

基于混合方式的贝叶斯网络等价类学习算法

朱明敏¹, 刘三阳^{1,2}, 杨有龙¹

(1. 西安电子科技大学理学院数学系, 陕西西安 710071; 2. 西安电子科技大学综合业务网国家重点实验室, 陕西西安 710071)

摘 要: 贝叶斯网络(BN)是不确定知识表示和推理的主要方法之一,是人工智能中重要的理论模型.针对现有混合方法学习 BN 结构不稳定、容易陷入局部最优等问题,本文将图论中的最大主子图分解理论与条件独立(CI)测试相结合,同时引入少量的局部评分搜索,提出一种新的基于混合方式的 BN 等价类学习算法.新算法通过确定所有变量的 Markov 边界构造网络的无向独立图,并对无向图进行最大主子图分解,从而将高维的结构学习问题转化为低维问题,然后利用低阶 CI 测试和局部评分搜索识别子图中的 V 结构.理论证明以及实验分析显示了新算法的正确性和有效性.

关键词: 贝叶斯网络; 最大主分解; Markov 边界; 有向无环图; 条件独立

中图分类号: TP18 **文献标识码:** A **文章编号:** 0372-2112 (2013)01-0098-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2013.01.018

Structural Learning Bayesian Network Equivalence Classes Based on a Hybrid Method

ZHU Ming-min¹, LIU San-yang^{1,2}, YANG You-long¹

(1. Department of Mathematics, Xidian University, Xi'an, Shaanxi 710071, China;

2. State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: Bayesian Network (BN) is one of the most important methods for representing and inferring with uncertainty knowledge, and also a powerful theory model within the community of artificial intelligence. To solve the drawbacks of hybrid methods for learning BNs which are easy to fall into local optimum and unreliable for learning large data set, we propose a novel hybrid algorithm for learning BN equivalence classes which combines ideas from maximal prime decomposition (MPD) of graph theory, conditional independence (CI) tests, and local search-and-score techniques in an effective way. It first reconstructs the undirected independence graph of a BN and then performs MPD to transform the undirected graph into its subgraphs. Finally, the new algorithm uses only lower-order CI tests and local BDeu score to check the v-structure of each subgraph. Theoretical and experimental results show that the proposed algorithm is correctness and effective.

Key words: Bayesian network; maximal prime decomposition; Markov boundary; directed acyclic graph; conditional independence

1 引言

贝叶斯网络 (Bayesian Network, BN), 又称贝叶斯信念网络 (Bayesian Belief Network, BBN) 或信念网络, 是人工智能领域表示不确定性知识和推理的一种重要方法. 经过十几年的发展, 它已广泛地应用于故障检测、医疗诊断、交通管理、军事目标自动识别、信息融合等方面^[1~3]. 目前, 从数据中学习 BN 受到了国内外学者们的广泛关注, 人们相继提出了许多种 BN 结构学习算法, 这些算法大致可归为两类: 基于约束的方法^[4~6]和基于

随机搜索机制的评分搜索方法^[7~9]. 此外, 一些研究者结合上述两种方法提出一类混合算法^[10~12], 该类算法首先利用条件独立测试 (Conditional Independence, CI) 降低搜索空间的复杂度, 然后执行评分搜索找到最佳网络. 如 MMHC (Max-Min Hill-Climbing) 算法^[12], 该算法分为 2 个阶段, 在第 1 阶段利用 MMPC (Max-Min Parents and Children) 算法构建贝叶斯网络结构的框架, 在第 2 阶段执行评分贪婪搜索确定网络结构的边以及边的方向. 然而, 这类算法在降低搜索空间复杂度的同时也缩小了解空间的范围, 而在评分搜索阶段由于空间的限制

无法修正这个错误,从而容易陷入局部最优,进而难以学习到最佳的网络结构。

针对上述情况,本文将图论中的最大主子图分解(Maximal Prime Decomposition, MPD)技术^[13]与 MMHC 算法结合,提出一种新的贝叶斯网络等价类混合学习算法(Structural Learning Bayesian Network Equivalence Classes based on a Hybrid Method,简称 BNEA)。新算法首先应用 CI 测试和 MMPC 算法得到所有变量的 Markov 边界并以此构造网络的无向独立图;然后利用 MPD 技术对无向图进行分解,理论证明,学习整个网络中的 V 结构等价于学习分解后的每个子图中的 V 结构,从而将一个高维的网络结构学习问题转化为低维问题;最后,利用低阶 CI 测试和局部 BDeu 得分搜索确定每个子图中的 V 结构。新算法不仅利用低阶 CI 测试有效地限制候选结构的选取范围,而且通过 MPD 技术对搜索空间进行分解,避免了不必要的统计计算和结构评分,有效地提高了算法的时间性能。理论和大量实验表明:与 MMHC 算法相比,新算法的学习速度有较大改进且具有更强的处理大规模数据的能力。

