

A Bayesian Network Structure Learning Algorithm Based on the Combination of PSO and Sub-graph Decomposition

Shaorong Feng

School of Information Science and Engineering, Xiamen University, 361005 Xiamen, China

shaorong@xmu.edu.cn

Keywords: Bayesian network; Structure learning; Maximal prime decomposition technology; Particle swarm optimization algorithm

Abstract. In this paper we mainly focus on how to reduce the searching spaces when accelerating the efficiency of searching for the best Bayesian network structure and a corresponding algorithm named MPD-PSO based on the combination of MPD(maximal prime decomposition technology) and PSO(particle swarm optimization algorithm) was proposed. In the algorithm we proposed, we firstly use the MBDA (Markov Boundary Discovery Algorithm) algorithm to achieve the Markov boundary, and construct the undirected independent graph, then we decomposed the large undirected independent graph into several maximal prime sub-graphs by using the MPD algorithm, therefore we transform a high-dimensional structure learning problem into a low-dimensional structure (sub-graph structure) learning problem and reduce the searching space relatively. After that we use the PSO algorithm to learn the sub-graph, and we merge the learned sub-structure by correcting the wrong edges to achieve the best Bayesian structure. We validate the proposed algorithm by using Asia and Alarm network, and the results verify the superiority of the proposed algorithm in learning effect and running time over the GA (Genetic Algorithm) and PSO algorithms.

基于子图分解和 PSO 的贝叶斯网络结构学习算法

冯少荣

厦门大学 信息科学与技术学院, 中国 厦门 361005

shaorong@xmu.edu.cn

摘要: 本文研究通过减少搜索空间来加速寻找最优贝叶斯网络结构算法的效率, 提出了一种最大主子图分解技术和粒子群算法相结合的贝叶斯网络结构学习算法 MPD-PSO, 该算法首先利用 MBDA 算法获得变量的 Markov 边界, 并以此构建无向独立图, 再利用最大主子图分解技术 (MPD) 将大的无向图分解为多个主子图, 从而将高维的结构学习问题转化为低维的结构学习问题, 降低了搜索空间的大小。最后, 利用粒子群 (PSO) 算法分别对每个子图进行学习, 再将学习后的子网络通过合并及边的调整, 最终获得最优的贝叶斯网络结构。通过在 Asia 网络和 Alarm 网络上与遗传算法 (GA) 和粒子群算法 (PSO) 进行实验验证, 结果表明, MPD-PSO 算法在学习效果、运行时间上均优于所比较的算法。

关键词: 贝叶斯网络; 结构学习; 最大主子图分解; 粒子群算法

1. 引言

贝叶斯网络 BN (Bayesian Network) 以图形的方式来表示一组随机变量之间的依赖关系, 它仅描述变量之间已知的条件依赖关系, 并没有考虑领域中的所有变量, 因而它提供了获得和展示领域知识的有效方法。从数据中学习贝叶斯网络可以独立的分为结构学习和参数学习。贝叶斯网络结构的学习就是对给定的数据集, 学习得到一个与数据集拟合最好的网络。目前, 常用的结构学习方法大致可以分为三类:

(1) 基于依赖性测试的方法[1]，它是在给定数据集中评估变量之间的条件独立性关系，进而构建网络结构。该方法可以把条件独立性测试和网络结构的搜索分离开，但是对条件独立性测试产生的误差非常敏感，且在一些特定的情况下条件独立性测试的次数相对于变量的数目成指数增长。

(2) 基于评分搜索的方法[2]，其基本思想是利用打分函数寻找最高得分的贝叶斯网络结构，典型算法有 MMHC[3]、蚁群优化算法[4]等。基于评分的搜索方法在准确性方面较第一种方法有所提高，但仍存在当网络节点较多时，网络结构的搜索空间呈指数级增长的缺点，Robinson 等人证明，若 BN(Bayesian network) 包含 n 个节点，则可能的 BN 结构数量为：

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \frac{n!}{i!(n-1)!} 2^{i(n-1)} f(n-i) \quad (1)$$

