

贝叶斯网络结构学习的CMIHC算法

李晓晴, 于海征

引用本文

李晓晴, 于海征. 贝叶斯网络结构学习的CMIHC算法[J]. 计算机科学, 2023, 50(11A): 220800046-7.

LI Xiaoqing, YU Haizheng. CMIHC Algorithm for Bayesian Network Structure Learning[J]. Computer Science, 2023, 50(11A): 220800046-7.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合物品关系的图神经网络推荐算法](#)

Graph Neural Network Recommendation Algorithm Based on Item Relations

计算机科学, 2023, 50(11A): 230100019-9. <https://doi.org/10.11896/jsjcx.230100019>

[参数全局耦合的基因调控网络建模研究](#)

Modeling Gene Regulatory Networks with Global Coupling Parameters

计算机科学, 2023, 50(11A): 221100088-7. <https://doi.org/10.11896/jsjcx.221100088>

[基于演化序搜索的混合贝叶斯网络结构学习方法](#)

Hybrid Bayesian Network Structure Learning via Evolutionary Order Search

计算机科学, 2023, 50(10): 230-238. <https://doi.org/10.11896/jsjcx.221000046>

[一种基于两步搜索策略的K2改进算法](#)

Improved K2 Algorithm Based on Two-step Search Strategy

计算机科学, 2023, 50(9): 303-310. <https://doi.org/10.11896/jsjcx.220700253>

[基于子网融合的贝叶斯网络结构学习算法](#)

Sub-BN-Merge Based Bayesian Network Structure Learning Algorithm

计算机科学, 2022, 49(11A): 210800172-7. <https://doi.org/10.11896/jsjcx.210800172>

贝叶斯网络结构学习的 CMIHC 算法

李晓晴 于海征

新疆大学数学与系统科学学院 乌鲁木齐 830046

(812915644@qq.com)

摘要 贝叶斯网络源于对人工智能领域不确定问题的研究,是进行不确定问题推理和数据分析的重要工具。自贝叶斯网络结构学习诞生以来,已有众多成熟的结构学习算法,包括基于依赖分析的方法、基于评分搜索的方法和混合搜索的方法。其中利用信息论进行结构修剪已成为常用手段,但条件互信息中条件集的选取并没有统一标准,导致网络结构的修剪不一致。爬山算法利用3种搜索算子对网络结构进行局部更新,通过评分函数得到最优结构。结合信息论和爬山算法思想,提出一种新的结构学习算法——CMIHC(Conditional Mutual Information Hill Climbing)算法。该算法利用互信息和创建的条件集修剪初始连通图,对其进行定向,进而得到初始网络结构,结合评分函数和爬山算法的贪婪搜索策略得到最优网络结构。通过实验分析,在精度和效率上,CMIHC算法效果优于其他对比算法。

关键词: 贝叶斯网络;结构学习;条件互信息;爬山算法;CMIHC算法

中图分类号 TP181

CMIHC Algorithm for Bayesian Network Structure Learning

LI Xiaoqing and YU Haizheng

College of Mathematics and Systems Science, Xinjiang University, Urumqi 830046, China

Abstract Bayesian network originates from the research on uncertain problems in the field of artificial intelligence. It is an important tool for reasoning and data analysis of uncertain problems. Since the birth of Bayesian network structure learning, there have been many mature structure learning algorithms, including dependency analysis based method, score based search method and hybrid search method. Among them, structure pruning by information theory has become a common method, but there is no unified standard for the selection of condition set in conditional mutual information, resulting in inconsistent pruning of network structure. The hill climbing algorithm uses three search operators to update the network structure locally, and obtains the optimal structure through the scoring function. Combined with the idea of information theory and hill climbing algorithm, a new structure learning algorithm—conditional mutual information hill climbing(CMIHC) algorithm is proposed. The proposed algorithm prunes the initial connected graph by using mutual information and the created condition set, and orients it to obtain the initial network structure. Combined with the scoring function and the greedy search strategy of hill climbing algorithm, the optimal network structure is obtained. Experimental analysis shows that CMIHC algorithm is superior to other comparison algorithms in accuracy and efficiency.

Keywords Bayesian network, Structure learning, Conditional mutual information, Hill climbing algorithm, CMIHC algorithm

1 引言

贝叶斯网络是概率论与图论相结合的产物,它能够在复杂领域进行不确定推理和数据分析^[1]。近年来,贝叶斯网络已被应用于故障监测^[2]、风险分析^[3]、决策支持^[4]、医学诊断^[5]和信息检索^[6]等不同领域。

贝叶斯网络结构学习是通过学习给定样本,从候选网络结构中选出适应度最高的网络结构。传统的贝叶斯网络结构学习方法大致分为3类:基于依赖分析的方法、基于评分搜索的方法和混合搜索方法^[7]。

(1) 基于依赖分析的方法:通过条件独立性检验或专家先

验知识得到节点间的相关性,从而构建对应的网络结构;经典算法有SGS算法^[8]、PC算法^[9]和TPDA算法^[10]。2020年,Xu等^[11]通过 L_1 惩罚学习道义图,基于道义图贪婪搜索最优网络结构。Wang等^[12]基于自助法提出BPKL算法,该算法基于 B 个自助样本得到 B 个有向无环图,挑选与 B 个极大似然估计平均的惩罚KL距离最近的网络结构为最优。Zhao等^[13]提出基于马尔可夫毯的结构学习算法,减少条件独立性的检验次数,在定向阶段提出两种解决方案,有效提高了算法的学习效率。

(2) 基于评分搜索的方法:给定评分函数,利用搜索算法取评分最优的网络结构;经典算法有GS(Greedy Search)

基金项目:国家自然科学基金(61662079,11761070,U1703262);新疆维吾尔自治区自然科学基金面上项目(2021D01C078)

This work was supported by the National Natural Science Foundation of China(61662079,11761070,U1703262) and Xinjiang Natural Science Foundation(2021D01C078).