本文内容安排如下:第 2 节介绍贝叶斯网络及其最大主子图分解的基本概念;第 3 节给出基于条件独立测试和评分搜索的混合学习算法并给出算法的理论证明;第 4 节为数值实验结果;最后为结束语,同时给出了今后的研究方向。

2 贝叶斯网络及其结构学习

符号约定:在论述网络中的节点或概率分布中的随机变量时,大写字母 U, V, W 表示节点集合或变量集合,带下标的字母或单个字母,如 v_i, v 表示单个节点或变量。

2.1 贝叶斯网络

定义 1 BN 是一个二元组,即 $BN = (G, P)$, $G = (V, E)$ 为有向无环图(Directed Acyclic Graph, DAG),其中 V 为节点集,与领域的随机变量一一对应, E 为有向边集,反映节点变量之间的因果依赖关系; P 为节点的概率分布,表示节点之间因果影响强度。

任意给定的 BN 均满足 Markov 条件,即对任意变量 v_i 在给定其父节点集合 $pa(v_i)$ 情况下, v_i 独立于除 $pa(v_i)$ 之外的所有非子孙节点。我们用 $Ind(X; Y|Z)$ 表示给定变量集 Z , 变量集 X 和 Y 在概率分布 P 下是相互独立的, $X, Y, Z \subseteq V$ 。

注:若 BN 中存在有向边 $v_i \rightarrow v_j$, 则称 v_i 是 v_j 的父节点, v_j 是 v_i 的子节点,用 $pa(v_i)$ 表示 v_i 的父节点集合, $pc(v_i)$ 表示 v_i 的父节点和子节点构成的集合。若存在有向路 $v_i \rightarrow \dots \rightarrow v_j$, 则称 v_i 是 v_j 的祖先节点, v_j 是 v_i 的子孙节点,用 $an(v_i)$ 表示 v_i 的祖先节点集合。

定义 2 任意变量子集 $Z \subseteq V$, Zd -分离 ρ 当且仅当下面两个条件之一成立:(1) ρ 包含形如 $v_i \rightarrow v_j \rightarrow v_k$ 或 $v_i \leftarrow v_j \rightarrow v_k$ 的子链,并且 $v_j \in Z$; (2) ρ 包含形如 $v_i \rightarrow v_j \leftarrow v_k$ 的子链,并且 v_j 和 v_j 的子孙节点都不属于 Z 。

其中,有向图 $G = (V, E)$ 中的链是指一个有限非空序列 $\rho = v_1 e_1 v_2 \dots e_{l-1} v_l (2 \leq l \leq n)$, 它的各项交替地是节点和有向边, e_i 表示有向边 $v_i \rightarrow v_{i+1}$ 或 $v_i \leftarrow v_{i+1} (i = 1, 2, \dots, l-1)$ 。若 $v_i \rightarrow v_j \leftarrow v_k$ 是 G 中的子链,且 v_i 和 v_k 在 G 中不相邻,则称 $v_i \rightarrow v_j \leftarrow v_k$ 是 G 中的 V 结构。

特别地, G 中任意两个不同的子集 X, Y 被集合 Zd -分离当且仅当任意一条经过 $x_i \in X$ 到达 $y_i \in Y$ 的链被 Zd -分离。我们用 $Dsep(X; Y|Z)$ 表示变量集 X 和 Y 在 G 中被 Zd -分离, $X, Y, Z \subseteq V$ 。

对于任意 $BN = (G, P)$, 若概率分布 P 上所有变量之间的条件独立都可以通过有向图 G 上的 d -分离表示出来,则称 BN 满足忠实性条件。假设下文所述的 BN 均满足这一条件。

2.2 最大主子图分解

BN 结构学习就是在给定一个样本数据集 D 的前提下,寻找一个与训练集 D 匹配最好的网络结构, Robinson^[14]等人证明了包含 n 个节点可能的 BN 结构数目 $f(n)$ 为:

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

可见,随着变量数目 n 的增多,搜索空间的维数呈指数级增长,因此,搜索最优的网络结构是一个 NP 难问题。为此,本文受到图论中分解理论的启发,提出一种基于分解技术、条件独立测试和局部评分搜索的混合结构学习算法,该算法能够有效降低搜索空间的维数并提高计算效率。下面给出与图分解相关的基本概念。

定义 3 设 $G = (V, E)$ 是一个图,节点集 $U \subseteq V$ 。以 U 为节点集,以 G 中两 endpoint 均属于 U 的所有边作为边集所构成的子图,称为 G 的由节点集 U 导出的子图,记为 $G[U]$ 。

定义 4 设 $G^m = (V, E^m)$ 是一个无向图,若存在 V 的非空子集 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。

对一个无向图进行最大主子图分解,实质上是寻找这个图的所有最大主子图,一个图的最大主子图分

解是唯一的^[13]. BN 上的最大主子图是指 BN 的道德图的所有最大主子图. 有向图 G 的道德图是通过将 G 中的有向边转变为无向边并在具有共同子节点的节点之间添加无向边得到的, 这些被添加的边称为道德边, 我们用 G^m 表示 G 的道德图. 以图 1(a) 中的 Asia 网络为例, 它的道德图及其所有最大主子图分别如图 1(b) 和图 1(c) 所示. 根据定义 5 可知, 图 1(c) 中的每个子图都是不可分解的.