即随着变量数目 n 的增多，搜索空间的维数呈指数增长。因此，搜索最优的网络结构是一个 NP-hard 问题[5]，一般需要使用启发式算法。代表性算法有 K2 算法[6]等。

(3) 混合方法，混合方法融合了上述两类算法的思想。首先利用依赖统计分析，缩减网络结构空间的大小，产生初始的架构，然后利用评分的方法对网络空间进行搜索得到最优的网络结构。典型的算法有最大最小爬山算法 MMHC(Max-Min Hill-Climbing)[7]等。该类方法可以降低搜索空间的复杂度，加速收敛速度，但算法在降低搜索空间复杂度的同时容易陷入局部最优。

本文提出了一种最大主子图分解技术和 PSO[7] 相结合的算法(Maximal Prime Sub-graph Decomposition and Particle Swarm Optimization, MPD-PSO)。该算法结合了条件独立测试、分解技术和局部搜索方法，融合了基于依赖性测试方法和基于评分方法各自的优点，在不破坏变量间的统计关系的基础上实现了对网络结构的分解，从而降低了算法的时间复杂度，同时也提高了学习的准确度。

2. 基本概念

2.1. 贝叶斯网络

贝叶斯网络可以表示成一个二元组 $BN=(G, P)$ ，其中 $G=(V, E)$ 为有向无环图(Directed Acyclic Graph, DAG)，其中， V 表示节点集，与领域内的变量一一对应， E 表示有向边集，表示节点之间的概率依赖关系， P 表示节点的概率分布和节点间的依赖强度。

2.2. 贝叶斯结构学习原理

在样本数据集中，学习贝叶斯网络结构即找到一个能最好地解释样本数据集的有向无环图。假设有数据集 D ， S^h 表示网络结构的某种假设，从数据集 D 中学习最优的网络结构 S 等价于最大化 $P(S^h|D)$ 的过程，由贝叶斯定理可得：

$$P(S^h|C) = \frac{P(S^h, C)}{P(C)} = \frac{P(S^h)P(C|S^h)}{P(C)} \quad (1)$$

$P(C)$ 代表与结构无关的常数； $P(S^h)$ 代表网络结构的先验概率，也为常量； $P(C|S^h)$ 代表边界似然。所以 BN 结构学习问题可以定义为使 $P(S^h|D)$ 最大的 BN 结构[8]。

3. 最大主子图分解

在给出最大主子图的分解过程之前，下面先给出相关的定义。

3.1. MBDA 算法

本文利用 MBDA 算法获得每个节点的 Markov 边界，并以此构建道德图 G^m 。该算法首先通过低阶测试(0 阶和 1 阶)剔除与目标变量 t 无关的多余变量获得 $M(t)$ 的上界 $\overline{pc} \cup \overline{sp}$ ，再通过子集 $\overline{pc} \cup \overline{sp}$ 中利用 MMHC 算法进行搜索获得变量 t 的父节点和子节点。下面为 MBDA 算法^[9]的具体描述：

MBDA 算法

- (1) Input: Data set D ; Variable set $V=\{v_1, v_2, \dots, v_n\}$; Target variable t .
- (2) Initialization: $\overline{pc} = V \setminus t$, $\overline{sp} = \phi$.
- (3) Order-0 CI test: for each variable $v_i \in \overline{pc}$, if $Ind(v_i, t)$ is hold, then $\overline{pc} = \overline{pc} \setminus v_i, S(v_i) = \phi$.
- (4) Order-1 CI test: for each variable $v_i \in \overline{pc}$, if there is a variable $v_j \in \overline{pc} \setminus v_i$ such that $Ind(v_i, t | v_j)$, then $\overline{pc} = \overline{pc} \setminus v_i, S(v_i) = S(v_i) \cup v_j$.
- (5) Find superset of spouses: for each variable $v_i \in \overline{pc}$, if there is a variable $v_j \in V \setminus t \cup \overline{pc}$, such that $\neg Ind(v_j, t | S(v_j) \cup v_i)$, then $\overline{sp} = \overline{sp} \cup v_j$.
- (6) Find parents and children of t : call the MMHC algorithm to get: $pc(t) = MMHC(t, D(\overline{pc} \cup \overline{sp}))$.
For each $v_i \in \overline{pc} \setminus pc(t)$ if $t \in MMPC(v_i, D)$, then $pc(t) = pc(t) \cup v_i$.
- (7) Find spouses of t : for each variable: $v_i \in pc(t)$,
if there is a variable $v_j \in MMPC(v_i, D) \setminus \{pc(t) \cup t\}$
and a subset: $Z \subset \overline{pc} \cup \overline{sp} \setminus \{t \cup v_j\}$, such that
 $Ind(t, v_j | Z)$ and $\neg Ind(t, v_j | Z \cup v_i)$
then $sp(t) = sp(t) \cup v_j$.
- (8) return $M(t) = pc(t) \cup sp(t)$.
- (9) output: A Markov boundary $M(t)$ of t .