通信作者:于海征(yuhaizheng@xju.edu.cn)

算法、K2 算法^[14]和 HC(Hill Climbing)算法^[15]。GS 算法的搜索思想是在空图的基础上,不断加边,直到评分函数值不再改变为止。2020 年,Guo 等^[16]提出对候选父集大小进行剪枝的策略,该策略作为基于评分搜索算法的一部分,有助于产生解决高维问题的近似解。Behjati 等^[17]提出基于数据学习节点序的结构学习算法,该算法首先通过数据构建初始图,然后提取初始图的强连接组件 SCC 用于生成节点序;最后,将节点序输入 K2 算法学习最优的网络结构。2021 年,Lv 等^[18]提出基于 BIC 评分函数的节点序改进算法,该算法首先证明最依赖的节点间是无向的,从而构造无向子图 U_G ;其次,对 U_G 进行检查并将其连接成整体的无向图 U_{GC} ;最后,将 U_{GC} 中所有边定向转化为有向无环图,得到节点序,从而搜索最优网络结构。

(3) 混合搜索方法:利用依赖分析方法缩减搜索空间,结合评分搜索方法得到最优网络结构;典型算法有 MMHC(Max-Min Hill Climbing)算法^[19]。2021 年,Liu 等^[20]提出 ILSM 算法,由 Leaf 操作符、Root 操作符和 Swap 操作符组成 Momentum 因子,进行迭代局部搜索,遍历搜索空间,有利于脱离局部最优。Sun 等^[21]提出 PSO-GA 算法学习网络结构,该算法首先利用 PC 算法生成初始结构,构造新的变异交叉算子,并在搜索过程中结合粒子群算法和遗传算法寻找最优网络结构。

结合信息论和搜索方法进行结构学习的研究发展如下:2011 年,Wang 等^[22]提出基于互信息的结构学习算法,该算法由互信息构造无向边,通过引入人类属性集作为条件集,计算条件互信息并对边定向,结合专家知识修剪网络得到最优结构。Li 等^[23]提出基于互信息的改进算法,该算法由互信息构造最大生成树精简初始网络结构,并通过条件独立性检验对边定向,最后利用贪婪算法得到最优网络结构。2012 年,Jin 等^[24]提出 MI&HC(Mutual Information & Hill Climbing)算法,该算法利用互信息构建初始网络结构,再结合爬山算法进行结构学习。2014 年,Wu 等^[25]提出 K2&HC 算法,该算法结合 K2 评分函数和爬山算法,在爬山策略中融入回溯原理进行结构学习。2015 年,Liu^[26]提出简化贪婪(Simplified Greedy,SG)算法,该算法由互信息构建最大生成树,结合评分函数得到初始网络结构,对其补充和优化,从而得到最优的网络结构。2017 年,Liu 等^[27]提出改进爬山算法,该算法由互信息链得到初始网络结构,利用条件独立性检验以及对孤立节点加边得到完全结构,结合改进的爬山搜索算子进行结构学习。

上述文献的不足之处在于选取条件集导致修剪网络结构不一致。为解决这一问题,本文基于条件互信息和爬山算法提出 CMIHC 算法,综合互信息和条件互信息进行结构调整,进而得到完整的网络结构。该算法首先通过互信息构造条件集,在其基础上结合条件互信息对全联通图进行删减,对孤立节点加以处理得到的无向图取上三角阵定向作为初始网络结构;之后在初始网络结构上利用贪婪搜索的爬山算法,结合 BIC(Bayesian Information Criterion)评分函数进行结构学习,从而得到最优网络结构。实验将 CMIHC 算法与 GS 算法、HC 算法和 SG 算法进行对比,结果证明 CMIHC 算法的学习更加有效。

2 贝叶斯网络

贝叶斯网络由网络结构 G 和网络参数 P 两部分组成。

网络结构 $G = (V, E)$ 为一个有向无环图 DAG(Direct Acyclic Graph)。其中,节点集合表示为 $V = \{X_1, X_2, \dots, X_n\}$,节点 $X_i (i=1, 2, \dots, n)$ 代表第 i 个变量; $E = \{\langle X_i, X_j \rangle | X_i \neq X_j, X_i, X_j \in V\}$ 表示为有向边集合,边 $\langle X_i, X_j \rangle$ 表示节点 X_i 和 X_j 间具有 $X_i \rightarrow X_j$ 的依赖关系。

网络参数 $P = \{p(X_i | pa(X_i))\}$ 为一个条件概率表 CPT(Conditional Probability Table)。其中,给定 X_i 的父节点集 $pa(X_i)$ 时, X_i 的条件概率分布为 $p(X_i | pa(X_i))$ 。贝叶斯网络满足马尔可夫条件:节点 X_i 在给定其父节点 $pa(X_i)$ 的条件下独立于它的所有非后代节点。那么,贝叶斯网络所有节点的联合概率就可以表示为各节点概率的乘积,即:

$$\begin{aligned} P(X_1, X_2, \dots, X_n) &= P(X_1)P(X_2 | X_1) \cdots P(X_n | X_1, \\ &\quad X_2, \dots, X_{n-1}) \\ &= \prod_{i=1}^n P(X_i | X_1, X_2, \dots, X_{i-1}) \\ &= \prod_{i=1}^n P(X_i | pa(X_i)) \end{aligned} \quad (1)$$

贝叶斯网络实质就是一个联合概率分布 $p(X_1, X_2, \dots, X_n)$ 所有条件独立性的图形化表示。采用图的邻接矩阵 $dag = (A_{ij})$ 表示网络结构, $A_{ij} = 1$ 表示有向边由节点 X_i 指向节点 X_j , $A_{ij} = 0$ 表示节点 X_i 和节点 X_j 之间没有边连接,每一个邻接矩阵就代表一个网络结构。

