

# 基于条件独立测试的链图结构学习算法

王静云,刘三阳,朱明敏

(西安电子科技大学数学与统计学院,陕西西安 710071)

**摘要:** 链图是贝叶斯网络和马尔科夫网络的自然推广,具有较强的表达能力.但目前关于链图结构学习算法的研究较少.本文基于贝叶斯网络结构学习的 Grow-Shrink 算法思想,提出一种链图等价类结构学习算法.该算法首先利用网络中结点的局部邻域信息,学习结点的邻接结点恢复网络骨架;然后根据链图复合体有向边的特点,利用条件独立测试确定网络的复合体有向边,从而恢复链图结构.理论分析和实验结果表明了该算法的正确性和有效性.

**关键词:** 链图; 结构学习; 条件独立测试; 马尔科夫性

**中图分类号:** TP18      **文献标识码:** A      **文章编号:** 0372-2112 (2017)10-2443-06

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2017.10.019

## Structure Learning of Chain Graphs Using the Conditional Independence Tests

WANG Jing-yun, LIU San-yang, ZHU Ming-min

(School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi 710071, China)

**Abstract:** Chain graphs (CGs) including both Bayesian networks (BNs) and Markov networks (MNs) as special cases, can express more independence models compared to the basic probabilistic graphical models. Today there exist several researches on structure learning of chain graphs. In this paper, we propose an algorithm for learning the equivalence classes of a chain graph based on the Grow-Shrink (GS) algorithm for structure learning of Bayesian networks. The algorithm works by first learning the adjacent nodes of each node in a chain graph for recovering its skeleton, and then discovering its complexes using the conditional independence tests and the property of the complexes. Theoretical analysis and experimental results demonstrate the effectiveness and correctness of the proposed algorithm.

**Key words:** chain graph; structure learning; conditional independence test; Markov property

### 1 引言

链图是贝叶斯网络和马尔科夫网络两种经典模型的推广,但不仅限于这两种模型.它包含两种类型的边,使其能够表示邻接变量之间的对称关系和非对称关系,因此具有更强的表示功能<sup>[1]</sup>.依赖于第二种边的解释,链图主要有三种不同的解释<sup>[2-5]</sup>,文献[6]中详细叙述了它们之间的关系,本文主要考虑 Lauritzen, Wermuth 和 Frydenberg 提出的 LWF CG<sup>[5]</sup>.

结构学习作为概率图模型研究的一个分支倍受关注,多种结构学习算法相继被提出,大致可归为两类<sup>[7]</sup>:基于约束的方法和基于随机搜索机制的评分搜索方法,由于这两种方法各有优缺,一些学者也提出将这两种方法进行融合的结构学习算法<sup>[8,9]</sup>,但这种方法

主要用于马尔科夫网络以及贝叶斯网络的结构学习.目前关于链图结构学习算法主要有:PC like 算法, M Studeny (1997)<sup>[10]</sup>将该算法用于学习 LWF CG,提出了 LCG 算法. J M Peña<sup>[11]</sup>和 D Sonntag<sup>[12]</sup>分别将该算法用于学习另外两种解释的链图; Ma (2008)<sup>[13]</sup>等提出的基于分解技术的结构学习算法 (LCD 算法),是对 Xie<sup>[14]</sup>提出学习贝叶斯网络分解算法的推广,该算法采用了分而治之的策略提高了搜索效率; J M Peña (2014)<sup>[15]</sup>等在扩展了 Meek 猜测的基础上提出了学习包含最优链图结构的 CKES 算法,该算法基于学习贝叶斯网络结构的 KES<sup>[16]</sup>算法思想,放松了对概率分布忠实性的假设,但由于链图等价类空间不可枚举,因此较难处理高维问题; D Sonntag (2015)<sup>[17]</sup>等提出了一种应用 ASP 学

习最优链图结构的算法,该算法在不假设概率分布的任何前提下能够找到包含最优的网络结构.上述算法均为贝叶斯网络结构学习算法在链图结构学习中的推广应用.由于目前很难找到有效的得分函数,链图结构恢复主要基于条件独立测试,然而类似于贝叶斯网络结构学习,在链图结构学习问题中,边的存在性判断需要指数阶的条件独立测试,因此很难处理较大规模的网络结构.随着马尔科夫边界理论知识的不断完善,涌现了大量利用马尔科夫边界学习 BNs 结构的算法. Margaritis 和 Thrun 提出的 Grow-Shrink (GS)<sup>[18,19]</sup> 是第一个理论上正确的马尔科夫边界学习算法,该算法将网络结构恢复问题转化为局部学习马尔科夫边界问题,利用结点的马尔科夫毯信息,通过学习结点的局部结构恢复整个网络结构,有效地减少了条件独立测试的次数,提高了结构学习效率.鉴于此,本文基于 GS 算法思想,在概率分布满足忠实性假设条件下,提出一种学习链图结构的 Grow-Shrink 算法 (GSCG),利用链图中结点的邻域信息,将链图结构学习问题转化为结点的局部结构恢复问题,减少结构学习的条件独立测试次数,并分别从理论和实验上说明了该算法的正确性和有效性.