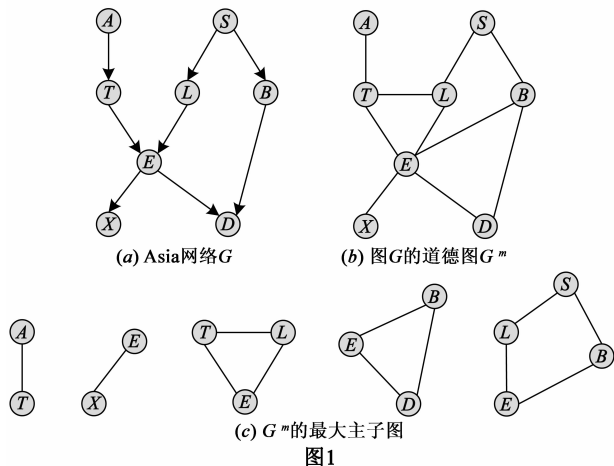


图1

3 基于分解混合方式的贝叶斯网络等价类结构学习

3.1 Markov 边界学习算法

如前所述, BN 以图结构的形式反映了节点变量间潜在的条件独立关系, 是随机变量之间联合概率分布的图形化表示. 如果任意两个不同的 BN s 定义了相同的联合概率分布, 则称这两个 BN s 是 Markov 等价的. 对于大部分评分函数, 如 BDeu 评分函数^[15], 等价的 BN s 具有相同的函数得分, 因此, 对 BN 结构的学习可以简化为学习其等价类. 该方法分为两个阶段实现. 第一阶段, 根据所观察到的数据通过 CI 测试或领域知识建立 BN 的道德图(最小无向独立图); 第二阶段, 根据道德图确定 BN 的 V 结构和其余的有向边. 本节针对第一阶段的学习, 给出一种学习 Markov 边界的算法(A Markov Boundary Discovery Algorithm, 简称 MBDA), 重复执行 MBDA 算法得到每个变量的 Markov 边界, 由这些 Markov 边界构成的无向图即为 BN 的无向独立图. 下面首先给出与 MBDA 算法相关的基本概念和定理, 然后给出算法的具体描述及其理论证明.

定义 5 设 t 是 $BN = (G, P)$ 中的任意变量且 $G = (V, E)$, 若存在最小的子集 $U \subset V$ 使 $Ind(t; V \setminus \{U \cup t\} | U)$ 成立, 则称 U 是 t 的 Markov 边界, 记 $U = M(t)$. 若 (G, P) 满足忠实性条件, 则变量 t 的父节点, 子节点及其配偶节点构成的集合是 t 唯一的 Markov 边界. 其中 t

的配偶节点集是指与 t 拥有共同子节点的节点构成的集合, 记为 $sp(t)$.

定义 6 设 $G = (V, E)$ 是有向无环图, P^W 是关于变量子集 W 上的联合概率分布, 其中 $W \subset V$. 若 G 满足子分布 P^W 上的所有条件独立, 则称 (G, P^W) 满足嵌入忠实性条件.

定理 1 设 $G = (V, E)$ 是有向无环图, P 是变量集 V 上的联合概率分布, 若 (G, P) 满足忠实性条件且 P^W 是子集 W 上的边缘分布, 则 (G, P^W) 满足嵌入忠实性条件.

定理 1 的证明可参阅文献^[16]. 定理 1 表明, 在忠实性条件成立的情况下, 对网络进行分解后子图上的概率分布满足嵌入忠实性条件, 从而可以有效降低数据维数并进行局部结构学习. 进一步, 确定变量 t 在变量集 V 上的 Markov 边界可转化为在 V 的某个子集上进行局部 CI 测试, 该子集包含变量 t 及与之相关的变量. 本文采用 χ^2 统计量^[17] 进行 CI 测试. 以下给出了 MBDA 算法的具体描述.

