)来执行变异操作。相较于传统的变异算子,该算子主要用于表示两个粒子间不同数值的位数和,通过判断汉明距离的大小,确定两个粒子间的差异程度。计算汉明距离公式如下:

$$r_{ij} = |X_i - X_j| = \sum_{d=1}^{n \times n} |(X_{id} - X_{jd})| \quad (18)$$

$$R_{r_{ij}} = \frac{r_{ij}}{n^2} \quad (19)$$

式中: r_{ij} 表示粒子 X_i 和粒子 X_j 之间不相等的位数和; $R_{r_{ij}}$ 表示数值不同的位置在行向量中所占的比例。

该变异方法输入参数是粒子本身的基因向量和该粒子个体最优值的基因向量。在变异过程中,通过随机产生概率值与变异概率作比较,来确定是否变异操作。如进行变异操作,再与 $R_{r_{ij}}$ 进行比较,确定变异程度。算法流程如图 2 所示。

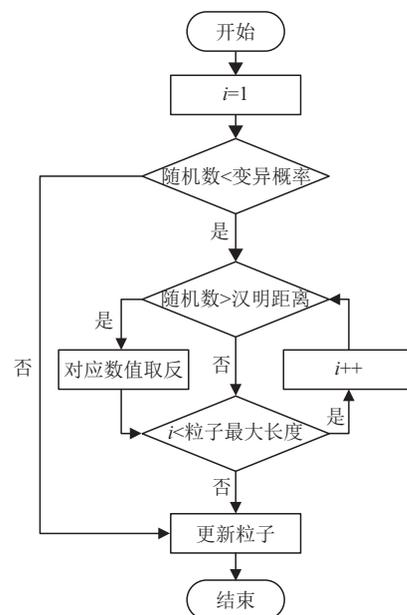


图 2 变异算子的流程图

2.2.2 基于萤火虫的交叉算子

由于传统的粒子群算法不适用于贝叶斯网络结构寻优,出现了基于遗传编码的粒子寻优策略。但是

该方法的交叉策略在与个体最优或者全局最优位置进行交叉时,不能很好地继承优秀基因,导致收敛速度较慢。因此,本文在传统基于粒子群的寻优策略中引入萤火虫算子^[23],该方法不仅能保留部分传统粒子群算法位置更新公式,还能有效地加快收敛速度。

如果粒子 X_j 的评分高于粒子 X_i 的评分,则粒子 X_i 与粒子 X_j 交叉操作的位置更新公式如下:

$$X_i^{\text{new}} = X_i + R_{\text{rand}} \otimes \beta \otimes (X_j - X_i) \quad (20)$$

式中: $R_{\text{rand}} = \{r_1, r_2, \dots, r_{n^2}\}$, 其中每一个元素 r_k 代表一个介于 0 和 1 之间的随机数。 $(X_j - X_i) = \{X_j(1) - X_i(1), \dots, X_j(n^2) - X_i(n^2)\}$, 其中每一个元素的值是 0, 1, -1 的其中一个。0 代表粒子 X_j 和粒子 X_i 在该位置有相同的网络结构, 粒子 X_i 倾向于不变。1 代表粒子 X_j 在该位置有边, 粒子 X_i 在该位置没有边, 粒子 X_i 倾向于添加这条边。-1 代表粒子 X_j 在该位置没有边, 粒子 X_i 在该位置有边, 粒子 X_i 倾向于删除这条边。因此公式(20)分解如下:

$$R_{\text{rand}} \otimes \beta \otimes (X_j - X_i) = \begin{cases} X_j(k) - X_i(k), & r_k \leq \beta \\ 0, & r_k > \beta \end{cases} \quad (21)$$

$$k = 1, 2, \dots, n^2$$

式中:吸引力 β 定义如下:

$$\beta = \frac{1}{1 + \beta_0 e^{-r_{ij}}} \quad (22)$$

式中: γ 为光吸收系数;吸引力 β 本质是一个 0 到 1 之间的数,在当前粒子与个体最优粒子或者全局最优粒子相差较大时,也就是汉明距离 r_{ij} 较大时,则需要交叉操作,这就要求吸引力 β 随着汉明距离的增大而增大。

