计算机科学
COMPUTER SCIENCE

一种基于两步搜索策略的K2改进算法

徐苗, 王慧玲, 梁义, 綦小龙, 高阳

引用本文

徐苗, 王慧玲, 梁义, 綦小龙, 高阳. 一种基于两步搜索策略的K2改进算法[J]. 计算机科学, 2023, 50(9): 303-310.

XU Miao, WANG Huiling, LIANG Yi, QI Xiaolong, GAO Yang. Improved K2 Algorithm Based on Two-step Search Strategy [J]. Computer Science, 2023, 50(9): 303-310.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220700253

一种基于两步搜索策略的 K2 改进算法

徐 苗[1] 王慧玲[1,2] 梁 义[1] 綦小龙[1] 高 阳[2]
1 伊犁师范大学网络安全与信息技术学院 新疆 伊宁 835000
2 南京大学计算机科学与技术系计算机软件国家重点实验室 南京 210023
(xm_1192@163.com)

摘 要 贝叶斯网络由于其强大的不确定性推理能力和因果可表示性越来越受到研究者的关注。从数据中学习一个贝叶斯网络结构被称为 NP-hard 问题。其中,针对 K2 算法强依赖于变量拓扑序的问题,提出了一种组合变量邻居集和 v-结构信息的 K2 改进学习方法 TSK2(Two-Step Search Strategy of K2)。该方法有效减小了序空间搜索规模,同时避免了过早陷入局部最优。具体而言,该方法在约束算法定向规则的启示下,借助识别的 v-结构和邻居集信息可靠调整汇点的邻居在序中的位置;其次,在贝网基本组成结构的启发下,借助变量邻居集信息,通过执行顺连、分连、汇连 3 个基本结构的搜索,准确修正父节点与子节点的序位置,获得最优序列。实验结果表明,在 Asia 和 Alarm 网络数据集上,与对比方法相比,所提算法的准确率得到显著提升,可以获得更准确的网络结构。

关键词:K2 算法;PC 算法;v-结构;邻居集;结构学习
中图法分类号 TP181

Improved K2 Algorithm Based on Two-step Search Strategy

XU Miao[1],WANG Huiling[1,2],LIANG Yi[1],QI Xiaolong[1] and GAO Yang[2]
1 School of Cybersecurity & Information Technology,Yili Normal University,Yining,Xinjiang 835000,China
2 State Key Laboratory for Novel Software Technology,Department of Computer Science and Technology,Nanjing University,Nanjing 210023,China

Abstract Bayesian network receiving increasing attention from researchers because of its strong uncertainty reasoning ability and causal representability. Learning a Bayesian network structure from data is an NP-hard problem. Among them,for the problem that the K2 algorithm strongly depends on the topological order of variables,an improved K2 learning method TSK2 is proposed,which combines variable neighbor sets and v-structure information. The proposed method effectively reduces the search scale in the order space and avoids prematurely falling into local optimum. Specifically,inspired by the orientation rules of the constraint algorithm,the method reliably adjusts the in-order position of the neighbors of the sink with the help of the identified v-structure and neighbor set information. Secondly,inspired by the basic structure of the shell net,with the help of the variable neighbor set information,the optimal sequence is obtained by performing the search of the three basic structures of shun-connection,sub-connection,and confluence-connection to accurately correct the order positions of parent nodes and child nodes. Experimental results show that the accuracy of the proposed algorithm is significantly improved compared with the comparison methods on the Asia and Alarm network datasets. A more accurate network structure can be learned.

Keywords K2 algorithm,PC algorithm,v-structure,Neighbor set,Structure learning

1 引言

贝叶斯网络是由概率推理和图论理论结合而成的图形化模型,通过条件概率表和有向无圈图定量地描述了变量之间的独立关系和依赖程度,在处理不确定性知识和数据分析方面具备独特的优势,是人工智能领域的有力工具。由于贝叶