Algorithm 1 MBDA

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

根据算法 1 可知, MBDA 首先利用 0 阶和 1 阶 CI 测试删除与目标变量 t 无关的冗余变量, 得到 $M(t)$ 的上界 $\bar{p}c \cup \bar{p}sp$, 根据定理 1, $(G[\bar{p}c \cup \bar{p}sp], P^{\bar{p}c \cup \bar{p}sp})$ 满足嵌入忠实性条件. 然后利用 MMPC 算法在子集 $\bar{p}c \cup \bar{p}sp$ 中搜索 t 的父节点和子节点, 与 MMPC 算法直接在整个变量集搜索相比, MBDA 的这一策略有效限制了候选节点的选取范围, 避免了一些不必要的搜索. MBDA 算法的证明由以下定理给出.

引理 1 设 $G = (V, E)$ 是有向无环图, P 是变量集 V 上的联合概率分布, (G, P) 满足忠实性条件. 对任意目标变量 t , 若 $\bar{p}c \cup \bar{p}sp$ 是由算法 MBDA 得到的集合, 则

$M(t)$ 是 $p\bar{c} \cup s\bar{p}$ 的子集.

证明 根据定理 1, $M(t) = pc(t) \cup sp(t)$, 其中 $pc(t)$ 表示由变量 t 的父节点和子节点构成的集合, $sp(t)$ 表示变量 t 的配偶节点集合. 我们只需要证明 $pc(t) \subseteq p\bar{c}$ 和 $sp(t) \subseteq s\bar{p}$ 成立. 取任意节点 $v_i \in pc(t)$, 由于 (G, P) 满足忠实性条件, 因此, 任意变量子集 $Z \subseteq V$, 均有 $\neg \text{Ind}(v_i; t \mid Z)$ 成立, 由 MBDA 算法的第 3 步和第 4 步, 节点 v_i 不会被删除, 即 $v_i \in p\bar{c}$. 同理, 取任意节点 $v_j \in sp(t)$, 则存在节点 $v_i \in pc(t) \subseteq p\bar{c}$ 和子集 $Z \subseteq V$ 使得 $\text{Ind}(t; v_j \mid Z)$ 和 $\neg \text{Ind}(t; v_j \mid Z \cup v_i)$ 成立. 令 $Z = S(v_j)$, 则 $S(v_j)$ 在 G 中 d -分离 t 和 v_j . 根据算法的第 5 步可得 $v_j \in s\bar{p}$, 因此 $sp(t) \subseteq s\bar{p}$. 证毕

注: $\neg \text{Ind}(X; Y \mid Z)$ 给定变量集 Z , 变量集 X 和 Y 不是条件独立的.

引理 2 设 $G = (V, E)$ 是有向无环图, P 是变量集 V 上的联合概率分布, (G, P) 满足忠实性条件. 对任意目标变量 t , 若 $M(t) \subseteq W \subseteq V$, 则 t 在子集 W 上存在唯一的 Markov 边界 $M_W(t)$ 且 $M_W(t) = M(t)$.

证明 根据定理 1 和定理 2 易知变量 t 在 W 上存在唯一的 Markov 边界, 我们只需要证明当 $M(t) \subseteq W \subseteq V$ 成立时 $M_W(t) = M(t)$. 由 $M(t) \subseteq W \subseteq V$ 可得, $M(t) \subseteq M_W(t)$. 下面我们反证法证明 $M_W(t) \subseteq M(t)$.

不失一般性, 假设存在变量 $v_i \in M_W(t)$ 且 $v_i \notin M(t)$. 根据 $M(t)$ 的定义, 则有 $\text{Ind}(v_i; t \mid M(t))$ 在概率分布 P 上成立. 又因为 $M_W(t) = pc_W(t) \cup sp_W(t)$, 当 $v_i \in pc_W(t)$ 时, 则 $Dsep(v_i; t \mid M(t))$ 在 G 中不成立, 这与忠实性条件矛盾; 当 $v_i \in sp_W(t)$, 设 v_j 是 v_i 和 t 在子集 W 中的共同子节点, 如果 $v_j \in M(t)$, 同样有 $Dsep(v_i; t \mid M(t))$ 在 G 中不成立, 矛盾. 如果 $v_j \notin M(t)$, 则 $\text{Ind}(v_j; t \mid M(t))$ 在概率分布 P 上成立, 又因为 t 在 G 中是 v_j 的父节点, 从而有 $Dsep(v_j; t \mid M(t))$ 在 G 中不成立, 矛盾.

证毕

定理 2 设 $G = (V, E)$ 是有向无环图, P 是变量集 V 上的联合概率分布, (G, P) 满足忠实性条件. 对任意目标变量 t , 算法 MBDA 得到 t 的 Markov 边界.

证明 根据配偶节点的定义, 由 MBDA 算法的第 7 步可得到 t 在 G 中的配偶节点集. 下面证明由 MBDA 算法的第 6 步可以得到变量 t 在 G 中的父子节点集合, 即 $pc(t)$.