2.2.3 粒子局部优化

为了提高粒子的开发能力获得更高质量的网络结构,本文结合 HC 算法^[24]中的弧操作搜索策略,利用增删改 3 种操作对粒子进行修改。在粒子完成交叉变异操作后,分别对粒子执行增删改操作。随机选择一个没有连接边的节点对,对其加边。删除和反转行为随机选择一对有边的节点对,删除该边或者修改边的方向,得到 3 个有效的粒子,通过计算评分值,选出评分值最高的粒子进入下一代。

2.3 算法的执行步骤

本文提出了一种基于 MMPC-FPSO 贝叶斯网络混合结构学习方法,简称 M-FPSO 算法。该算法执行步骤如下:

步骤 1 初始化参数:种群规模、变异和交叉概率、光吸收系数等。

步骤 2 通过改进的 MMPC 算法得到骨干网

络框架,通过随机增改操作构建初始粒子种群。

步骤 3 通过计算每个粒子的 BIC 评分,更新个体最优值和全局最优值。

步骤 4 执行变异和交叉操作,更新粒子位置,并对粒子进行局部优化。

步骤 5 计算每个粒子的 BIC 评分,更新个体最优值和全局最优值。

步骤 6 检查是否满足终止条件,如果满足则输出当前结果,否则,返回步骤 4。

算法流程如图 3 所示:

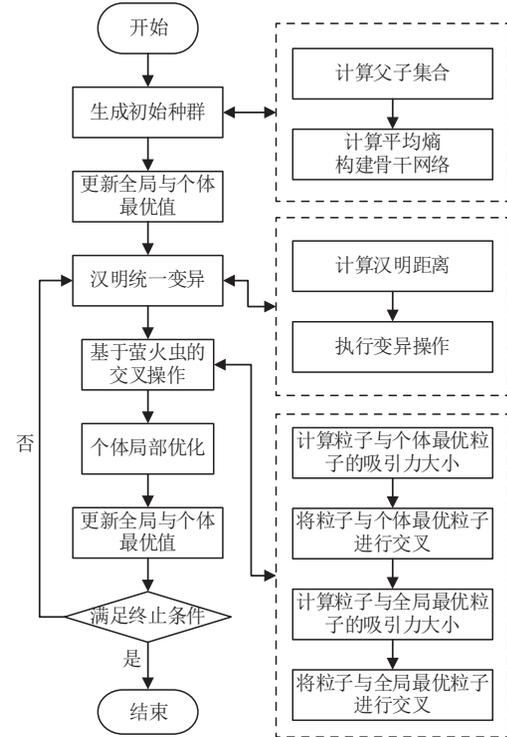


图 3 算法整体流程图

对于改变粒子位置的每一步操作,都需要保证该粒子所代表的网络结构是有向无环图。因此,在粒子每一次改变位置后,都需要对粒子进行去环处理,保证粒子的有效性。

3 实例分析

3.1 数据集

本文仿真实验的硬件环境是内存 8G, CPU 为 Intel(R) Core(TM), i5-11300H 3.10 GHz。实验平台为 Windows10, 程序采用 matlab R2021a 进行编写, 工具箱采用 FullBNT-1.0.7。为了测试本文提出的 M-FPSO 算法的性能, 使用 3 种标准网络 CANCER、ALARM 和 ASIA。本文使用 BNT 工具箱从标准网络 CANCER、ASIA 和 ALARM 网络中进行采样, 分别构建样本数量为 500、1 000、3 000、5 000 的样本数据集。数据集的具体构成如表 1 所示。