到稿日期:2022-07-26 返修日期:2022-11-09
基金项目:新疆维吾尔自治区自然科学基金(2021D01C467);伊犁师范大学博士科研启动项目(2020YSBS007);学实高层次人才岗位(YSXSQN22007)
This work was supported by the Natural Science Foundation of Xinjiang Uygur Autonomous Region,China(2021D01C467),Doctoral Research Start-up Project of Yili Normal University(2020YSBS007) and Promising Scholar of Academic Integrity(YSXSQN22007).
通信作者:綦小龙(qxl_0712@sina.com)

斯网络具有坚实的数学基础、灵巧的学习体系和直观形象的语义表述等特点，引起了国内外众多研究学者的极大兴趣。近几十年来，贝叶斯网络成为了研究热点，并在医疗诊断、机器视觉、因果推断、工业控制、学习预测和数据挖掘等领域[1-5]得到了广泛应用。

贝叶斯网络最早主要是由领域专家的先验知识进行构建，在网络规模较小时可行，但是面对较大网络规模时，变量间的因果关系变得十分复杂，这时再依靠领域专家来构建贝叶斯网络会非常困难且容易出错。如何从已知数据中学习贝叶斯网络已成为现阶段的研究方向，这是一个 NP 难问题[6]。目前，结构学习算法主要分为：基于约束的学习算法、基于评分-搜索的学习算法和混合学习算法。

约束的学习算法通过信息论进行条件的独立性检测，判断变量间关系从而确定网络结构，例如 TPDA 算法[7]和 SGS 算法[8]等，它们具备严格的理论基础，学习效率较高，但依赖于独立检测的准确度，在多维数据或者数据不足时模型学习精度不可靠。基于评分-搜索的学习算法通过评分函数对网络模型结构的优劣进行评判，利用搜索策略寻找最佳评分的模型结构，例如爬山法[9]和粒子群算法[10]等，它们可以学习到较精确的结构模型，但在复杂结构场景下搜索空间过大，导致搜索效率大幅降低。混合学习算法将上述两种算法结合，首先利用基于约束的学习算法降低网络空间的搜索规模，然后执行评分-搜索的学习算法寻找最佳模型结构。例如 MMACO 算法[11]，其步骤分为两步：先通过 MMPC 算法获得基本模型，再使用蚁群算法确定基本模型的边和方向。混合学习算法已成为现阶段受欢迎的研究方向之一[12]。

基于评分-搜索的 K2 算法[13]将假设的随机变量顺序作为输入，利用变量拓扑序列约束搜索过程，在有效减小搜索空间的同时还能降低非法结构的概率，整体表现优于大多数经典算法。因此，K2 算法的效率很大程度上依赖于变量顺序，不准确的变量顺序无法学习到正确的结构模型。如何获得准确的变量顺序是使 K2 算法更加实用的主要挑战，本文提出的充分利用 v-结构和邻居集信息的混合学习算法为此提供了新的思路。

本文的主要贡献如下：

1）提出了基于 v-结构与汇点邻居集信息的序搜索策略。在学习过程中，使用该信息，无需执行评分搜索计算，直接对初始序列进行调整，实现部分节点定序，在减少计算量的同时也合理约减了序搜索空间。

2）受到贝叶斯网络结构的 3 个基本组成成分启发，提出了基于邻居集置换的序搜索策略。对未定序节点进行 3 种置换搜索，可靠地从给定的数据中学习出高质量顺序，有效避免了过早陷入局部最优，实现全局搜索。