3 基础知识

3.1 信息论基础

从信息论的角度,互信息用于衡量两个随机变量之间的相互关系。在观测到变量 Y 之前,变量 X 的不确定性用熵表示,记为 $H(X)$;观测到 Y 之后, X 的不确定性变成了 $H(X|Y)$ 。因此, $H(X)$ 与 $H(X|Y)$ 的差就是 X 和 Y 之间的互信息,记作 $MI(X, Y)$ 。

定义^[28] 对任意的两个离散型随机变量 X 和 Y ,互信息定义为在已知变量 Y 后,变量 X 的熵的减少量,称为 X 和 Y 之间的互信息,公式为:

$$\begin{aligned} MI(X, Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X, Y) \\ &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \end{aligned} \quad (2)$$

条件互信息 CMI(Conditional Mutual Information)描述了 3 个变量之间的条件独立关系,即在给定一个变量的情况下另外两个变量之间的互信息。假设 $H(X|Z)$ 是给定 Z 条件下 X 的不确定性, $H(X|Z, Y)$ 是给定 Z 条件下观测到 Y 后 X 剩余的不确定性。 $H(X|Z)$ 与 $H(X|Z, Y)$ 之间的差就是在给定 Z 条件下 X 和 Y 之间的条件互信息 $CMI(X, Y|Z)$ 。

定义^[28] 已知变量 Z 的条件下,变量 X 和 Y 的互信息称为条件互信息,公式为:

$$\begin{aligned} CMI(X, Y|Z) &= MI(X, Y, Z) - MI(X, Z) \\ &= H(X|Z) - H(X|Y, Z) \\ &= \sum_{x \in X} \sum_{y \in Y} \sum_{z \in Z} p(x, y, z) \log \frac{p(z)p(x, y|z)}{p(x|z)p(y|z)} \end{aligned} \quad (3)$$

性质^[29] 互信息 $MI(X, Y)$ 和条件互信息 $CMI(X, Y|Z)$

Z)具有如下性质:

(1)对称性,即 $MI(X,Y)=MI(Y,X)$ 和 $CMI(X,Y|Z)=CMI(Y,X|Z)$;

(2)非负性,即 $MI(X,Y)\geq 0$ 和 $CMI(X,Y|Z)\geq 0$ 。当且仅当 X 和 Y 条件独立时,有 $MI(X,Y)=0$;同理,当且仅当给定条件 Z 时, X 和 Y 条件独立有 $CMI(X,Y|Z)=0$ 。

利用互信息或条件互信息判断变量间的相关性,其取值范围为 $[0,1]$ 。值越接近于0,变量间越独立;值越接近于1,变量间越依赖。通常需要预先设置一个阈值 α 来判断边是否存在;若两节点间的(条件)互信息 $MI(X,Y)\leq \alpha$,则认为该节点间是(条件)独立的,并且在网络结构中表现为两节点间不存在边;反之,节点是相关的,并且在网络结构中表现为两节点间存在边。常选 α 值为0.05或0.01。

3.2 爬山算法

爬山算法采用启发式的贪婪搜索策略,进行结构学习的目标是寻找评分最高的网络结构。该算法基于一个给定的初始网络结构,利用3种操作算子对当前网络结构进行局部更新,从而获得一系列候选网络结构,然后计算这些候选网络结构的评分,选择评分最高的候选网络结构与当前网络结构进行对比;如果候选网络结构的评分大于当前网络结构,则将候选网络作为当前网络并开始下一轮的搜索;否则,认为当前网络结构是最优网络结构,停止搜索并返回当前网络结构^[30]。

爬山算法对边进行操作的3种搜索算子:加边,即对网络结构中两个未连接的节点间添加一条边;删边,即删除当前网络结构中已经存在的一条边;转边,即反转当前结构中已存在的某一条边的指向。在执行加边和转边算子的操作时,需保证不能在网络中产成有向环,因此需要增加一个判断是否产生有向环,若是则此操作无效^[30]。

爬山算法不依赖先验信息,只是利用3种搜索算子对初始网络中的边进行局部操作,借此产生一系列候选网络结构,结合评分函数对候选网络结构进行评估,把评分值最大的候选网络结构作为最终的网络结构。在实践中,一般选择多次执行爬山算法来得到其中评分最高的贝叶斯网络结构,因为爬山算法的贪婪搜索策略很容易陷入局部最优。

4 基于条件互信息的爬山算法

4.1 基于条件互信息的爬山算法原理

由信息论中的互信息和条件互信息衡量节点间的依赖关系,从而在节点间添加或者删除边构建网络结构。但是对于条件互信息中条件集的选取应用,文献方法并没有统一标准。爬山算法常使用空图作为初始网络结构,从而增加搜索的复杂度;文献中常利用最大生成树作为初始网络结构,一定程度上避免了陷入局部最优,但由于最大生成树构建的边数过于单薄,反而会忽视一些必要的边。

为解决以上问题,提出一种基于互信息构造无向图选择条件集的方法,便于缩小搜索空间;并且对结合互信息和条件互信息修剪的无向图进行定向,定向规则为取无向图邻接矩阵的上三角阵作为初始网络结构,结合评分函数和爬山搜索策略进行结构学习,获得最优网络结构,从而达到提高效率的目的。CMIHC算法具体思想如下:

(1)首先构造一个完全连通图 G_0 ,根据节点间互信息值与阈值的关,对边进行删减,从而产生初始无向图 G_1 ;

(2)处理无向图 G_1 中的孤立节点,将其与互信息值最大

的节点连接,构造一个初始连通图 G_2 ;