3.2. 基本定义

定义 1: 道德图

假设 $G=(V, E)$ 为有向无环图, 若 G 中的任意两个节点 Y_i, Y_j 有共同的子节点, 则添加一条新的无向边 $e=(Y_i, Y_j)$ 到边集 E , 同时去掉 E 中所有边的方向由此形成的图称为 G 的道德图 G^m 。

定义 2: 极大团和 Markov 树

如果节点集 V 的一个子集 V' 在 $G=(V, E)$ 中是完全连接的, 则称 V' 为 G 的一个团, 记为 C , 若任何包含 V' 的节点集都不是团, 则 C 为 G 的一个极大团。

设团树 $T=\{V'', E''\}$, 若 T 的节点集 V'' 由构成 G 的所有极大团 $\{C_1, C_2, \dots, C_k\}$ 组成, 且 $\{C_1, C_2, \dots, C_k\}$ 包含的节点等于节点集 V , T 中任意边 $(C_i, C_j) \in E''$ 当且仅当分离集 $C_i \cap C_j \neq \phi$ 存在。对于 T 上任意团 C , 如果 C 在团 C_i, C_j 之间的路径上, 并且包含 $C_i \cap C_j$, 则该团树为一个 Markov 树(Markov Tree, MT)。

定义 3: 子图

设 $G=(V, E)$ 是一个图, 节点集 $V' \subset V$, 以 V' 为节点集, 以 G 中两端点均属于 V' 的边为边集所构成的图, 称为 G 的由节点集 V' 导出的子图, 记为 $G[V']$ 。

定义 4: 最大主子图

设 $G^m=(V, E^m)$ 是一个无向图, 若存在 V 的子集 U, S, W 且 $U \cup S \cup W = V$, 使得 G^m 中每一条经过点 $u_i \in U$ 到达点 $w_i \in W$ 的无向路上至少存在一个点在 S 中, 则称 S 分离子集 U 和 W ; 如果 $G^m[S]$ 在 G^m 中是完全的, 则称 G^m 是可分解的, 且分解为子图 $G^m[U \cup S]$ 和 $G^m[S \cup W]$ 。

若 $G^m[U]$ 是 $G^m=(V, E^m)$ 的一个导出子图且 $G^m[U]$ 是不可分解的, 则称 $G^m[U]$ 是 G^m 的主子图, 若对任意子集 $w \supset U$, $G^m[W]$ 都是可分解的, 则称 $G^m[U]$ 是 G^m 的最大主子图。

定义 5: MPD Markov 树(MPD-MT)

对于 MT 中任意分割两个团结点 C', C'' 的分割集 C , 若分割集 C 导出的无向子图在 G^m 中是不完全的, 则合并 C 和 C', C'' , 最终得到 MPD-MT。

4. PSO 算法

PSO 算法最早的提出为了解决连续域的目标优化问题,由于实际问题往往是一些离散域问题,Kennedy 和 Eberhart^[10]首先在原来的算法上进行改进,提出了二值粒子群优化算法(BPSO),并且在具体的实践应用中,不同的学者对其进行了不同的改进和拓展。