根据定理 2 和引理 1 可知, $(G, P^{pc(t) \cup s\bar{p}(t)})$ 满足变量集 $p\bar{c}(t) \cup s\bar{p}(t)$ 上的嵌入忠实性条件. 令 $p\bar{c}(t) \cup s\bar{p}(t) = W$, 则 $W \subseteq V$. 由 MMPC 算法可知, 算法第 6 步可得到 t 在子集 W 上的父子节点集合, 记为 $pc_W(t)$. 下面证明 $pc_W(t) = pc_V(t)$. 由引理 2 可知, $M_W(t) = M_V(t) \subseteq W$, 因此, $pc_V(t) \subseteq pc_W(t)$. 与引理 2 的证明类似, 我们

只需要证明 $pc_W(t) \subseteq pc_V(t)$. 不失一般性, 假设存在变量 $v_i \in pc_W(t)$ 且 $v_i \notin pc_V(t)$. 由于 G 中所有与 t 不相邻的节点均被 $M_V(t)$ 的某个子集 d -分离, 因此, 存在子集 $Z \subseteq M_V(t) \setminus v_i$ 使得 $\text{Ind}(v_i; t \mid Z)$ 成立. 又因为 $M_W(t) = M_V(t)$, v_i 与 t 被子集 $W \setminus \{v_i, t\}$ d -分离, 从而 v_i 与 t 在由子集 W 导出的子图中不相邻, 这与 $v_i \in pc_W(t)$ 矛盾, 定理得证.

证毕

3.2 等价类学习算法及其理论证明

在给出本文的主要算法之前, 首先给出两个定理, 作为算法的理论依据.

定理 3 设 $G^m = (V, E^m)$ 表示有向图 $G = (V, E)$ 的道德图, 任意 $v_i, v_j \in V$, 如果存在变量子集 S 在 G 中 d -分离 v_i 和 v_j 当且仅当 S 在 $G^m[an(v_i) \cup an(v_j) \cup \{v_i, v_j\}]$ 中分离 v_i 和 v_j .

定理 3 的证明可参阅文献[18]. 定理 3 表明, 有向图中变量间的 d -分离关系等价于其对应道德图中变量间的分离关系.

定理 4 设 $G^m = (V, E^m)$ 表示有向图 $G = (V, E)$ 的道德图, 假设 G^m 可以分解成它的最大主子图 $G_1^m, G_2^m, \dots, G_k^m$, 任意 $v_i, v_k, v_j \in V$, 如果 $v_i \rightarrow v_k \leftarrow v_j$ 是 G 中的 V -结构, 则至少存在 G^m 的一个最大主子图包含圈 $v_i - v_k - v_j - v_i$, 记作 $C(v_i, v_k, v_j)$.

证明 根据 G^m 的连通性, 一定存在 G^m 的子图包含边 $v_i - v_k$ 或 $v_j - v_k$. 下面用反证法证明. 假设边 $v_i - v_k$ 和 $v_j - v_k$ 分别属于 G^m 的两个不同子图 G_p^m 和 G_q^m , 由定义 5 可知, 一定存在子集 $S \subseteq V$ 且 $v_k \in S$ 使得 S 分离 G_p^m 和 G_q^m , 从而 v_k 在 G^m 中分离 v_i 和 v_j , 根据定理 3, v_k 在 G 中 d -分离 v_i 和 v_j , 这与 $v_i \rightarrow v_k \leftarrow v_j$ 是 G 中的 V 结构矛盾. 从而圈 $C(v_i, v_k, v_j)$ 至少包含在 G^m 的一个最大主子图中.

证毕

定理 4 表明, 确定 BN 中的所有 V 结构等价于将其道德图进行最大主子图分解, 对包含 3 个变量的圈进行 CI 测试, 因此, 定理 4 将一个高维的结构学习问题转化为低维的 CI 测试问题, 基于这一思想, 下面给出本文的主要算法.

Algorithm 2 BNEA

1. Input: Data set D ; Variable set $V = \{v_1, v_2, \dots, v_n\}$.
2. Construct an undirected independence graph G^m from the data set D :
 - * For each variable $v_i \in V$, call the MBDA to get the Markov boundary of v_i and represent it as an undirected sub-graph $\bar{G}_i (i = 1, 2, \dots, n)$.
 - * Combine $\bar{G}_1, \bar{G}_2, \dots, \bar{G}_n$ into an undirected graph $G^m = (V, E^m)$.
3. Call the junction tree algorithm to obtain maximal prime sub-graphs $G_1^m, G_2^m, \dots, G_k^m$ from the undirected independence graph G^m .
4. Set $P^c = G^m, C = \emptyset$.