## 2 基本定义和性质

**定义 1** 链图 (Chain Graph) 是一个二元组  $(G, P)$ , 其中  $G = (V, E)$ ,  $V$  是有限的结点集合,与领域的随机变量一一对应,  $E$  是由  $V$  中成对结点组成的集合,称为边集,  $P$  是定义在有限集合  $V$  上的联合概率分布.

链图中,任意互异结点  $u, v \in V$ , 如果  $(u, v) \in E$  且  $(v, u) \in E$ , 则称结点之间的边为无向边, 表示为  $u - v$  或者  $v - u$ , 其中  $u$  (或  $v$ ) 称为  $v$  (或  $u$ ) 的邻居结点,  $u$  的邻居结点  $Ne(u) = \{v | (u, v) \in E, (v, u) \in E\}$ ;

如果  $(u, v) \in E$  但  $(v, u) \notin E$ , 则称从  $u$  到  $v$  的边为有向边, 表示为  $u \rightarrow v$ , 其中  $u$  称为  $v$  的父结点,  $v$  称为  $u$  的子结点, 表示为  $Pa(v) = \{u | (u, v) \in E, (v, u) \notin E\}$ ,  $Ch(u) = \{v | (u, v) \in E, (v, u) \notin E\}$ .  $G$  中两个互异结点  $u, v \in V$  是邻接的, 如果  $u$  和  $v$  之间存在一条有向或无向边,  $u$  的邻接  $Adj(u) = Pa(u) \cup Ch(u) \cup Ne(u)$ , 定义  $u$  的边界  $Bd(u) = Pa(u) \cup Ne(u)$ .

图  $G$  中从  $u$  到  $v$  长度为  $n$  ( $n \geq 2$ ) 的路径是互异结点序列  $\{u \equiv u_0, \dots, u_n \equiv v\} \subseteq V$ , 满足  $(u_{i-1}, u_i) \in E$ ,  $\forall i = 1, \dots, n$ . 路径  $u_0, \dots, u_n$  ( $n \geq 0$ ) 称为下降路径, 如果对所有  $i = 1, \dots, n$ ,  $u_{i-1} \rightarrow u_i$  或者  $u_{i-1} - u_i$  在该路径中, 称  $u_n$  为  $u_0$  的后代结点, 相应的  $u_0$  称为  $u_n$  的祖先结点, 如果  $u_n$  是  $u_0$  的后代节点, 但  $u_0$  不是  $u_n$  的后代结点, 则称  $u_n$  是  $u_0$  的严格后代结点.  $V$  的一个子集  $X$  的后代结点是  $X$  中所有结点的后代结点集合, 表示为  $De(X)$ ; 集

合  $X$  的祖先结点是  $X$  中所有结点的祖先结点集合, 表示为  $An(X)$ ; 集合  $X$  的严格后代结点是  $X$  中所有结点的严格后代结点集合, 表示为  $Sd(X)$ .

概率分布  $P$  是定义在乘积空间  $\prod_{v \in V} X_v$  上的函数,  $P: \prod_{v \in V} X_v \rightarrow [0, 1]$ ,  $\sum \{P(x); x \in \prod_{v \in V} X_v\} = 1$ , 若  $P(x) > 0, \forall x \in \prod_{v \in V} X_v$ , 则称概率分布  $P$  是严格正的, 本文只考虑严格正的的概率分布. 令  $X, Y, Z$  是  $V$  中三个互不相交的子集,  $X$  和  $Y$  在给定条件  $Z$  时关于概率分布  $P$  是相互独立的, 表示为  $X \perp Y | Z [P]$ .

**定义 2** 图  $G$  中称互异结点序列  $(u_1, \dots, u_n)$  ( $n \geq 3$ ) 产生的子图为复合体, 若  $u_1 \rightarrow u_2, u_{n-1} \leftarrow u_n, u_i - u_{i+1}$  ( $i = 2, \dots, n-2$ ), 且  $u_1, \dots, u_n$  中其他结点对是不邻接的. 其中结点  $u_1$  和  $u_n$  称为复合体的父结点. 如果结点  $u$  和  $v$  为相同复合体的父结点则称  $u$  (或  $v$ ) 为  $v$  (或  $u$ ) 的配偶结点, 记  $u$  的所有配偶结点为集合  $SP(u)$ .