本文利用二值 PSO 算法搜索给定数据集所构成的全部 DAG 网络空间。我们用邻接矩阵来表示 DAG,如图 1 和图 2 所示:

0 代表两个节点之间没有联系,1 代表两个节点间存在一条有向边。

为了更好地理解 PSO 算法在网络结构学习中的应用,下面对一些基本术语对应含义进行介绍。

粒子: 一个候选 DAG

粒子速度: 候选解在空间中移动。

种群: 选定的一组候选解,即为子图的 DAG 候选空间。

个体极值: 粒子从开始搜索到当前迭代的最优解。

全局极值: 种群从开始搜索到当前迭代的最优解。

适应度: 适应度函数对应对 DAG 的评分结果。

下面我们对每个粒子的位置、速度以及怎样得到更新后的位置和速度进行具体的介绍。

我们假设有 n 个粒子组成的群体在 D 维空间中以一定的速度飞行,则 PSO 算法的迭代公式可描述如下:

$$v_{id}^{k+1} = \mu v_{id}^k + c_1 r_1 (pb_{id}^k - x_{id}^k) + c_2 r_2 (gb_d^k - x_{id}^k) \quad (2)$$

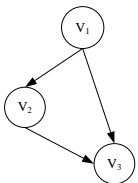
$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (3)$$

其中 μ 代表惯性系数, c_1, c_2 为学习因子 r_1, r_2 为随机数, i ($1 \leq i \leq N$) 代表种群的粒子数。PSO 初始化为一群随机粒子。然后通过迭代寻找最优解,在每一次迭代中,粒子通过跟踪个体极值 pb 和全局极值 gb 来更新自己。在找到这两个最优值时,粒子根据公式(2)和(3)来更新自己的速度和新的位置。

在整个 DAG 搜索空间中,每一个粒子对应一个候选的贝叶斯网络结构。其中粒子的速度和粒子的位置我们用 $n \times n$ 邻接矩阵来表示, n 表示节点的个数,每个 DAG 中的节点的个数为粒子的维度。速度用来决定粒子新的位置,同可以根据两次位置的差异来移动粒子,速度矩阵我们用离散值来表示:

$$V_{ij} = \begin{cases} 1 & \text{添加节点间的边} \\ 0 & \text{两节点间边保持不变} \\ -1 & \text{删除节点间有向边} \end{cases} \quad (4)$$

其中-1 表示删除节点 i 和节点 j 之间的边,0 表示点 i 和节点 j 之间的依赖关系保持不变,1 表示在节点 i 和节点 j 之间添加一条有向边。变换过程如图 1 至 4 所示:



$$\begin{bmatrix} & V_1 & V_2 & V_3 \\ V_1 & 0 & 1 & 1 \\ V_2 & 0 & 0 & 1 \\ V_3 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} & V_1 & V_2 & V_3 \\ V_1 & 0 & -1 & 0 \\ V_2 & 0 & 0 & 0 \\ V_3 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} & V_1 & V_2 & V_3 \\ V_1 & 0 & 0 & 1 \\ V_2 & 0 & 0 & 1 \\ V_3 & 0 & 0 & 0 \end{bmatrix}$$

图 1 贝叶斯网络

图 2 对应位置矩阵

图 3 速度矩阵

图 4 当前位置矩阵

在第一代时,我们随机初始速度矩阵,从第二代开始,我们将根据公式(2)、(3)对速度进行更新,同时将更新后的速度转换为 $\{-1, 0, 1\}$ 中的某一离散值[11]。

5. MPD-PSO 算法

本章利用图论中的最大主子图分解技术结合粒子群算法提出了一种贝叶斯网络结构学习算法 MPD-PSO(maximal prime sub-graph decomposition)。该算法利用分解不会破坏随机变量的统计信息的特点,通过对网络结构进行最大主子图分解将高维的结构学习问题转变为学习其子图的问题。其中子图的学习我们利用粒子群算法完成。实验结果表明,MPD-PSO 算法在 Asia 网络和 Alarm 网络数据集上得到的学习效果均优于遗传算法(GA)和粒子群算法(PSO)。