表 1 训练数据集的构成

样本类型	节点个数	连接边数	最大入/出度	取值范围
CANCER-500	5	5	2	2
CANCER-1 000	5	5	2	2
CANCER-3 000	5	5	2	2
CANCER-5 000	5	5	2	2
ASIA-500	8	8	2	2
ASIA-1 000	8	8	2	2
ASIA-3 000	8	8	2	2
ASIA-5 000	8	8	2	2
ALARM-500	37	46	4/5	2~4
ALARM-1 000	37	46	4/5	2~4
ALARM-3 000	37	46	4/5	2~4
ALARM-5 000	37	46	4/5	2~4

3.2 参数设置及评价指标

为验证本文提出算法的可行性,使用以上 3 种数据集对算法进行实验分析。为验证该算法的优越性,分别将其与基于遗传的结构学习算法(bayesian network construction algorithm using GA,BNC-GA)^[25]、BNC-PSO^[13]算法、最大最小蚁群算法(max-min ACO,MM-ACO)^[26]、最大最小爬山算法(max-min hill-climbing,MM-HC)^[27]进行实验仿真时,需要对不同的算法进行参数设置,BNC-GA 算法的种群规模为 50,交叉概率因子 $p_c=0.9$,变异概率因子 $p_m=0.1$; BNC-PSO 算法的种群规模为 50,惯性权重 $\omega_{start}=0.9$, $\omega_{end}=0.35$,学习因子 c_1 从 0.84 下降到 0.52,学习因子 c_2 从 0.38 上升到 0.81;MM-ACO 算法种群规模为 10,信息挥发因子 $\rho=0.4$,权重参数 $\alpha=\beta=2$; MMHC 算法使用 BNT 工具箱中的默认参数。

为了评估算法的性能,本文采用 2 种性能评价指标分别是 BIC 评分和结构汉明距离(SHD),使用以上 2 种指标对不同的算法分别评估学习之后的网络结构与标准网络结构之间的差异程度。

1)BIC 评分,当学习到的网络结构越接近于标准网络结构时,BIC 评分会越高。

2)结构汉明距离(SHD),汉明距离定义为 $H(G)=M(G)+A(G)+I(G)$ 分别代表与标准网络结构相比,缺失边个数、多余边个数和反转边个数之和,通常一个优秀的搜索算法可以从数据集中学习到尽可能与原始网络相同的结构,即 SHD 指标越小表明算法的准确性越高。

3.3 实验结果分析

3.3.1 M-FPSO 算法性能分析

为了验证本文算法改进策略的有效性,基于 3.1 节中 CANCER、ASIA、ALARM 3 种不同的网络结构数据集,首先为验证基于改进 MMPC 的种群约束方法的有效性,设置初始化种群个数为 100,计算种群平均 BIC 得分,将基于改进 MMPC 的种

群约束方法分别与随机初始化、基于互信息(MI)、基于最大权生成树(MWST)、基于 PC 算法的种群约束方法进行比较,在 3 种不同数据集上的评分数据如表 2~4 所示,数据对比柱形图如图 4~6 所示。

表 2 基于 CANCER 数据集种群 BIC 值对比

数据集	随机初始化	MI	MWST	PC	改进 MMPC
500	-965.8	-932	-895.4	-903.7	-871.07
1 000	-1 971.6	-1 798.1	-1 800.7	-1 789.2	-1 749.5
3 000	-5 745.6	-5 623.2	-5 399.7	-5 142.2	-4 875.6
5 000	-9 823.4	-9 021.1	-8 765.8	-8 691.3	-8 528.7

表 3 基于 ASIA 数据集种群 BIC 值对比

数据集	随机初始化	MI	MWST	PC	改进 MMPC
500	-1 437.1	-1 373.1	-1 370.1	-1 308.9	-1 305.5
1 000	-2 765.8	-2 548.4	-2 555.7	-2 538.1	-2 486.9
3 000	-7 775.3	-7 281.1	-7 205.4	-7 349.8	-7 178.2
5 000	-13 080.9	-12 300.5	-11 919.2	-12 328.7	-11 666.2