链图  $G$  的骨架图  $G^* = (V, E^*)$  是一个无向图, 其中  $E^*$  是  $G$  中去掉所有有向边的方向后得到的边集, 由  $G$  所决定的道德图  $G^m = (V, E^m)$  为无向图, 其中  $E^m$  是  $E^*$  与  $G$  中连接每个复合体父结点的无向边的并.

**性质 1**<sup>[2]</sup> 给定链图  $(G, P)$ ,  $G = (V, E)$ , 概率分布  $P$  是满足局部马尔科夫性的, 如果:

$$X \perp V \setminus \{X\} \setminus Sd(X) \setminus Bd(X) | Bd(X) [P], X \in V.$$

**性质 2**<sup>[10]</sup> 给定链图  $(G, P)$ ,  $G = (V, E)$ , 概率分布  $P$  满足全局马尔科夫性, 如果图  $(G_{An(X \cup Y \cup Z)})^m$  中  $Z$  分离  $X$  和  $Y$ , 则  $X \perp Y | Z [P]$ .  $(G_{An(X \cup Y \cup Z)})^m$  是  $An(X \cup Y \cup Z)$  在道德图  $G^m$  中的导出子图.

## 3 学习链图结构的 GSCG 算法

本文将链图结构学习分为两个阶段实现, 在第一阶段中首先基于 GS 算法思想, 根据观测数据恢复链图骨架, 并证明该方法的正确性. 第二阶段提出了一种根据链图中复合体有向边的特点, 通过条件独立测试确定复合体有向边的方法, 并证明该方法的正确性. 在算法介绍之前, 首先给出相关的概念和定理.

**定义 3**<sup>[20,21]</sup> 给定链图  $(G, P)$ , 概率分布  $P$  忠实于图  $G = (V, E)$ , 当且仅当:

$$X \perp Y | Z [G] \Leftrightarrow X \perp Y | Z [P].$$

**定理 1**  $G = (V, E)$  为链图, 如果概率分布  $P$  忠实于  $G$ , 则  $G$  中的每一对结点  $X$  和  $Y$  是邻接的, 当且仅当  $\forall Z \subseteq V, X \perp Y | Z [P]$ . 即  $X \in Adj(Y)$  或者  $Y \in Adj(X) \Leftrightarrow \exists Z \subseteq V \setminus \{X, Y\}$  s. t.  $X \perp Y | Z [P]$ .

**性质 3**  $G = (V, E)$  为链图, 如果概率分布  $P$  忠实于  $G$ , 结点  $X$  和结点  $Y$  在  $G$  中不邻接, 则  $X \perp Y | Bd(X) \cup Bd(Y) [P]$ .

定理 1 及性质 3 的证明可参阅文献[10].

**定理 2**  $G = (V, E)$  为链图, 如果概率分布  $P$  忠实于  $G$ , 则  $G$  中的任意结点  $X$  满足:

$$X \perp V \setminus \{X\} \setminus Adj(X) \setminus SP(X) \mid SP(X) \cup Adj(X) [P].$$

**证明** 在由链图  $G$  所决定的道德图  $G^m$  中, 由无向图的局部马尔科夫性可知, 对于任意结点  $X \in V, X \perp V \setminus \{X\} \setminus Ne_{G^m}(X) \mid Ne_{G^m}(X)$ ,  $Ne_{G^m}(X)$  是道德图中  $X$  的邻居结点集合. 由于道德图  $G^m$  中结点  $X$  的邻居  $Ne_{G^m}(X) = Adj(X) \cup SP(X)$ , 故由忠实性假设有

$$X \perp V \setminus \{X\} \setminus Adj(X) \setminus SP(X) \mid SP(X) \cup Adj(X) [P].$$

**性质 4**  $G = (V, E)$  为链图, 如果概率分布  $P$  忠实于图  $G$ , 则在  $G$  中  $\forall T \notin V \setminus \{X\} \setminus SP(X) \setminus Adj(X), X \perp T \mid SP(X) \cup Adj(X) [P]$  成立.

性质 4 很容易由概率分布  $P$  的分解性质得到<sup>[15]</sup>.