3）在标准数据集上对所提算法 TSK2 进行实验对比分析，以评估其学习性能。实验结果表明了 TSK2 算法的有效性。

2 相关工作

近年来，针对 K2 算法需要准确变量拓扑序的问题，研究者们相继给出了很多解决方案。文献[14]提出了基于 MWST-K2 的算法，通过 MWST 算法构建有向生成树，将生成的排序作为 K2 算法的先验拓扑序列，由于 MWST 算法的定向精度较低，该算法在构建基准网络模型时结果不理想。文献[15]提出了基于因果效应的算法，该算法通过改进版因果效应法和 BDe 评分获得节点优先次序，利用删除边操作和反向调节修正结果，提升了网络结构学习性能，但是面对大型网络时易陷入局部最优。文献[16]提出了 K2ACO 算法，通过种群中的初始个体随机创建节点顺序，由蚁群算法进行优化学习得到最佳序列，在蚁群算法过程中，运行 K2 搜索算法计算各排序的适应度，寻找最优排序并直接获得相对应的网络结构，该算法存在时间复杂度较高的问题。文献[17]提出了 ITNO-K2PC 算法，利用信息理论测试节点间的依赖关系，通过确定父子节点关系对生成的矩阵进行验证，获得最终结构模型，该算法更适用于中小规模网络。文献[18]提出了优先级排序算法，通过 MCMC 算法[19]学习拟合观测值的结构模型，利用定义的节点排序评分函数获得变量先验序，该算法减少了穷举所需花费的较长时间，但受到 MCMC 算法的约束。文献[20]提出了重构先验信息的算法，将 MCMC 算法与模拟退火算法相结合，根据邻域查找规则配合专家知识和数据修补等方法获得最佳网络结构，该算法在小数据集上表现较好，但是需要具备完整且准确的领域知识，在很多场景下难以实现。文献[21]提出了 PSOK2 算法，利用粒子群算法更新变量拓扑序，通过 K2 算法构建贝叶斯网络，使用 AIC 评分函数进行判定取得对应拓扑序列的适应度函数并进行种群迭代，该算法的运行速度较慢，学习效率较低。文献[22]提出了基于强连通的改进算法，利用每个变量的最佳父节点集来构建强连通图，从强连通分量中推断变量的顺序，该算法降低了时间复杂度，但最佳父集的准确性对算法结果影响较大。

3 算法的构建

TSK2 算法分为 4 步：首先利用改进版因果效应法学习节点优先排序，将其作为初始顺序；然后通过 PC 算法获得 v-结构和邻居集，使用 v-结构与汇点邻居集信息对初始序列进行调整，修正不符合父子节点关系的节点顺序；接着利用邻居集置换策略，对邻居节点进行置换搜索，进一步修正节点顺序获得最佳节点序；最后利用 K2 算法学习高质量贝叶斯网络结构模型。通过 v-结构和邻居集信息的配合使用，能够快速而有效地检测出错误节点顺序并做出修正。

3.1 节点优先排序

利用改进版因果效应法学习节点优先排序，可以定性地表达节点之间的父子关系。对于二值数据网络和多值数据网络，分别从两种方法开始，得到所有节点的优先度并降序排列，获得节点优先排序。

3.1.1 二值数据网络

Pearl 提出的因果理论仅考虑 X 和 Y 在 $Y=1$ 处的因果效应，改进版因果效应法[15]在此基础上进行了拓展，对于任意两个节点 X_i 和 X_j，考虑 $X_i \rightarrow X_j$ 在 $X_j=1$ 和 $X_j=0$ 两处的因果效应，因果效应 $CE_{X_i \rightarrow X_j}$ 计算式如下：

$$CE_{X_i\to X_j}=\frac{N(X_j=1)}{N}\times[P(X_j=1|X_i=1)-P(X_j=1|X_i=0)]-\frac{N(X_j=0)}{N}\times[P(X_j=0|X_i=1)-P(X_j=0|X_i=0)] \tag{1}$$

其中，N 表示总样本数目，$N(X_j=1)$表示 $X_j=1$ 的样本数目，$N(X_j=0)$表示 $X_j=0$ 的样本数目。

算法从一个节点开始，计算任意两个节点之间的因果效应，直至遍历网络中的所有节点。对于 X_i 和 X_j，如果 $CE_{X_i\to X_j}>CE_{X_j\to X_i}$，那么将 X_i 的优先度加 1；如果 $CE_{X_i\to X_j}<CE_{X_j\to X_i}$，那么将 X_j 的优先度加 1。对最终的节点优先度进行降序排列，获得节点优先排序，将其作为初始序列。

3.1.2　多值数据网络

对于多值数据网络，搜索每个节点最多的两个状态值，使用相对应的值计算任意两个节点的因果效应。对于任意两个节点 X_i 和 X_j，若 X_i 的最多状态值是 a 与 b，X_j 的最多状态值是 c 与 d，则需要考虑 $X_i\to X_j$ 在 $X_j=c$ 和 $X_j=d$ 两处的因果效应，因果效应 $CE_{X_i\to X_j}$ 计算式如下：

$$CE_{X_i\to X_j}=\frac{N(X_j=c)}{N}\times[P(X_j=c|X_i=a)-P(X_j=c|X_i=b)]-\frac{N(X_j=d)}{N}\times[P(X_j=d|X_i=a)-P(X_j=d|X_i=b)] \tag{2}$$