(3)在 G_2 中,寻找每个节点的邻居节点,将任意两个节点的邻居节点的交集作为条件集,并将条件独立性检验中的条件集大小限制为1;

(4)在给定条件集的前提下,计算节点间的条件互信息值,根据条件互信息值与阈值的关系,对边进行删减,得到一个修剪后的无向图 G_3 ;

(5)将 G_3 存储为邻接矩阵,其元素包含节点间的依赖关系,并且元素关于主对角线对称,取上三角阵作为定向规则,得到一个初始有向图 G_4 ;

(6)由 G_4 作为初始网络结构,通过爬山算法的搜索策略,基于无环性约束对初始网络结构进行加边、删边、转边的操作,并根据BIC评分函数计算网络评分值,最后返回评分最优的网络结构。

在常用评分函数中,BIC评分因为在精确度和复杂度之间的选择较为均衡,所以经常被用于结构学习。通常来说,BIC评分不仅能挑选出与训练数据集拟合程度高的网络结构,而且可以使得学习得到的网络结构简单。评分函数公式如下^[31]:

$$BIC(G|D)=\sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \theta_{ijk} - \frac{1}{2} \sum_{i=1}^n q_i (r_i - 1) \log m \quad (4)$$

其中, G 是含有 n 个节点的贝叶斯网络结构, D 是给定的样本数据集,样本量为 m ; r_i 表示节点 X_i 的取值状态, q_i 表示节点 X_i 的父节点 $pa(X_i)$ 的取值状态; m_{ijk} 表示节点 X_i 的父节点 $pa(X_i)$ 取第 $j(j=1,2,\dots,q_i)$ 个节点时, X_i 处于第 k 种状态时的样本个数; $\theta_{ijk}=m_{ijk}/m_{ij}$ 为似然条件概率,满足 $0\leq\theta_{ijk}\leq 1$, $\sum_{k=1}^{r_i} \theta_{ijk}=1$ 。

4.2 基于条件互信息的爬山算法举例

图1给出一个构造初始网络结构的过程。如图1(a)所示:对于全连通图中的节点 $\{A,B,C,D,E\}$,计算所有节点对间即 $MI(A,B),MI(A,C),\dots,MI(D,E)$ 的互信息值,并与阈值 α 比较,若互信息值 $MI(X,Y)<\alpha$,则删减对应节点间的边。

如图1(b)所示:倘若互信息值中有 $MI(A,B)<\alpha,MI(A,C)<\alpha,MI(A,D)<\alpha,MI(A,E)<\alpha,MI(C,D)<\alpha$ 和 $MI(D,E)<\alpha$;则将对应节点间的边删除,发现此刻图中产生孤立节点 A 。

如图1(c)所示:将节点 A 与各节点的互信息值进行排序,有 $MI(A,B)>MI(A,C)>MI(A,D)>MI(A,E)$,则将节点 A 与互信息值最大的节点 B 连接,此时图中没有孤立节点,形成完全连接图。

如图1(d)所示:基于图1(c)得到的无向图,找到每个节点的邻居节点并储存,得到 $nbnode(A)=\{B\},nbnode(B)=\{A,C,D,E\},nbnode(C)=\{B,E\},nbnode(D)=\{B\},nbnode(E)=\{B,C\}$ 。计算节点间的条件互信息 $CMI(X,Y|Z)$,使用如下规则:(1)变量 X 可为节点 $\{A,B,C,D,E\}$;(2)在选定变量 X 后的邻居节点集中选取变量 Y ;(3)条件变量 Z 为变量 X 和变量 Y 的邻居节点的交集,并且将条件变量大小限制为1。则对于图1(d)中只需计算条件互信息 $CMI(B,C|E),CMI(B,E|C)$ 和 $CMI(C,E|B)$,图中由标红的边显示。