### 3.1 学习 BNs 的 GS 马尔科夫毯算法

随机变量  $X \in V$  的马尔科夫毯 (Markov Blanket)  $BL(X) \subseteq V$  是满足对于任意  $Y \in V \setminus BL(X) \setminus \{X\}, X \perp Y \mid BL(X)$  成立的集合. 变量  $X$  的马尔科夫毯是不唯一的, 最小的马尔科夫毯称为马尔科夫边界 (Markov Boundary). 在贝叶斯网络  $G = (V, E)$  中, 结点  $X \in V$  的马尔科夫边界  $B(X) \subseteq V$ , 包括其父结点、子结点及子结点的父节点, 即  $B(X) = Pa(X) \cup Ch(X) \cup \cup_{Y \in Ch(X)} Pa(Y)$ , 在忠实性条件下  $B(X)$  是  $X$  的唯一马尔科夫边界<sup>[22]</sup>.

算法 1 描述了 GS 算法的具体流程, 该算法由增厚和削薄两个阶段组成, 增厚阶段从一个空集  $S$  出发, 以当前的马尔科夫边界为条件集, 如果变量  $Y \in V \setminus \{X\}$  和目标变量  $X \in V$  不独立, 则将变量加入当前马尔科夫边界  $S$ , 直至所有候选变量都被测试. 由于这一过程可能有一些误报结点进入当前马尔科夫边界, 所以接下来进一步修正. 削薄阶段用当前马尔科夫边界作为条件集来删除误报结点, 如果目标变量  $X$  与当前马尔科夫毯中的变量  $Y \in S$  在  $S \setminus \{Y\}$  条件下独立, 则从  $S$  中删除变量  $Y$ , 最终返回  $X$  的马尔科夫边界  $B(X)$ .

#### 算法 1 GS 马尔科夫毯算法

```

1:  $S \leftarrow \emptyset$ 
2: repeat
3:   for each  $Y \in V \setminus \{S\} \setminus \{X\}$  do
4:     if then  $X \not\perp Y \mid S [P]$  then
5:        $S \leftarrow S \cup \{Y\}$ 
6:     end if
7:   end for
8: until  $S$  does not change
9: for each  $Y \in S$  do
10:  if  $X \perp Y \mid S \setminus \{Y\} [P]$  then
11:     $S \leftarrow S \setminus \{Y\}$ 
12:  end if

```

```

13: end for
14:  $B(X) \leftarrow S$ 

```

### 3.2 GSCG 算法描述

学习链图结构的 GSCG 算法采用分而治之的策略, 将整个网络结构学习分解为结点的局部结构恢复问题. 具体按如下步骤进行:

步骤 1: 基于算法 2 学习所有结点  $X \in V$  的配偶结点  $SP(X)$  和邻接结点  $Adj(X)$ ;

步骤 2: 利用算法 3 对步骤 1 中学习的  $SP(X)$  和  $Adj(X)$ ,  $X \in V$  做进一步修正;

步骤 3: 基于算法 4 恢复复合体有向边, 最终得到网络结构.

下面对各算法做进一步的解释和证明.

#### 算法 2 学习 $SP(X)$ 和 $Adj(X)$ 算法

```

1:  $SP(X) \leftarrow \emptyset; Adj(X) \leftarrow \emptyset$ 
2:  $SA \triangleq SP(X) \cup Adj(X)$ 
3: repeat
4:   for each  $Y \in V \setminus \{X\}$  do
5:     if  $X \not\perp Y \mid SA [P]$ 
6:        $SA \leftarrow SA \cup \{Y\}$ 
7:     end if
8:   end for
9: until  $SA$  does not change
10: for each  $Y \in SA$  do
11:   if  $X \perp Y \mid SA \setminus \{Y\} [P]$  then
12:      $SA \leftarrow SA \setminus \{Y\}$ 
13:   end if
14: end for
15: for each  $Y \in SA$  do
16:   if  $X \not\perp Y \mid B [P], \forall B \subseteq SA$  then
17:      $Adj(X) \leftarrow Adj(X) \cup \{Y\}$ 
18:   else
19:      $SP(X) \leftarrow SP(X) \cup \{Y\}$ 
20:   end if
21: end for

```

算法 2 首先基于 GS 算法思想, 在步骤 1 - 14 中不加区分地学习结点的配偶及邻接结点; 其次, 在步骤 15 - 21 中根据链图中结点的邻接结点以及配偶结点的性质, 确定该结点的邻居结点和配偶结点集合. 定理 3 给出了算法 2 的正确性证明.

**定理 3** 链图  $G = (V, E)$ , 概率分布  $P$  忠实于图  $G$ , 假定独立性测试是正确的, 且数据集  $D$  是独立抽样的样本集, 则算法 2 能够得到结点  $X \in V$  的配偶及邻接结点集合.