5. repeat
6. if there exists a sub-graph G_l^m such that a cycle $C(v_i, v_k, v_j) \subseteq G_l^m$ then
7. if $Ind(v_i; v_j)$ and $\neg Ind(v_i; v_j | v_k)$ then
8. Orient $v_i \rightarrow v_k \leftarrow v_j$ and delete $v_i - v_j$ in P^c ;
9. else if $Ind(v_i; v_j | v_l)$ and $\neg Ind(v_i; v_j | v_k)$, $l \neq i, j, k$. then
10. Orient $v_i \rightarrow v_k \leftarrow v_j$ and delete $v_i - v_j$ in P^c ;
11. else
12. $C = C \cup C(v_i, v_k, v_j)$;
13. end if
14. end if
15. until P^c has not changed;
16. Orient the undirected edges in P^c by using local BDeu score search method if they appear in the list C ;
17. Orient other edges in P^c if each opposite of them creates either a directed cycle or a new v-structure;
18. return $P^c = (V, E^m)$
19. Output: PDAG $P^c = (V, E^m)$.

BNEA 算法首先通过 MBDA 子程序得到以每个变量 Markov 边界为节点集的无向子图, 将这些无向子图合并后的图结构即为对应 BN 的无向独立图; 然后, 对无向独立图进行最大主子图分解并检测分解后的每个子图是否包含 3 个变量的圈; 同时, 通过 CI 测试和局部 BDeu 评分搜索确定每个子图中的 V 结构, 并在无向独立图中进行更新, 直到不存在满足条件的子图; 最后对等价类中其余的无向边定向. 由算法 2 可见, BNEA 不仅利用 MPD 技术来分解搜索空间, 而且通过低阶 CI 测试和局部 BDeu 评分确定 V 结构, 从而极大地减少了许

多冗余边引起的统计因子的计算、结构的评分和比较的数量, 因此, 能够有效地提高算法的时间性能. 事实上, 与其他基于约束的算法一样, BNEA 算法的主要运算量在于子程序 MBDA, 即利用 CI 测试确定无向独立图, 最坏情况下, MBDB 与 MMPC 算法具有相同的时间复杂度. 因此, BNEA 算法的时间复杂度低于 MMHC 算法.

4 试验与分析

为了测试本文算法的性能, 我们采用标准的 Asia 网络^[19]和 Alarm 网络^[20]来完成下面的实验, 并将测试结果与 MMHC 算法进行了比较. 实验的运行环境为: 操作系统 Windows XP, CPU 为 Pentium4 3.06GHz, 内存为 0.99GB.

本文采用 BDeu 函数作为评分标准, BDeu 函数是基于 Bayesian Dirichlet 先验分布的贝叶斯评分函数的一种特殊形式, 如果取先验等价样本量为 10, 结构的先验信息为 0.001^σ , σ 表示网络中自由参数的数目, 则 BDeu 函数可表示为:

$$F_{BDeu}(BN, D) = \log \prod_{i=1}^n 0.001^{(r_i-1)q_i} \prod_{j=1}^{q_i} \frac{\Gamma(10/q_i)}{\Gamma(10/q_i + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(10/(r_i \cdot q_i) + N_{ijk})}{\Gamma(10/(r_i \cdot q_i))}.$$

其中 q_i 表示 V_i 父节点的取值数目, $\Gamma(\cdot)$ 表示 Gamma 函数, N_{ijk} 表示 V_i 取第 k 个值, 同时 V_i 的父节点取第 j 个值的样本数目, $N_{ij} = \sum_k N_{ijk}$.

表 1 两种算法在 Asia 数据集上的实验结果

Alg(Level α)	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$	$n = 10000$
BNEA(0.01)	(1.7, 1.7, 3.7)	(1.7, 1.6, 3.6)	(1.1, 0.9, 2.8)	(0.7, 0.4, 1.8)	(0.1, 0.1, 0.8)
BNEA(0.05)	(1.6, 1.3, 3.8)	(1.5, 1.4, 3.5)	(1.0, 1.0, 3.0)	(0.6, 0.5, 2.0)	(0.1, 0.2, 1.3)
MMHC(0.01)	(2.9, 1.6, 5.7)	(1.9, 1.4, 4.6)	(1.1, 0.8, 3.0)	(0.7, 0.3, 2.3)	(0.2, 0.1, 1.9)
MMHC(0.05)	(2.0, 1.1, 3.8)	(1.7, 1.4, 4.2)	(1.0, 0.8, 3.0)	(0.5, 0.5, 2.3)	(0.1, 0.2, 2.0)