如图1(e)所示:倘若条件互信息中只有 $CMI(C,E|B)<\alpha$,则删除该边,得到了由互信息和条件互信息

修剪后的无向连接图。

如图 1(f)所示;对图 1(e)中得到的无向图进行定向,定向规则为取邻接矩阵的上三角阵,则将得到的定向图作为初始网络结构用于接下来的结构学习。

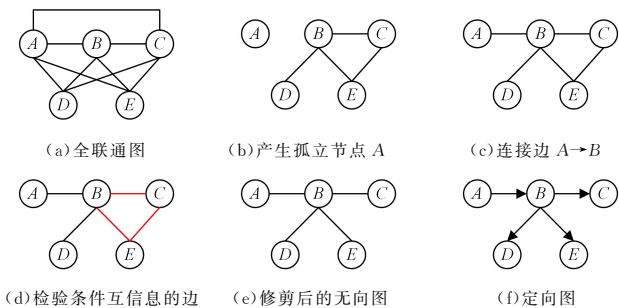


图 1 构造初始网络结构

Fig. 1 Construct initial network structure

4.3 基于条件互信息的爬山算法

算法 1 基于条件互信息的爬山算法伪代码

输入:包含 n 个节点 X_1, X_2, \dots, X_n 的完整数据集 $data$, n 维邻接矩阵

G_0 , 各节点取值范围 ns , 阈值 α

输出:有向无环图 DAG

```

1. MI ← 计算互信息(data); /* 互信息检验 */
2. For {i=1→n};
3. For {j=1→n};
4. If {MI(i,j) < α};
5. G0(i,j)=0;
6. EndIf;
7. EndFor;
8. EndFor;
9. lonelynode ← G0内的孤立节点; /* 消除孤立节点 */
10. For {i=1→length(lonelynode)};
11. lonelyline = MI(lonelynode(:, i), :);
12. lianjiencode = find(lonelyline == max(lonelyline));
13. G0(lonelynode, lianjiencode) = 1;
14. G0(lianjiencode, lonelynode) = 1;
15. EndFor;
16. For {i=1→n}; /* 构造条件集 */
17. nbnode1 = find(G0(i, :));
18. nbnode(i, 1:length(nbnode1)) = nbnode1;
19. EndFor;
20. For {i=1→n-1};
21. For {j=i+1→n};
22. jiaoji1 = intersect(nbnode(i, :), nbnode(j, :));
23. jiaoji(j, 1:length(jiaoji1), i) = jiaoji1;
24. EndFor;
25. EndFor;
26. For {i=1→n-1}; /* 条件互信息检验 */
27. x ← Xi;
28. y ← nbnode(x);
29. z ← jiaoji(x, y);
30. CMI ← 条件互信息(x, y|z);
31. If {CMI < α};
32. G0(x, y) = 0;
33. G0(y, x) = 0;
34. EndIf;
35. EndFor;
36. seeddag ← 取 G0上三角阵作为初始网络结构; /* 对边定向并由

```

爬山算法搜索最优网络结构 */

```

37. temp_score ← BIC(seeddag|data);
38. For {分别对 seeddag 加边、删边、转边获得候选结构 G'};
39. new_score ← BIC(G'|data);
40. If {temp_score < new_score};
41. G ← G';
42. temp_score ← new_score;
43. Else {return(G')};
44. EndIf;
45. EndFor;

```

输入:完整数据集 $data$; G_0 是一个主对角线元素为 0, 其余元素为 1 的 n 维矩阵; 节点取值范围 ns 是评分函数计算需要; 阈值 α 检验互信息和条件互信息。输出: 一个完整的有向无环图 DAG, 即代表得到的最优贝叶斯网络结构。

第 1—8 行是根据互信息值删减图中的边。计算得到节点间的互信息值被存储为一个邻接矩阵, 其索引位置 (i, j) 代表节点 X_i 和节点 X_j 的互信息值。根据其和阈值 α 的比较, 就可以改变代表结构的邻接矩阵 G_0 索引的值, 即若 $MI(i, j) < \alpha$, 则 $G_0(i, j) = 0$, 表现在图结构中就是节点 X_i 和节点 X_j 之间没有边。

第 9—15 行是消除删边过程中产生的孤立节点。将孤立节点存储为一个向量 $lonelynode$, 然后将 $lonelynode$ 与各节点间的互信息值也存储为一个向量 $lonelyline$, 其中最大值的索引 $lianjiencode$ 就是与 $lonelynode$ 有最大互信息的节点, 则将图结构 G_0 中的相应索引位置变为 1, 以此消灭孤立节点。

第 16—25 行是构造邻居节点集 $nbnode$ 和条件集 $jiaoji$ 。将结构 G_0 中的每个节点的邻居节点存储为一个向量 $nbnode_1$, 将向量 $nbnode_1$ 汇总为一个邻居矩阵 $nbnode$, 其中第 i 行就代表着节点 X_i 的邻居节点集。提取邻居矩阵 $nbnode$ 中的各两行的交集形成一个三维矩阵 $jiaoji$, 其中 $jiaoji(j, 1:length(jiaoji_1), i)$ 表示为节点 X_i 与邻居节点 X_j 的条件节点, 并且条件节点集的大小为 $length(jiaoji_1)$ 。

第 26—35 行是进行条件互信息检验削减结构中的边。其中, 变量 x 选择任一节点 X_i , 变量 y 要在节点 X_i 的邻居节点 $nbnode(X_i)$ 中进行选择, 变量 z 要在节点 X_i 和邻居节点 $nbnode(X_i)$ 的交集 $jiaoji(X_i, nbnode(X_i))$ 中选择。利用构造的条件互信息函数计算各节点间的相应值与阈值 α 做比较, 从而将结构 G_0 中的相应索引位置变为 0, 达到削减网络的目的。

第 36—45 行是对削减产生的无向边图进行定向并由爬山搜索算法进行结构学习。由于互信息和条件互信息的对称性, 削减后的结构 G_0 是沿主对角线对称的矩阵, 将该矩阵的上三角阵取为初始网络结构 $seeddag$, 即索引 $(i, j) = 1$, 而 $(j, i) = 0$; 反映到图结构中就是由 $X_i - X_j$ 转变为 $X_i \rightarrow X_j$ 。由 BIC 评分函数计算初始网络结构 $seeddag$ 的评分值, 分别对其进行加边、删边和转边获得候选结构 G' , 将 G' 大于 $seeddag$ 的评分值和结构保留并进行循环搜索, 最后对比出评分最优的网络结构 DAG。

5 实验

5.1 实验配置

实验平台的基本配置: 操作系统 Windows10, 处理器 Intel(R)Core(TM)i5-4210U, 主频 2.40GHz, 内存 8.00GB。实验环境: Matlab R2019a, 使用 FullBNT-1.0.7 工具箱。

为了评价算法的学习性能,选择经典的 Asia 网络、Insurance 网络和 Alarm 网络进行实验,其网络信息如表 1 所列,标准网络结构如图 2—图 4 所示。在仿真实验中,各网络的样本量取 $N=1000,3000,5000$ 。

表 1 标准贝叶斯网络信息

Table 1 Standard Bayesian network information

Bayesian network	the number of nodes	the number of edges	the node sizes
Asia	8	8	2
Insurance	27	52	2,3,4,5
Alarm	37	46	2,3,4