**证明** 首先证明如果在图  $G$  中结点  $Y$  是  $X$  的邻接结点或配偶结点, 则应用算法 2 后, 有  $Y \in SA$ . 在假定独

立性检验正确的前提下,如果在图  $G$  中  $Y$  是  $X$  的邻接结点,由定理 1 可知步骤 5 的条件满足,因此在步骤 6 中,结点  $Y$  被添加到集合  $SA$  中.如果在图  $G$  中  $Y$  是  $X$  的配偶结点,由性质 4 及  $SA$  的定义,步骤 5 的条件满足,同样在步骤 6 中结点  $Y$  被添加到集合  $SA$  中.而且在步骤 12 中均不会被移除.因此集合  $SA$  中包含了结点  $X$  的所有配偶以及邻接结点.

下证,若在图  $G$  中  $Y \in SA$  是  $X$  的邻接结点,应用算法 2 后  $Y \in Adj(X)$ .如果在图  $G$  中  $Y \in SA$  是  $X$  的邻接结点,则由定理 1 可知,在任意子集  $B \subseteq SA$  的条件下,步骤 16 中的条件满足,因此  $Y$  被添加到  $Adj(X)$  中.如果在图  $G$  中  $Y \in SA$  不是  $X$  的邻接点,由配偶结点的定义可知  $Y$  和  $X$  是某一个复合体的父节点,因  $Y$  不是  $X$  的严格后代结点,由性质 1 可知存在子集  $Bd(X) \subseteq SA$  使得  $X \perp Y | Bd(X)$ ,步骤 16 中的条件不满足,因此  $Y$  不会被添加到  $Adj(X)$  中.

最后证明,如果在图  $G$  中  $Y \in SA$  是  $X$  的配偶结点,应用算法 2 后  $Y \in SP(X)$ ,由上述证明可知,该结论显然成立.

算法 3 作为修正阶段是由于如果  $X, Y$  是邻接的则一定有  $Y \in Adj(X)$  且  $X \in Adj(Y)$ ,同样的,如果  $X, Y$  互为配偶结点则  $Y \in SP(X)$  且  $X \in SP(Y)$ .因此算法 3 显然是正确的.

---

#### 算法 3 进一步学习 $SP(X)$ 和 $Adj(X)$ 算法

---

```

1: for each  $X \in V$  do
2:   if  $Y \in Adj(X), X \notin Adj(Y)$  then
3:      $Adj(X) \leftarrow Adj(X) \setminus \{Y\}$ 
4:   end if
5:   if  $Y \in SP(X), X \notin SP(Y)$  then
6:      $SP(X) \leftarrow SP(X) \setminus \{Y\}$ 
7:   end if
8: end for

```

---

定理 4 给出了算法 4 的正确性证明.

---

#### 算法 4 结构恢复算法

---

```

1: for each  $X \in V$  do
2:   if  $Y \in Adj(X)$  then
3:     add edge  $X - Y$ 
4:   end if
5: end for
6: for each  $T \in SP(X)$  do
7:   if  $\exists Y \in Adj(X)$  s. t.  $X \not\perp T | B \cup \{Y\} [P]$  for
     all  $B \subseteq Adj(X) \setminus \{Y\}$  then
8:     orient  $X \rightarrow Y$ 
9:   end if

```

10: end for

**定理 4** 链图  $G = (V, E)$ , 概率分布  $P$  忠实于图  $G$ , 假定独立性测试正确,且数据集  $D$  是独立抽样的样本集,则算法 4 能够恢复图  $G$  的骨架和所有复合体.

**证明** 先证明算法 4 能够恢复  $G$  的骨架,由算法 2 和算法 3 的证明可知该结论成立.

下证算法 4 能够恢复所有复合体有向边.首先说明如果复合体有向边  $X \rightarrow Y$  在  $G$  中则算法 4 一定能够恢复该有向边.如果在图  $G$  中  $X \rightarrow Y$  是复合体有向边,由链图复合体的定义,则一定存在  $T \in SP(X)$ ,使得算法 4 中步骤 7 的条件满足,从而能够恢复方向  $X \rightarrow Y$ .下证,如果算法 4 恢复了复合体有向边  $X \rightarrow Y$ ,则  $X \rightarrow Y$  一定是图  $G$  中的复合体有向边.考虑三元组  $\langle X, T, Y \rangle, T \in SP(X)$ ,采用反证法证明,如果  $X \leftarrow Y$  或  $X - Y$  在图  $G$  中,则  $Y \in Bd(X) \subseteq Adj(X), T \in SP(X)$  因而  $T \notin Sd(X)$ ,由性质 1 可知  $X \perp T | Bd(X)$ ,即存在包含  $Y$  的子集  $Adj(X)$  使得  $X$  和  $T$  在该子集条件下独立,显然算法 4 中步骤 7 的条件不满足,因此不能指向  $X \rightarrow Y$ ,这与已知矛盾,故复合体有向边  $X \rightarrow Y$  在  $G$  中.