其中，$N(X_j=c)$表示 $X_j=c$ 的样本数目，$N(X_j=d)$表示 $X_j=d$ 的样本数目。

多值数据网络方法与二值数据网络方法的规则相同，依次计算每两个节点之间的因果效应，获得初始序列。

3.2　基于 PC 算法的节点序搜索策略

3.2.1　PC 算法

PC 算法[8]是 Spirtes 等提出的基于约束的学习算法，是对 SGS 算法的进一步改良。PC 算法不需要完全计算指数级别下的条件独立性，它从空图开始，确定节点间的依赖关系得到无向图，然后再确定节点间的依赖方向，把无向图学习变为部分有向无圈图。其中，变量的邻居集决定了变量对的条件集，算法在遍历过程中会动态更新变量的邻居集。

文献[23]中的定理 3.3 阐明了 PC 算法可以输出体现有向无圈图 G 的 CPDAG，即 PC 算法可以输出准确的 v-结构。该算法利用 PC 算法学习的 v-结构和邻居集对最佳变量拓扑序进行搜索，达到减小序列搜索空间规模、准确判定序列正确性的目的。

3.2.2　节点序搜索策略

改进版因果效应法删除了“do-操作”，所以获得的初始序列是一个近似正确的序列[15]，仍可能存在不正确的父子序列顺序。

考虑到之前的算法，如贪婪爬山法，在调整节点序时，节点的移动是随机的，没有结合额外的信息，学习效果不理想。为了提高序列的准确度，学习最优网络结构，本文配合使用两种节点序搜索策略：基于 v-结构与汇点邻居集的序搜索策略和基于邻居集置换的序搜索策略。

情况 1　v-结构信息可用时，采用基于 v-结构与汇点邻居集的序搜索策略。

定义 1(邻居)　如果节点 X_i 和节点 X_j 之间存在一条边，则节点 X_i 是节点 X_j 的邻居，反之亦然，记作 $Adj(X_i)=\{X_j\}$。

对于三元组$\langle X_i,X_j,X_k\rangle$，如果存在 $X_i\to X_j$，且 $X_k\to X_j$，即 X_i 与 X_k 都存在指向 X_j 的有向边，构成 $X_i\to X_j\leftarrow X_k$ 的形式，就称其为 v-结构。如图 1 所示，三元组$\langle X_i,X_j,X_k\rangle$构成了 v-结构，其中 X_i 和 X_k 作为父节点，X_j 作为子节点，v-结构清晰而直观地展现了 3 个节点之间的父子关系。在节点序的排列中，父节点必须排在子节点之前，所以根据 v-结构的信息可以确定，节点 X_i 和节点 X_k 的序列位置需要排在节点 X_j 之前。

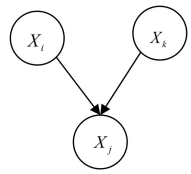


图 1　v-结构

Fig. 1　v-structure

基于 v-结构的特性，将学习到的相关信息作用于初始序列的调整，对可能存在的不正确排序进行更新。具体而言，对于任意 3 个节点构成的 v-结构 $X_i\to X_j\leftarrow X_k$，搜索 X_i，X_j 和 X_k 在节点序中的位置，检测 X_i 和 X_k 是否排在 X_j 之前，若未满足，则将 X_i 和 X_k 移动到 X_j 之前，以满足父节点在前、子节点在后的准则；接着根据节点 X_j 的邻居信息，直接将位于 X_j 前方的其余邻居节点置于 X_j 后方。示例如图 2 所示，节点优先序列为(X_1,X_2,X_3,X_4,X_5)，邻居信息为 $Adj(X_3)=\{X_4,X_5,X_2\}$，v-结构信息为 $X_4\to X_3\leftarrow X_5$，根据调整策略得到新节点序列(X_1,X_4,X_5,X_3,X_2)。