下面我们对 MPD-PSO 算法进行详细的描述,并对仿真实验结果进行分析。

5.1. MPD-PSO 算法步骤

MPD-PSO 算法

(1) 首先通过 MBDA 算法得到每个节点的 Markov 边界,以此构建道德图 G^m ;

(2) 根据 2 中介绍的最大主子图分解方法,对 G^m 进行最大主子图进行分解;

(3) 对分解后的每个子图运用粒子群算法学习其结构,学习过程如下:

①生成子图对应数据集的所有有向无环图 $\{DAG_1, DAG_2, \dots, DAG_n\}$, 并对其进行矩阵编码,以此为初始种群,设置最大迭代次数 \max_r ;

② for each particle:

初始化粒子参数,速度矩阵 V 和位置矩阵 P ;

do {

for each particle:

{ 计算适应度(采用 MDL 评分^[12]) p_i

if $p_i > pb$ $pb = p_i$

}

$gb = \max(p_1, p_2, \dots, p_n)$ (n 为粒子数)

for each particle:

根据公式(2)、(3)更新速度(进行离散)矩阵和位置矩阵;

检验此时形成的结构是否符合 BN 定义,若存在环路则删除此次迭代,进入下一次迭代

$r = r + 1$;

} while $r < \max_r$

(4) 将学习到的各个子图合并修正得到最优的贝叶斯网络结构。

5.2. 仿真实验

为了验证本文所提方法的有效性,本文选用 Asia 网和 Alarm 网数据为实验数据进行模拟实验。实验环境为 Windows 7, CPU 主频 3GHz, 实验工具为 Matlab2012a 和 FullBNT-1.0.4 工具箱。

我们将本文提出的 MPD-PSO 算法与 PSO 算法及 GA 算法进行网络结构学习得到的结果进行比较。我们通过比较实验所得的网络与标准网络的汉明距离(汉明距离=多余的边+丢失的边+反向的边)来衡量算法在学习贝叶斯网络时的准确度,由于算法在实现过程中第一代数据为随机初始化数据,所以我们选取 10 次实验所得结果的平均值来进行比较。表 2 和表 3 分别给出了在 1000、3000、5000 组 Asia 和 Alarm 网络数据上应用 GA、PSO、MPD-PSO 算法得到的贝叶斯网络结构与标准网络结构的汉明距离。图 5 和图 6 分别给出了 GA、PSO、MPD-PSO 算法在 Asia 和 Alarm 网络上运行的时间。从实验结果分析,我们可以看出,无论是从汉明距离还是从运行时间的角度来看,MPD-PSO 算法均优于 GA 算法和 PSO 算法。

MPD-PSO 在两组数据集的不同样本大小上均获得了较小的结构海明距离,明显提高了贝叶斯网络结构学习的准确性。从 Asia 网络数据集转换为 Alarm 网络数据集,三种算法得到的结构汉明距离均有所提高,即随着变量数目的增加,三种算法在学习效果上均有所退化,但从结构汉明距离的增长程度来看,本文提出的 MPD-PSO 算法相对更稳定。

表 2 Asia 网实验结果			
	GA	PSO	MPD-PSO
1000	3.2	2.4	1.6
3000	1.5	2.6	1.2
5000	1.4	1.4	0.9

表 3 Alarm 网实验结果			
	GA	PSO	MPD-PSO
1000	23.6	22.1	19.7
3000	21.2	20.3	15.4
5000	19.0	18.2	12.9

从图 5 和图 6 提供的数据整体来看,MPD-PSO 算法在不同大小的样本容量和不同的变量数目上的运行时间均小于其他两种算法。尤其在 Asia 网络上运行时,MPD-PSO 算法要明显优于 GA 和 PSO 算法,且在样本数量增大的过程中仍然保持这种优势。