表 4 基于 ALARM 数据集种群 BIC 值对比

数据集	随机初始化	MI	MWST	PC	改进 MMPC
500	-13 209.7	-13 136	-7 884.9	-8 340.9	-7 788.1
1 000	-22 337.9	-22 273.6	-14 169.4	-14 766.3	-13 236.8
3 000	-57 736.3	-57 708.6	-39 488.6	-37 707.2	-34 163.1
5 000	-93 543.9	-93 020.8	-65 331.2	-55 055.2	-53 471.1

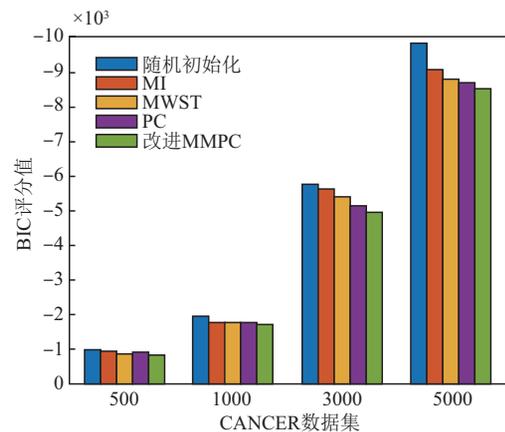


图 4 基于 CANCER 数据集对比图

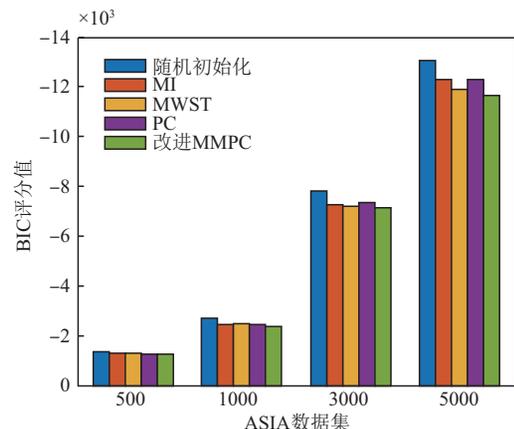


图 5 基于 ASIA 数据集对比图

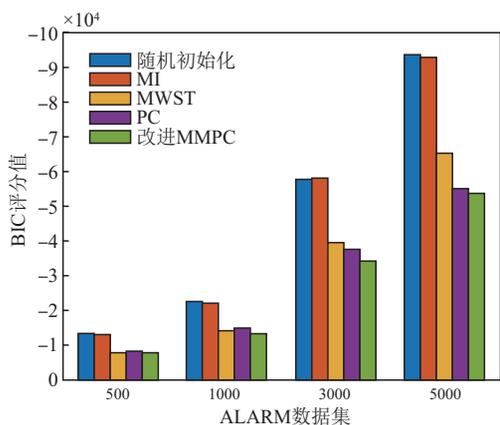


图6 基于ALARM数据集对比图

表2、表3、表4分别给出了基于CANCER、ASIA、ALARM 3种网络数据集下, BIC评分在不同种群初始化算法中的结果。由数据可以看出, 传统随机初始化种群方法得到的BIC评分值均小于其他方法。而本文提出的基于改进MMPC的种群约束方法得到的BIC评分值, 在3种不同数据集上均大于基于互信息(MI)、基于最大权生成树(MWST)、基于PC算法的种群约束方法得到的BIC评分值。正如在图4、图5、图6中本文的改进方法的柱形图始终最短, 因此可见其种群质量相对较高, 大大缩减了种群搜索空间, 提高了算法的收敛精度和速度。

为验证本文改进后算法的收敛性能, 与未改进的传统基于粒子群的贝叶斯网络结构学习方法(BNC-PSO)进行对比, 在ASIA-1000的数据集上两种算法分别运行50次, 求取平均值, 构建收敛曲线如图7所示。