表 2 两种算法在 Alarm 数据集上的实验结果

Alg(Level α)	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$	$n = 10000$
BNEA(0.01)	(5.4, 5.2, 19.8)	(3.2, 3.6, 16.7)	(2.0, 3.6, 15.2)	(0.9, 1.9, 10.2)	(0.6, 1.6, 8.7)
BNEA(0.05)	(3.8, 5.7, 19.8)	(2.6, 4.0, 16.7)	(1.4, 4.0, 15.2)	(0.4, 2.3, 12.2)	(0.4, 2.1, 11.3)
MMHC(0.01)	(6.5, 4.2, 19.8)	(4.5, 2.5, 16.7)	(3.2, 2.5, 16.7)	(1.7, 2.1, 11.2)	(0.8, 1.3, 9.6)
MMHC(0.05)	(5.4, 4.7, 21.8)	(3.5, 2.9, 15.0)	(2.6, 2.9, 15.2)	(1.3, 2.3, 13.3)	(0.7, 2.6, 13.1)

表 3 两种算法在 Alarm 数据集上的评分比较

Score	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$	$n = 10000$
$Sc(G_0)$	-5.2004e+003	-1.0000e+004	-1.9440e+004	-4.7737e+004	-9.4515e+004
$Sc(G_1)$	-4.9831e+003	-1.0652e+004	-2.0502e+004	-4.7558e+004	-9.4514e+004
$Sc(G_2)$	-6.8038e+003	-1.9824e+004	-3.0308e+004	-4.7826e+004	-9.4533e+004

表 1 和表 2 分别给出了本文算法和 MMHC 算法在不同样本容量数据集上独立运行 10 次的实验结果, 其中表 1 为 Asia 网络的实验结果, 表 2 为 Alarm 网络的实验结果, 表中括号中的 3 个数据分别表示所学到的最

佳网络结构与标准网络结构相比, 缺失边的数目、额外边的数目以及结构海明距离(指将算法得到的图结构转化为标准的网络结构所需要的平均运算总数目: 包括添加边、删除边、反转边). 从表 1 中可以看出, 当样本

量小于 2000 时, BNEA 与 MMHC 得到的实验结果很接近, 随着样本量的增大, BNEA 算法能够获得更好的求解质量. 当变量数目增大时, 得到表 2 的实验结果与表 1 类似, 并且 BNEA 比 MMHC 算法得到更小的结构海明距离值.

表 3 为不同样本容量数据集上的 BDeu 函数评分值, 其中 $Sc(G_0)$ 表示原标准网络的函数评分, $Sc(G_1)$ 和 $Sc(G_2)$ 分别表示由 BNEA 算法和 MMHC 算法得到的网络结构的函数评分. 从表 3 可见, 随着样本量的增加, $Sc(G_1)$ 和 $Sc(G_2)$ 逐渐向 $Sc(G_0)$ 逼近. 然而, 与 MMHC 算法相比, 本文算法在样本量小于 5000 时得到网络结构评分值与真实网络的评分值更加接近.

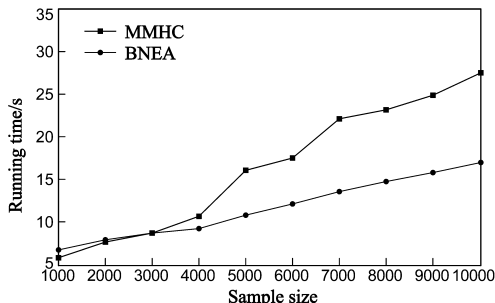


图2 两种算法在Alarm数据集上的时间性能比较

图 2 为相对应的运行时间比较, 由图 2 可知, 本文算法的运行时间在整个样本容量的范围上都好于 MMHC 算法, 而且在样本容量大的情况下时间性能的优势更加明显. 同时, 实验结果也表明: 随着样本量的增加, MMHC 算法的运行时间显著增长, 而本文算法对样本量的增加不敏感, 这说明与 MMHC 算法相比, 本文算法具有更好的处理较大数据集的能力. 总之, 从以上实验结果可见: 与 MMHC 算法相比, 本文算法无论是在结构误差, 还是函数评分值上都有一定的改善, 所以本文算法能够获得更好的结构学习质量; 另一方面, 在运行时间和实际结构评分数量比较上均有明显降低, 故算法的计算复杂度得到了显著改进.

5 结论

贝叶斯网络的结构学习是目前机器学习和数据挖掘领域中的一个研究热点. 以基于约束满足的方法为基础, 本文提出一种基于最大主子图分解、条件独立测试和局部评分搜索的混合贝叶斯网络等价类学习算法, 该算法有效地将高维的结构学习问题转化为低维问题, 降低了了条件独立测试的统计计算量和结构评分对比量; 同时文章从理论上严格证明了算法的正确性和有效性, 并通过实际问题进行了实验分析, 为贝叶斯网络用于解决实际问题提出了一种新的方案. 下一步的工作重点是利用混合方法对数据不完备、含隐变

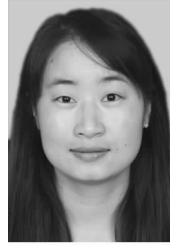
量及动态时变的贝叶斯网络结构进行系统地研究.