## 4 实验测试与分析

为了评价 GSCG 算法性能,本节给出了不同类型误差的响应曲线.并与文献[13]中提出的 LCD 算法进行了对比分析.文章中考虑三类误差:(1) TPR:恢复结构骨架中正确确定边占所有边的比例;(2) FPR:恢复结构骨架中错误确定边的个数与空缺边个数的比值;(3) SHD:恢复结构与标准结构相比,需要操作的次数,即(a)添加或者移除一条边,(b)增加、去除或者反转一条边的方向.实验环境:操作系统 Windows7, CPU 为 Pentium(R)3.20GHz,内存 2.00GB.所有实验均通过 R i386 3.1.3. lmk 实现.

实验参数  $p$ (结点个数),  $n$ (样本个数),  $N$ (每个结点平均邻接点的个数).取  $p \in (10, 20, 40), n \in (100, 300, 1000, 3000, 10000, 30000), N \in (2, 3)$ .

针对每组  $(p, N)$  依据文献[13]的方法随机生成不同链图,然后抽取样本容量为  $n$  的独立同分布样本.假设检验过程考虑显著性水平  $(\alpha = 0.01, 0.05)$ .在不同参数条件下,算法独立运行 30 次的平均实验结果如图 1~4 所示.

图 1~3 分别为本文算法和 LCD 算法在不同样本容量、不同结点个数及邻接点个数上的 TPR 和 FPR 误差分析曲线.图 4 为本文算法的 SHD 响应曲线.实验结果表明:TPR 随样本容量的增加而增大,尤其当结点个数  $p \in (10, 20)$  时,随样本容量增加 TPR 接近于 1.当显著性水平取 0.01 时 LCD 算法的 TPR 值较大,但当显著