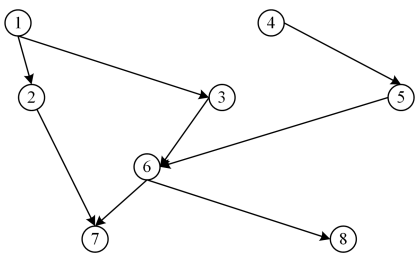


图 2 标准 Asia 网络

Fig. 2 Standard Asia network

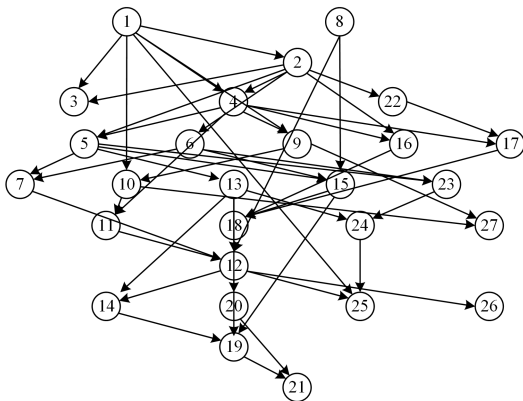


图 3 标准 Insurance 网络

Fig. 3 Standard Insurance network

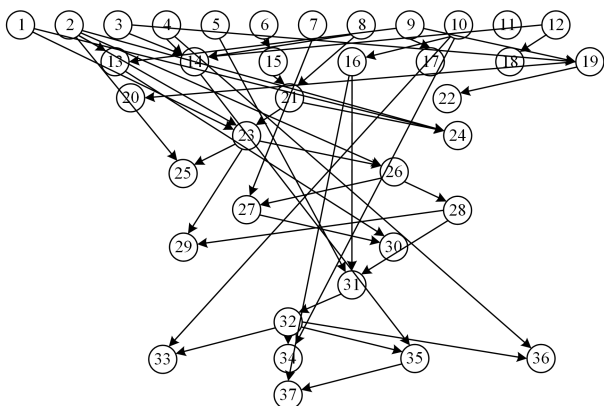


图 4 标准 Alarm 网络

Fig. 4 Standard Alarm network

实验对比算法为:GS 算法、HC 算法、SG 算法和 CMIHC 算法。

衡量网络结构优劣的指标如下。

(1)运行时间 T (Running Time):各个算法学习得到最终网络结构所需的时间,以秒(s)为计时单位。

(2) CE (Correct Edges)表示经结构学习得到的网络与标准网络相比正确的边数, IE (Increased Edges)表示经结构学习得到的网络比标准网络多出的边, ME (Missed Edges)表示缺失的边,而 RE (Reversal Edges)表示方向相反的反转边, GE (Graph Error)表示图错误,其值为 $GE=ME+IE+RE$ 。

(3)准确率 A (Accuracy):学得网络结构中正确的边数 CE 与标准网络的边数的比值。

比较不同规模和数据集下贝叶斯网络结构学习的结果,验证 CMIHC 算法的学习效果。为避免产生随机效应,实验过程中每种样本量随机生成 10 次,取结果的平均值进行衡量。

5.2 实验结果及分析

实验结果如图 5—图 7 所示。

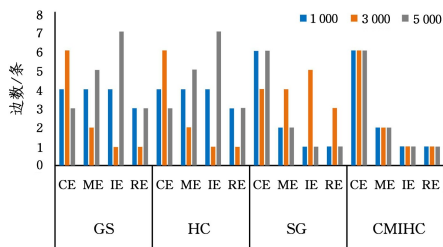


图 5 Asia 网络的学习效果对比

Fig. 5 Comparison of learning effects in Asia network

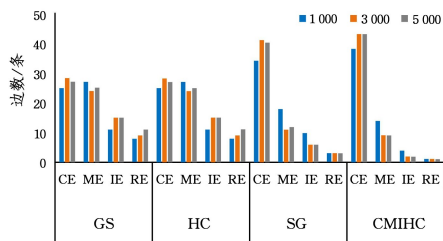


图 6 Insurance 网络的学习效果对比

Fig. 6 Comparison of learning effects in Insurance network

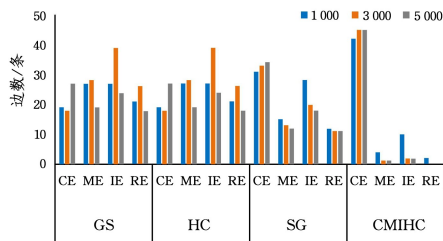


图 7 Alarm 网络的学习效果对比

Fig. 7 Comparison of learning effects in Alarm network

对于评价标准:正确边 CE 和准确率 $Accuracy$ 的值越高代表学习效果越好;运行时间 T 、缺边 ME 、多边 IE 、反转边 RE 和图错误 GE 的值越小代表学习效果越好。

(1)图结构的分析

对于图 5 的 Asia 网络来说,GS 算法和 HC 算法在样本量为 3000 时,学习效果最好, $CE=6,ME=2,IE=RE=1$ 。SG 算法在样本量为 1000 和 3000 时,学习效果最好, $CE=6,ME=2,IE=RE=1$;CMIHC 算法在 3 种样本量下,学习效果均衡, $CE=6,ME=2,IE=RE=1$ 。所以在小型 Asia 网络中,CMIHC 算法稳定且学习效果相对好。

对于图 6 的 Insurance 网络来说,GS 算法和 HC 算法在样本量为 3000 时,学习效果最好, $CE=28,ME=24,IE=15$,

$RE=9$;SG 算法在样本量为 3 000 时,学习效果最好, $CE=41,ME=11,IE=6,RE=3$;CMIHC 算法在样本量为 3 000 和 5 000 时,学习效果最好, $CE=43,ME=9,IE=2,RE=1$ 。所以在稠密的 Insurance 网络中,CMIHC 算法学习效果相对好。