参考文献

- [1] A Aussem, S R de Morais. A conservative feature subset selection algorithm with missing data[J]. Neurocomputing, 2010, 73(4-6): 585-590.
- [2] 王洪泊, 涂序彦. 一种面向最经济服务流的可视化动态贝叶斯网络模型. 电子学报, 2011, 39(6): 1331-1335.
Hongbo WANG, Xuyan TU. A visual dynamic bayesian network model oriented most economical service-flows, Acta Electronica Sinica, 2011, 39(6): 1331-1335. (in Chinese).
- [3] S R de Morais, A Aussem. A novel Markov boundary based feature subset selection algorithm[J]. Neurocomputing, 2010, 73(4-6): 578-584.
- [4] J P Pellet, A Elisseeff. Using Markov blankets for causal structure learning[J]. Journal of Machine Learning Research, 2008, 9: 1295-1342.
- [5] C Borgelt. A conditional independence algorithm for learning undirected graphical models[J]. Journal of Computer and System Sciences, 2010, 76(1): 21-33.
- [6] Xianchao Xie, Zhi Geng. A recursive method for structural learning of directed acyclic graphs[J]. Journal of Machine Learning Research, 2008, 9: 459-483.
- [7] Xuewen Chen, G. Anantha, Xiaotong Lin. Improving Bayesian network structure learning with mutual information-based node ordering in the K2 algorithm[J]. IEEE transactions on knowledge and data engineering, 2008, 20(5): 1-13.
- [8] L Bouchaala, A Masmoudi, F Gargouri, A Rebai. Improving algorithms for structure learning in Bayesian networks using a new implicit score[J]. Expert Systems with Applications, 2010, 37(7): 5470-5475.
- [9] E Perrier, S Imoto, S Miyano. Finding optimal Bayesian network given a super-structure[J]. Journal of Machine Learning Research, 2008, 9: 2251-2286.
- [10] 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[J]. 自动化学报, 2009, 35(3): 281-288.
Junzhong Ji, Hongxun Zhang, Renbing Hu, Chunnian Liu. A Bayesian network learning algorithm based on independence test and ant colony optimization[J]. Acta Automatica Sinica, 2009, 35(3): 281-288. (in Chinese).
- [11] 贾海洋, 陈娟, 等. 基于混合方式的贝叶斯网弧定向算法[J]. 电子学报, 2009, 37(8): 1842-1847.
Haiyang JIA, Juan CHEN, et al. A Hybrid method for orienting edges of Bayesian network[J]. Acta Electronica Sinica, 2009, 37(8): 1842-1847. (in Chinese).
- [12] I Tsamardinos, L F Brown, C F Aliferis. The max-min hill-climbing BN structure learning algorithm[J]. Machine Learning, 2006, 65(1): 31-78.

- [13] K Olesen, A Madsen. Maximal prime subgraph decomposition of Bayesian networks[J]. IEEE Trans Syst Man Cybern B, 2002, 32: 21 – 31.
- [14] R W Robinson. Counting unlabeled acyclic digraphs[J], Combinational Mathematics, 1977, 622: 28 – 43.
- [15] D Heckerman, D Geiger, D M Chickering. Learning BNs: The combination of knowledge and statistical data [J]. Machine Learning, 1995, 20: 197 – 243.
- [16] R E Neapolitan. Learning Bayesian Networks[M]. Prentice-Hall, Englewood Cliffs, NJ, 2004.
- [17] S Kullback. Information Theory and Statistics [M]. Dover Publication, 1968.
- [18] S L Lauritzen. Graphical Models[M]. Clarendon Press, Oxford, 1996.
- [19] S Lauritzen, D Spiegelhalter. Local computations with probabilities on graphical structures and their application on expert systems[J]. J Royal Statistical Soc, 1988, 50: 157 – 224.
- [20] I Beinlich, G Suermondt, R Chavez, G Cooper. The ALARM monitoring mystem: a case study with two probabilistic inference techniques for belief networks[A]. Proc Second European Conf Artificial Intelligence in Medicine[C]. NK USA, 1989. 247 – 256.

作者简介



朱明敏 女, 1985 年生于陕西咸阳. 西安电子科技大学数学系博士研究生. 研究方向: 最优化理论、算法及其在贝叶斯网络学习中的应用.
E-mail: zmmzhu2010@126.com



刘三阳(通信作者) 男, 1959 年生于陕西西安. 西安电子科技大学数学系教授、博士生导师. 研究方向: 最优化理论、算法及其应用.
E-mail: liusanyang@126.com



杨有龙 男, 1967 生于陕西西安. 西安电子科技大学数学系教授、博士生导师. 研究方向: 图模型学习及其应用.
E-mail: ylyang@mail.xidian.edu.cn