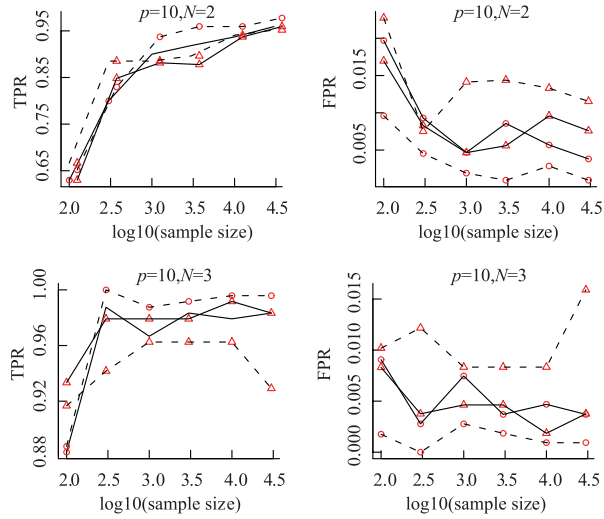


图1  $p=10, N=2,3$ 的TPR、FPR曲线, 实线表示本文算法, 虚线表示LCD算法, 圈和三角符号分别表示显著性水平0.01和0.05

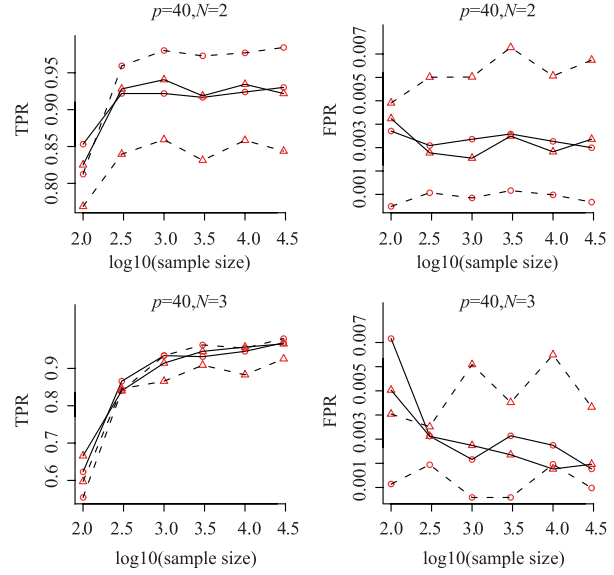


图3  $p=40, N=2,3$ 的TPR、FPR曲线, 实线表示本文算法, 虚线表示LCD算法, 圈和三角符号分别表示显著性水平0.01和0.05

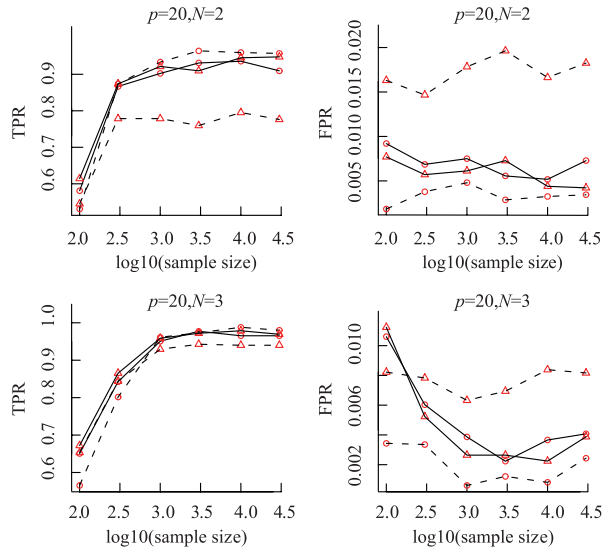


图2  $p=20, N=2,3$ 的TPR、FPR曲线, 实线表示本文算法, 虚线表示LCD算法, 圈和三角符号分别表示显著性水平0.01和0.05

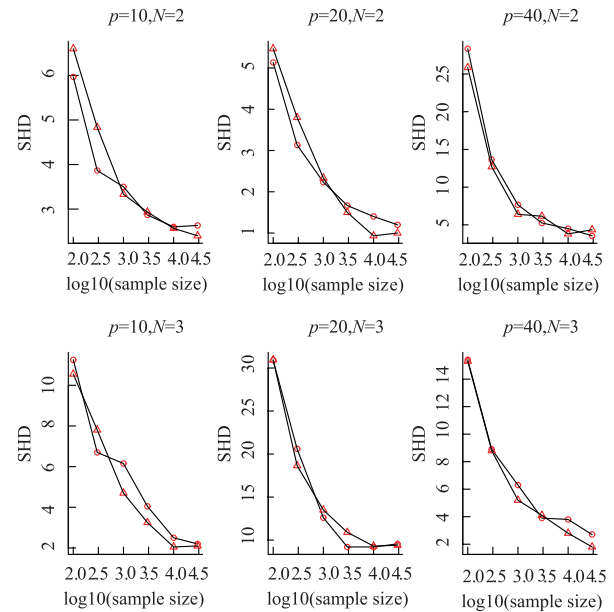


图4  $p=10, 20, 40; N=2,3$ 的SHD曲线, 圈和三角符号分别表示显著性水平0.01和0.05

性水平增大时 LCD 算法的 TPR 值明显小于本文算法; FPR 随样本容量的增加而减小, 但递减相对缓慢. 与 LCD 算法相比较, 在两种不同显著性水平下本文算法的 FPR 值都相对较小; 综合对比试验 LCD 算法受显著性水平的影响较大, 而本文算法受显著性水平的影响较小, 因此在实验中选择较为常用的显著性水平 0.05 即可; 由图 4 可以看出, 当样本容量增加时, SHD 呈明显的递减趋势. 由上述实验结果可见: 本文算法能够获得较好的结构学习质量, 尤其当结点个数相同时, 算法对  $N=2$  的网络结构恢复相对  $N=3$  时较好, 更适用于稀疏网络结构学习.

### 5 结论

链图模型作为一种工具, 日益得到重视. 从数据中学

习链图结构并进行推理和计算已成为该研究领域的重点和难点问题. 本文以基于约束满足的方法为基础, 提出了学习链图结构的 GSCG 算法, 文章从理论上证明了算法的正确性, 并通过对比实验说明了算法的有效性, 为链图结构学习提供了新的方案. 该算法的优势在于它限制了条件独立测试中条件集的大小, 减少了条件独立测试的次数, 而且基于独立测试的方法实现简单, 对于较大规模的稀疏网络结构, 该算法表现出良好的性能和精度. 但由于条件独立测试是以结点的整个邻接结点作为条件集的, 所以当网络中结点的邻接结点较多时, 高阶条件独立

测试计算复杂,且不可靠,故模型学习精度难以保证.因此,下一步的重点工作是针对大规模的复杂网络结构学习问题,提出相应的改进策略.