对于图 7 的 Alarm 网络来说,GS 算法和 HC 算法在样本量为 5 000 时,学习效果最好, $CE=27,ME=19,IE=24$,

$RE=18$;SG 算法在样本量为 5 000 时,学习效果最好, $CE=34,ME=12,IE=18,RE=11$;CMIHC 算法在样本量为 3 000 和 5 000 时,学习效果最好, $CE=45,ME=1,IE=2,RE=0$ 。所以在稀疏的 Alarm 网络中,CMIHC 算法学习效果相对好。

(2)运行时间 T 、准确率 $Accuracy$ 和图错误 GE 的分析实验结果如表 2—表 4 所列,其中的数据最优值用黑体显示。

表 2 Asia 网络学习结果对比

Table 2 Comparison of Asia networks learning results

Asia	GS			HC			SG			CMIHC		
	T/s	A	GE	T/s	A	GE	T/s	A	GE	T/s	A	GE
1 000	83.43	0.500	11	68.430	0.500	11	32.704	0.75	4	25.228	0.75	4
3 000	101.15	0.750	4	89.011	0.750	4	71.176	0.50	12	40.332	0.75	4
5 000	156.36	0.375	15	140.137	0.375	15	60.686	0.75	4	47.690	0.75	4

表 3 Insurance 网络学习结果对比

Table 3 Comparison of Insurance networks learning results

Insurance	GS			HC			SG			CMIHC		
	T/s	A	GE	T/s	A	GE	T/s	A	GE	T/s	A	GE
1 000	13 225	0.4808	46	12 885	0.4808	46	8 879.1	0.6538	31	6 542	0.7308	19
3 000	23 297	0.5385	48	22 698	0.5385	48	10 367	0.7885	20	7 015	0.8269	12
5 000	27 726	0.5192	51	26 825	0.5192	51	18 837	0.7692	21	10 666	0.8269	12

表 4 Alarm 网络学习结果对比

Table 4 Comparison of Alarm networks learning results

Asia	GS			HC			SG			CMIHC		
	T/s	A	GE	T/s	A	GE	T/s	A	GE	T/s	A	GE
1 000	44 756	0.4130	75	44 244	0.4130	75	41 168	0.6739	55	18 171	0.9130	16
3 000	77 611	0.3913	93	75 843	0.3913	93	48 745	0.7174	44	13 471	0.9783	3
5 000	157 000	0.5870	61	83 873	0.5870	61	57 086	0.7391	41	17 499	0.9783	3

对于表 2 的 Asia 网络来说:(1)1 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=25.228\text{ s},A=0.75,GE=4$;学习效果最差的是 GS 算法,其 $T=83.43\text{ s},A=0.5,GE=11$ 。(2)3 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=40.322\text{ s},A=0.75,GE=4$;学习效果最差的是 SG 算法,其 $T=71.176\text{ s},A=0.5,GE=12$ 。(3)5 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=47.69\text{ s},A=0.75,GE=4$;学习效果最差的是 GS 算法,其 $T=156.36\text{ s},A=0.375,GE=15$ 。

对于表 3 的 Insurance 网络来说:(1)1 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=6 542\text{ s},A=0.7308,GE=19$;学习效果最差的是 GS 算法,其 $T=13 225\text{ s},A=0.4808,GE=46$ 。(2)3 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=7 015\text{ s},A=0.8269,GE=12$;学习效果最差的是 GS 算法,其 $T=23 297\text{ s},A=0.5385,GE=48$ 。(3)5 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=10 666\text{ s},A=0.8269,GE=12$;学习效果最差的是 GS 算法,其 $T=27 726\text{ s},A=0.5192,GE=51$ 。

对于表 4 的 Alarm 网络来说:(1)1 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=18 171\text{ s},A=0.913,GE=16$;学习效果最差的是 GS 算法,其 $T=44 756\text{ s},A=0.413,GE=75$ 。(2)3 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=13 471\text{ s},A=0.9783,GE=3$;学习效果最差的是 GS 算法,其 $T=77 611\text{ s},A=0.3913,GE=93$ 。(3)5 000 的样本量下学习效果最优的为 CMIHC 算法,其 $T=17 499\text{ s},A=0.9783,GE=3$;学习效果最差的是 GS 算法,其 $T=157 000\text{ s},$

$A=0.587,GE=61$ 。

综上所述,与另外 3 种结构学习算法相比,CMIHC 算法具有更强的学习性能。该算法结合互信息和条件互信息构建初始网络结构,缩减搜索空间,从一定程度上节省了时间成本;在爬山算法中采用 BIC 评分函数,使得学得结构的正确边数多,错误边数少,更接近于标准网络结构,进一步说明了 CMIHC 算法的优越性。

结束语 由于结构学习是 NP-hard 问题,如何高效地学习出准确度高的网络结构一直都是研究目标。CMIHC 算法首先通过互信息和条件互信息对全联通图进行删减,其中利用改进的方法构造条件集,减少条件互信息的检验次数;修剪得到的无向图取上三角阵定向作为初始网络结构;之后在初始网络结构上利用爬山算法,经过加边、删边和转边的局部操作,得到一系列候选网络结构;结合 BIC 评分函数返回得到最优的网络结构。CMIHC 算法结合信息论和评分搜索,给出了一种确定节点条件集的改进方法,基于此进行结构学习缩小搜索空间,提高算法准确度。

但是,仍然存在一些研究问题:对于条件集的构造,虽然减少了条件独立性的检验次数,但也很有可能忽略条件依赖的节点,从而没有添加相应的边。未来的研究方向在于结构学习中对条件集的构造进行改进处理。

参考文献

- [1] ZHANG L W, GUO H P. Introduction to Bayesian Networks [M]. Beijing: Science Publishing House, 2006.
- [2] LOU C Y, LI X S, ATOU I M A. Bayesian network based on an