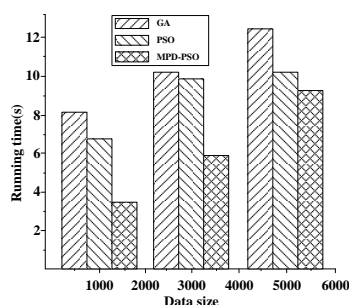


图 5 Asia 网上运行时间

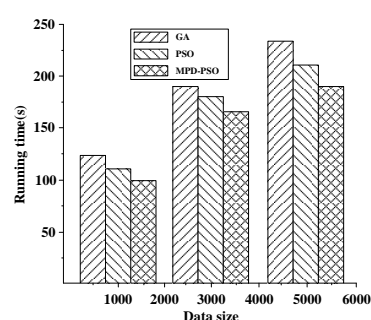


图 6 Alarm 网上运行时间

由以上分析可知,通过最大主子图分解技术对搜索空间进行限制可以在降低算法的复杂度的同时进一步提高学习的准确性,以较小的时间消耗获得较为精确的网络结构。

6. 结论

在贝叶斯网络结构学习过程中,搜索空间会随节点数的增加成指数增长,针对这一问题,本文提出了一种基于最大主子图分解技术和粒子群算法相结合的结构学习算法,并在 Asia 和 Alarm 数据集上与 GA 算法和 PSO 算法进行了比较,实验结果表明本文提出的 MPD-PSO 算法在相同的实验环境及相同的数据样本大小的前提下,消耗的运行时间较小且能得到较为准确的网络结构,从学习性能上有所提高。下一步的主要研究为,如何将遗传算法和粒子群算法的优势结合起来实现子图的学习,并将学习到的贝叶斯网络模型应用到分类当中。

7. 鸣谢

本文得到国家自然科学基金项目(No.61303004);国家社会科学基金重大项目(No.13&ZD148);福建省自然科学基金项目(No.2013J05099)的资助支持。

References

- [1] Dominik S. Degrees of conditional in dependence: a framework for approximate Bayesian networks and examples related to the rough set-based feature selection [J]. Information Sciences, 2009, 179(3):197-209.
- [2] Heckerman D. A tutorial on learning Bayesian networks MSR-T R-95-06[R]. Washington: Microsoft Research.
- [3] Jin yan, Hu yunan, Zhang jin, Song yanbo. Bayesian network structure learning with the combination of mutual information and mountain climbing method [J]. Computer applications and software, 2012, 29(9): 122-125.

- [4] Ji junzhong, Hu renbing, Zhang hong, Liu chun. A hybrid Bayesian network structure learning algorithm [J]. Computer research and development, 2009, 46(9):1498-1507.
- [5] R W Robinson. Counting unlabeled acyclic digraph [J]. Combinational Mathematics, 1977, Vol.622: 28-243.
- [6] Gregory F. Cooper, Edward Herskovits. A Bayesian Method for the Induction of Probabilistic Network from Data [J]. Machine Learning, 1992, 9(4):309-347.
- [7] Tsamardinos I, Brown L F, Aliferis C F. The max-min hill-climbing Bayesian networks structure learning algorithm [J]. Machine Learning, 2006, 65(1): 31-78.
- [8] Huang hexiao, Heng xingchen, Peng jianhan. Bayesian network structure learning algorithm based on particle swarm optimization [J]. Computer engineering and Applications, 2010, 46(20): 193-196.
- [9] Zhu mingmin, Liu sanyang, Yang youlong et al. Equivalence class learning algorithm based on hybrid method in Bayesian Networks [J]. Chinese Journal of Electronics, 2013, 41(1): 98-104.
- [10] Kennedy J, Eberhart R C. A Discrete Binary Version of the Particle Swarm Algorithm [C]. Proceeding IEEE International Conference System Man Cybericle, 1997: 4104-4109.
- [11] E.J. Clarke, B.A. Barton. Entropy and MDL Discretization of Continuous Variable for Bayesian Belief Networks [J]. International Journal of Intelligence Systems, Vol.15, 2000, pp.61-92.
- [12] J. Rissanen. Modelling by the shortest data description [J]. Automatica, Vol.14, 1978, pp.465-471.