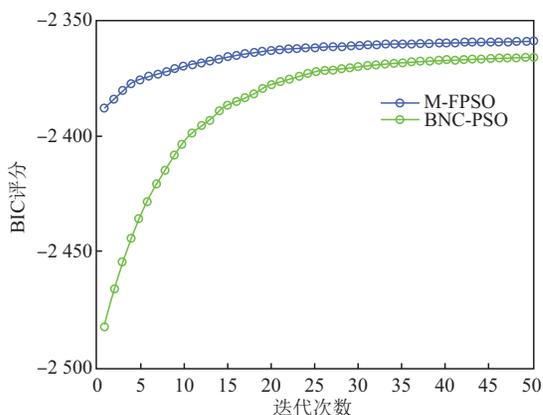


图7 算法改进前后的收敛曲线

由图7可知, 本文M-FPSO算法的收敛曲线起始阶段远高于传统的基于粒子群的结构学习方法, 其次M-FPSO算法迭代至20代左右就趋向于稳定, 而传统的基于粒子群的结构学习方法在35代左右才趋向于稳定。最后本文算法最终收敛结果也高于传统的基于粒子群的结构学习方法的收敛结果。由此, 可以看出使用基于改进MMPC算法的种群约束方法可以有效提高初始种群的质量, 并且基于

萤火虫算子的交叉变异操作加快了种群收敛速度。

为验证本文所提改进搜索策略的有效性, 与未改进的传统基于粒子群的贝叶斯网络结构学习方法(BNC-PSO)进行对比, 分别在CANCER、ASIA、ALARM 3种数据集, 采用同一初始化种群, 每种算法运行50次, 求取最终BIC评分的平均值, 结果如表5所示。

表5 算法改进前后的BIC评分结果

数据集	BNC-PSO	M-FPSO	标准网络
CANCER-500	-816.3	-815.4	-814.3
CANCER-1000	-1626.8	-1625.7	-1624.4
CANCER-3000	-4879	-4876.1	-4875.7
CANCER-5000	-8078.7	-8076.6	-8076.4
ASIA-500	-1223.1	-1218.9	-1216.1
ASIA-1000	-2365.5	-2358.8	-2357.4
ASIA-3000	-6847.8	-6840.3	-6840
ASIA-5000	-11420.2	-11411.9	-11409.4
ALARM-500	-7012.9	-6878.2	-6364.4
ALARM-1000	-12702.3	-12122.7	-11162.4
ALARM-3000	-33937.4	-32413.7	-29894.3
ALARM-5000	-54091.3	-51869.2	-48597.3

由表5可知, 在3种数据集上本文所提的M-FPSO算法的最终BIC得分均高于未改进的BNC-PSO算法。本文算法相较于传统BNC-PSO算法在CANCER、ASIA、ALARM 3种标准网络上学习的精度分别提高了约68.7%、65.5%、34.1%。可见本文所提出的改进方法有效地提高了算法的寻优精度。

3.3.2 不同算法的比较分析

为验证本文算法的性能, 分别与BNC-GA算法、BNC-PSO算法、MM-ACO算法、MMHC算法进行对比分析。在同一种网络结构上, 分别采用500、1000、3000、5000的4种不同样本个数的数据集, 在每个数据集上, 每种算法分别运行50次, 对50次的运行结果求取平均值, 构建表6~8。

表6 ASIA网络的学习结果

算法	指标类型	ASIA			
		500	1000	3000	5000
标准网络	BIC	-1216.1	-2357.4	-6840	-11409.4
BNC-PSO	BIC	-1229.5	-2366.1	-6850.8	-11426.5
	SHD	7	6.26	6.56	7.22
BNC-GA	BIC	-1249.5	-2403.2	-6903.5	-11492
	SHD	10.1	8.3	11.6	12.2
MM-ACO	BIC	-1220	-2289.7	-6841.8	-11414.7
	SHD	4.8	3.3	1.2	0.9
MMHC	BIC	-1228.6	-2379.4	-6866.1	-11484.9
	SHD	6.6	6.1	4.2	3.2
M-FPSO	BIC	-1218.9	-2358.8	-6840.3	-11411.9
	SHD	4.6	2.7	0.6	0.3