#### 参考文献

- [1] Dag Sonntag, J M Peña, Manuel Gomez-Olmedo. Approximate counting of graphical models via MCMC revisited [J]. *International Journal of Intelligence Systems*, 2015, 30(3): 384 – 420.
- [2] M Frydenberg. The chain graph Markov property [J]. *Scandinavian Journal of Statistics*, 1990, 17(4): 333 – 353.
- [3] S L Lauritzen, N Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative [J]. *Annals of Statistics*, 1989, 17(1): 31 – 57.
- [4] S A Andersson, D Madigan, M D Perlman. An alternative Markov property for chain graphs [J]. *Scandinavian Journal of Statistics*, 2001, 28(1): 33 – 85.
- [5] D R Cox, N Wermuth. Linear dependencies represented by chain graphs [J]. *Statistical Science*, 1993, 8(3): 204 – 283.
- [6] Dag Sonntag, J M Peña. Chain graph interpretations and their relations [J]. *International Journal of Approximate Reasoning*, 2015, 58: 39 – 56.
- [7] 王飞跃, 韩素青, 等译. 概率图模型原理与技术 [M]. 清华大学出版社, 2015.
- [8] Tsamardinos I, Brown L E, Aliferis C F. The max-min hill-climbing Bayesian network structure learning algorithm [J]. *Machine Learning*, 2006, 65(1): 31 – 78.
- [9] 朱明敏, 刘三阳, 杨有龙. 基于混合方式的贝叶斯网络等价类学习算法 [J]. *电子学报*, 2013, 41(1): 98 – 104.  
ZHU M M, Liu S Y, Yang Y L. Structural learning Bayesian network equivalence classes based on a hybrid method [J]. *Acta Electronica Sinica*, 2013, 41(1): 98 – 104. (in Chinese)
- [10] Milan Studeny. On recovery algorithm for chain graphs [J]. *International Journal of Approximate Reasoning*, 1997, 17(2 – 3): 265 – 293.
- [11] J M Peña. Learning AMP chain graphs under faithfulness [A]. *Proceedings of the 6<sup>th</sup> European Workshop on Probabilistic Graphical Models* [C]. Granada, Spain, 2012. 251 – 258.
- [12] D Sonntag, J M Peña. Learning multivariate regression chain graphs under faithfulness [A]. *Proceedings of the 6<sup>th</sup> European Workshop on Probabilistic Graphical Models* [C]. Granada, Spain, 2012. 299 – 306.
- [13] Zongming Ma, Xianchao Xie, Zhi Geng. Structure learning of chain graphs via decomposition [J]. *Journal of Machine Learning Research*, 2008, 9(9): 2847 – 2880.
- [14] Xianchao Xie, Zhi Geng, Qiang Zhao. Decomposition of structure learning about directed acyclic graphs [J]. *Artificial Intelligence*, 2006, 170(4 – 5): 422 – 439.
- [15] J M Peña, Dag Sonntag, Jens D Nielsen. An inclusion optimal algorithm for chain graph structure learning [A]. *Proceedings of the 17<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS)* [C]. Reykjavik, Iceland, 2014. 778 – 786.
- [16] J D Nielsen, T Kocka, J M Peña. On local optima in learning Bayesian networks [A]. *Proceedings of the 19<sup>th</sup> Conference on Uncertainty in Artificial Intelligence* [C]. Acapulco, Mexico, 2003. 435 – 442.
- [17] Dag Sonntag, Matti Jarvisalo, J M Peña, Antti Hyttinen. Learning optimal chain graphs with answer set programming [A]. *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence* [C]. Amsterdam, Netherlands, 2015. 822 – 831.
- [18] Dimitris Margaritis, Sebastian Thrun S. Bayesian network induction via local neighborhoods [A]. *Advances in Neural Information Processing Systems 12* [C]. Denver, USA, 2000. 505 – 511.
- [19] Dimitris Margaritis. Learning Bayesian network model structures from data [D]. Pittsburgh: Carnegie Mellon University, 2003.
- [20] J M Peña. Faithfulness in chain graph: The discrete case [J]. *International Journal of Approximate Reasoning*, 2009, 50: 1306 – 1313.
- [21] J M Peña. Faithfulness in chain graphs: the Gaussian case [A]. *Proceedings of the 14<sup>th</sup> International Conference on Artificial Intelligence and Statistics* [C]. Florida, USA, 2011. 588 – 599.
- [22] Richard E Neapolitan. *Learning Bayesian Networks* [M]. Pearson Prentice Hall, 2004.

#### 作者简介



王静云 女, 1988 年生于陕西咸阳, 西安电子科技大学数学与统计学院博士研究生. 研究方向: 最优化理论、算法及其在概率图模型学习中的应用.

E-mail: wangjingyun0607@163.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与统计学院教授、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