- adaptive threshold scheme for fault detection and classification [J]. *Industrial and Engineering Chemistry Research*, 2020, 59(34):15155-15164.
- [3] TAGHI-MOLLA A, RABBANI M, KARIMI GAVARESHKI M H, et al. Safety improvement in a gas refinery based on resilience engineering and macro-ergonomics indicators: a Bayesian network-artificial neural network approach [J]. *International Journal of System Assurance Engineering and Management*, 2020, 11(3):641-654.
- [4] WANG H, GUO X D, ZHU H. Research on causal reasoning method of domain ontology based on Bayesian network[J]. *Application Research of Computers*, 2019, 36(3):711-715.
- [5] LU E S, HAN M G, LIU F, et al. Analysis of Chinese medicine syndrome rules of severe pneumonia based on Bayesian network [J]. *Acta Chinese Medicine*, 2022, 37(1):173-179.
- [6] CHENG L X, YAN J N, JIAO Y, et al. Bayesian network method for remote sensing cloud user behavior authentication[J]. *Application Research of Computers*, 2019, 36(2):441-445.
- [7] LI S H, ZHANG J. Review of Bayesian networks structure learning[J]. *Application Research of Computers*, 2015, 32(3):641-646.
- [8] SPIRTE S P, GLYMOUR C, SCHEINES R. *Causality from Probability*[M]. London: Pima, *Evolving Knowledge in Natural Science and Artificial Intelligence*, 1990:181-199.
- [9] SPIRTE S P, GLYMOUR C. An algorithm for fast recovery of sparse causal graphs[J]. *Social Science Computer Review*, 1991, 9(1):62-72.
- [10] CHENG J, GREINER R, KELLY J, et al. Learning Bayesian networks from data: An information-theory based approach[J]. *Artificial Intelligence*, 2002, 137(1/2):43-90.
- [11] XU P F, YANG Z. Accelerating GDS algorithm with L1 penalty for Bayesian network[J]. *Journal of Changchun University of Technology*, 2020, 41(2):118-121.
- [12] XU P F, WANG S D, SHANG L X, et al. Structural learning of Gaussian Bayesian network by bootstrap[J]. *Journal of Changchun University of Technology*, 2020, 41(4):313-321.
- [13] ZHAO J Z, WU C N, WANG X W, et al. Structure Learning Algorithm of Bayesian Networks Based on Markov Blanket[J]. *Journal of Northeastern University(Natural Science)*, 2020, 41(4):464-469, 481.
- [14] HERSKOVITS E. *Computer-based probabilistic-network construction*[D]. Stanford University, 1991.
- [15] HECKERMAN D, GEIGER D, CHICKERING D M. Learning Bayesian networks: The combination of knowledge and statistical data[J]. *Machine learning*, 1995, 20(3):197-243.
- [16] GUO Z, CONSTANTINOU A C. Approximate learning of high dimensional Bayesian network structures via pruning of candidate parent sets[J]. *Entropy*, 2020, 22(10):1142.
- [17] BEHJATI S, BEIGY H. Improved K2 algorithm for Bayesian network structure learning[J]. *Engineering Applications of Artificial Intelligence*, 2020, 91:103617.
- [18] LV Y, MIAO J, LIANG J, et al. BIC-based node order learning for improving Bayesian network structure learning[J]. *Frontiers of Computer Science*, 2021, 15(6):1-14.
- [19] TSAMARDINOS I, BROWN E L, ALIFERIS C F. The maximum hill-climbing Bayesian network structure learning algorithm [J]. *Machine Learning*, 2006, 65(1):31-78.
- [20] LIU X, GAO X, WANG Z, et al. Improved local search with momentum for Bayesian networks structure learning[J]. *Entropy*, 2021, 23(6):750.
- [21] 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, 184:115237.
- [22] WANG Y, TAN S Q, LIU Y H. Bayesian network structural learning algorithm based on mutual information[J]. *Computer Engineering*, 2011, 37(7):62-64.
- [23] LI B H, GAO X L, LIU S Y, et al. Learning Bayesian network structures based on mutual information[J]. *CAAI Transactions on Intelligent Systems*, 2011, 6(1):68-72.
- [24] JIN Y, HU Y A, ZHANG J, et al. Bayesian network structure learning combining mutual information with hill climbing algorithm[J]. *Computer Applications and Software*, 2012, 29(9):122-125.
- [25] WU Y G, PANG S C. K2&HC Structure learning algorithm[J]. *Computer and Digital Engineering*, 2014, 42(7):1137-1140, 1145.
- [26] LIU H R, LI X, MA M, et al. Simplified Greedy Algorithm for Bayesian Network Structure Learning[J]. *Journal of Chinese Computer Systems*, 2015, 36(2):306-309.
- [27] LIU B, LIU Y J, LIU H R, et al. A study of fault diagnosis model of rotary kiln based on improved structural algorithm of Bayesian [J]. *China Mechanical Engineering*, 2017, 28(18):2143-2151.
- [28] LONG Y G. *Learning Bayesian network structure basing on Information theory and causal effect*[D]. Changchun: Jilin University, 2020
- [29] GAO X G, WANG C F, DI R H. A block learning algorithm with improved K-means algorithm for learning sparse BN optimal structure[J]. *Acta Automatica Sinica*, 2020, 46(5):923-933.
- [30] WU B W. *Hybrid uncertainty Bayesian networks learning model and implementation by R*[D]. Guangzhou: South China University of Technology, 2013.
- [31] ZHAO J Z, WU C N, WANG X W, et al. Structure learning algorithm of Bayesian networks based on markov blanket[J]. *Journal of Northeastern University(Natural Science)*, 2020, 41(4):464-469, 481.



LI Xiaoqing, born in 1996, postgraduate. Her main research interests include Bayesian network structure learning and so on.



YU Haizheng, born in 1976, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include big data analysis and so on.