表 7 ALARM 网络的学习结果

算法	指标类型	ALARM			
		500	1 000	3 000	5 000
标准网络	BIC	-6 364.4	-11 162.4	-29 894.3	-48 597.3
	SHD				
BNC-PSO	BIC	-7 049.1	-12 799.3	-35 406.2	-54 947.6
	SHD	41.3	37.5	31.3	19.8
BNC-GA	BIC	-6 873.3	-12 790.5	-35 652.8	-56 326.1
	SHD	38.7	37.3	32.7	24.1
MM-ACO	BIC	-7 095.9	-12 205.6	-32 835.2	-53 151.24
	SHD	42.5	23.9	16.7	14.2
MMHC	BIC	-6 970.3	-12 576.6	-34 402.5	-53 183.3
	SHD	35.6	32.4	25.6	14.3
M-FPSO	BIC	-6 878.2	-12 122.7	-32 413.7	-51 869.2
	SHD	28.2	21.9	14.3	10.2

表 8 CANCER 网络的学习结果

算法	指标类型	CANCER			
		500	1 000	3 000	5 000
标准网络	BIC	-814.3	-1 624.4	-4 875.7	-8 076.4
	SHD				
BNC-PSO	BIC	-817.2	-1 627	-4 879.2	-8 080.6
	SHD	2.2	1.8	2.5	2.3
BNC-GA	BIC	-818.9	-1 630.4	-4 884.8	-8 081
	SHD	3.2	3.3	2.9	2.6
MM-ACO	BIC	-817.7	-1 631	-4 886.3	-8 079.1
	SHD	3	4	3.1	1.7
MMHC	BIC	-820.3	-1 630.8	-4 878.1	-8 078.4
	SHD	4.6	3.9	1.7	1.1
M-FPSO	BIC	-815.4	-1 625.7	-4 876.1	-8 076.6
	SHD	1.8	1.6	0.3	0.1

表 6 表示不同 BN 结构学习算法在 ASIA 标准网络上的学习结果,本文算法在不同的数据量样本下最接近于标准 ASIA 网络 BIC 评分,并且具有最小的结构汉明距离。表 7 表示在 ALARM 标准网络上的学习结果,由于该网络具有 37 个节点,因此学习结果误差较大。但是本文算法得到的 BIC 评分仍然高于其它算法的 BIC 评分,最接近于 ALARM 标准网络评分,同时也拥有最小的结构汉明距离。表 8 表示在 CANCER 标准网络上的学习结果,该网络仅具有 5 个节点,虽然各算法的 BIC 评分并无较大差异,但本文算法的结果仍优于其它算法。

该仿真结果主要是因为本文提出的基于改进 MMPC 的种群约束方法相较于其他基于启发式的结构学习方法中随机初始化种群方法,能够得到较为接近真实网络结构的粒子种群,提高了初始种群的整体质量,在此基础上再利用基于萤火虫算子的交叉变异操作获得更为接近真实网络结构的 BIC 评分值。

为验证本文算法的收敛性,以容量为 3 000 的 ASIA 网络数据集为例,与 BNC-PSO、BNC-GA 以

及 MM-ACO 以上 3 种基于群智能的结构学习算法进行比较分析,得到的收敛曲线如图 8 所示。

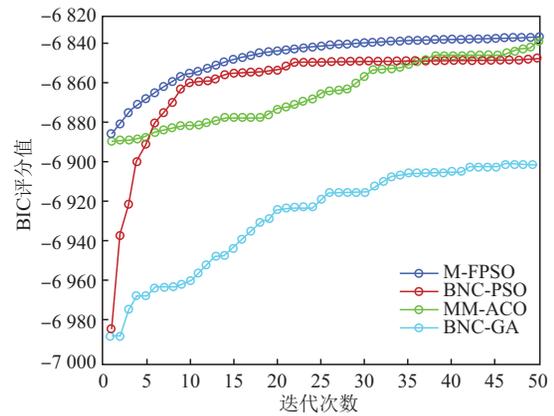


图 8 不同算法在 ASIA 网络上的收敛曲线

首先,图 8 验证了本文算法的收敛性能,M-FPSO 算法的收敛曲线一直位于其他 3 种收敛曲线上方,与其他算法相比具有更高的 BIC 评分值。其次,由图中可以看出 MM-ACO 算法与 M-FPSO 算法的起始位置较为接近,是因为这两种方法均对搜索空间进行约束,得到的初始种群较为优秀。由于经典的 BNC-PSO 算法与 BNC-GA 算法均没有对搜索空间进行约束,导致种群初始化随机性较强,初始种群最优个体的评分较低,种群质量较差,与 M-FPSO 算法相比开始时的收敛曲线位置相差较大。最后,M-FPSO 算法在迭代至 25 代左右趋于稳定,与其他对比算法相比收敛速度和精度都优于其他算法。

4 结论

1) 基于改进 MMPC 算法的种群约束方法通过构建含有局部信息的初始网络框架有效缩小了候选网络的搜索空间,提高了整个初始种群的质量,使算法在较高评分的网络结构中进行搜索。

2) 基于萤火虫算子的交叉变异策略,有效提高了算法的收敛效率和收敛精度。从实验结果可以看出,在样本量相同的情况下,该方法学习到的网络结构更接近于真实网络。

3) 虽然本文算法有着较为优秀的性能表现,但是随着节点数量的增加,在处理节点较多的网络时,时间开销与空间开销较大。未来,将会考虑研究包含一种基于最优父节点的结构学习算法,由此避免搜索过程中大量父节点带来的时间与空间巨大开销问题。

参考文献

[1] BARTLETT M, CUSSENS J. Integer Linear Pro-

- gramming for the Bayesian Network Structure Learning Problem [J]. *Artificial Intelligence*, 2017 (3): 258-271.
- [2] JIANG Y, LIANG Z, GAO H, et al. An Improved Constraint-Based Bayesian Network Learning Method Using Gaussian Kernel Probability Density Estimator [J]. *Expert Systems with Applications*, 2018, 113 (15):544-554.
- [3] BEHJATI S, BEIGY H. Improved K2 Algorithm for Bayesian Network Structure Learning[J]. *Engineering Applications of Artificial Intelligence*, 2020, 91 (5): 103617.
- [4] SUN X, CHEN C, WANG L, et al. Hybrid Optimization Algorithm for Bayesian Network Structure Learning[J]. *Information*, 2019, 10(10):294.
- [5] ASKARI M B A, AHSAAEE M G. Bayesian Network Structure Learning Based on Cuckoo Search Algorithm[C]//2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS). Iran: IEEE, 2018: 127-130.
- [6] LIU X, LIU X. Structure Learning of Bayesian Networks by Continuous Particle Swarm Optimization Algorithms [J]. *Journal of Statistical Computation and Simulation*, 2018, 88(8):1528-1556.
- [7] JI J, YANG C, LIU J, et al. A Comparative Study on Swarm Intelligence for Structure Learning of Bayesian Networks[J]. *Soft Computing*, 2017, 21 (11): 6713-6738.
- [8] 刘浩然, 张力悦, 范瑞星, 等. 基于改进鲸鱼优化策略的贝叶斯网络结构学习算法[J]. *电子与信息学报*, 2019, 41(6): 1434-1441.
- [9] 刘浩然, 苏昭玉, 张力悦, 等. 改进遗传-狼群对节点序寻优的贝叶斯网络结构算法[J]. *计量学报*, 2023, 44 (1):120-126.
- [10] 林朗, 张自力. 基于多头绒泡菌的贝叶斯网络结构学习[J]. *计算机科学*, 2019, 46(9):206-210.
- [11] 陈海洋, 尚珊珊, 任智芳, 等. 基于改进蚁狮优化的贝叶斯网络结构学习算法[J]. *空军工程大学学报*, 2023, 24(2):104-111.
- [12] WANG J, LIU S. A Novel Discrete Particle Swarm Optimization Algorithm for Solving Bayesian Network Structures Learning Problem[J]. *International Journal of Computer Mathematics*, 2019, 96 (12): 2423-2440.
- [13] GHEISARI S, MEYBODI M R. Bnc-Pso: Structure Learning of Bayesian Networks by Particle Swarm Optimization[J]. *Information Sciences*, 2016 (18): 272-289.
- [14] 高晓光, 邸若海, 郭志高. 基于改进粒子群优化算法的贝叶斯网络结构学习[J]. *西北工业大学学报*, 2014, 32(5): 749-755.
- [15] 尉永清, 陈小雪, 伊静, 等. 基于混合粒子群优化的贝叶斯网络结构学习方法[J]. *小型微型计算机系统*, 2018, 39(9): 2060-2066.
- [16] SUN B, ZHOU Y, WANG J, et al. A New PC-PSO Algorithm for Bayesian Network Structure Learning with Structure Priors[J]. *Expert Systems with Applications*, 2021(2):115237.
- [17] ANDREWS B, RAMSEY J, COOPER G F. Scoring Bayesian Networks of Mixed Variables[J]. *International Journal of Data Science and Analytics*, 2018, 6 (1):3-18.
- [18] 闫世瑛, 颜克斐, 方伟, 等. 基于差分进化邻域自适应的大规模多目标算法[J]. *系统工程与电子技术*, 2022, 44(7): 2112-2124.
- [19] 陈海洋, 张娜. 基于混合改进鸟群算法的贝叶斯网络结构学习[J]. *空军工程大学学报(自然科学版)*, 2021, 22(1):85-91, 98.
- [20] 苏树伟, 范科峰, 莫玮. 基于拓扑排序的贝叶斯网络结构学习方法[J]. *电视技术*, 2018, 42(5): 4-8.
- [21] 施伟勇. 基于蜻蜓优化算法的贝叶斯网络结构学习方法研究[D]. 长春: 吉林大学, 2022.
- [22] 姚洁, 朱响斌, 宋新方, 等. 基于节点排序的贝叶斯网络结构学习算法[J]. *计算机工程*, 2017, 34(5): 317-321.
- [23] WANG X, REN H, GUO X. A Novel Discrete Firefly Algorithm for Bayesian Network Structure Learning[J]. *Knowledge-Based Systems*, 2022(8):108426.
- [24] SONG W, QIU L, QING J, et al. Using Bayesian Network Model with MMHC Algorithm to Detect Risk Factors for Stroke[J]. *Math Biosci Eng*, 2022, 19(12): 13660-13674.
- [25] 郭文强, 毛玲玲, 黄梓轩, 等. 改进进化算法的贝叶斯网络结构学习及其应用[J]. *河南科技大学学报(自然科学版)*, 2022, 43(2):34-40.
- [26] JI J, HU R, ZHANG H, et al. A Hybrid Method for Learning Bayesian Networks Based on Ant Colony Optimization[J]. *Applied Soft Computing*, 2011, 11 (4): 3373-3384.
- [27] KHANTEYMOORI A R, OLYAEE M H, ABBA-SZADEH O, et al. A Novel Method for Bayesian Networks Structure Learning Based on Breeding Swarm Algorithm [J]. *Soft Computing*, 2018, 22 (7): 3049-3060.

(编辑:陈斐)