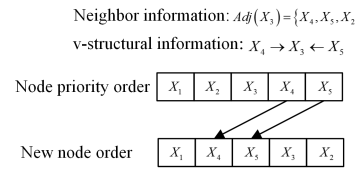


图 2　基于 v-结构与汇点邻居集的序搜索策略

Fig. 2　Search strategy for v-structure and sink neighbor set

基于 v-结构与汇点邻居集的序调整过程中，存在两种情况：

1）节点入度为 2 时，PC 算法会获得一个 v-结构信息 $X_i\to X_j\leftarrow X_k$。基于 v-结构信息和约束算法的定向准则，X_j 不存在第三个父节点。因此，节点 X_j 的其余邻居节点可以直接确定为它的子节点，换言之，这些邻居节点应该在节点 X_j 之后，如图 3(a)所示。

2）节点入度大于 2 时，PC 算法会获得关于节点 X_j 的多个 v-结构，如图 3(b1)所示。对于每一个 v-结构，依然根据 v-结构信息和约束算法的定向准则，调整 X_j 其余邻居节点的序，如图 3(b2)所示。节点 X_j 其余邻居集的计算式如下：

$$R_{\mathrm{Adj}}(X_j)=Adj(X_j)\backslash V_{\mathrm{Adj}}(X_j) \tag{3}$$

其中，$V_{\mathrm{Adj}}(X_j)$表示由以节点 X_j 为汇点的 v-结构中节点 X_j 的父集所构成的邻居集。

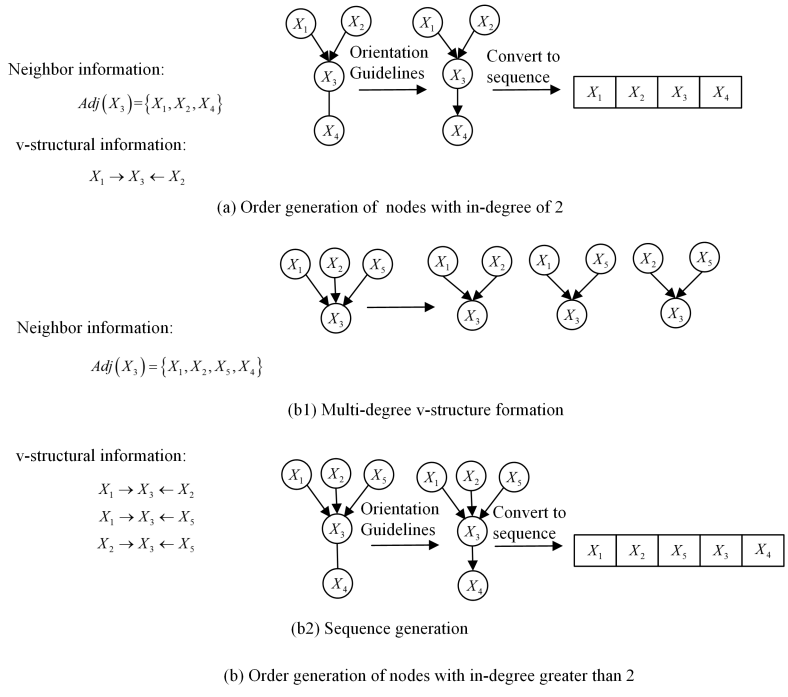


图 3 基于 v-结构与汇点邻居集的序调整情况

Fig. 3 Order adjustment of v-structure and sink neighbor set

命题 1 基于 v-结构与汇点邻居集的序搜索策略是可靠的。

证明：不失一般性，假设通过 PC 算法的学习得到 v-结构信息 $X_i \to X_j \leftarrow X_k$，$X_i \to X_j \leftarrow X_l$，$X_k \to X_j \leftarrow X_l$ 与 X_j 的邻居信息 $Adj(X_j)=\{X_i,X_k,X_l,X_m\}$。如果定向 $X_m \to X_j$，那么 X_i 和 X_m、X_k 和 X_m、X_l 和 X_m 的分隔集中一定不包含 X_j，这与 PC 算法获取的 X_i 和 X_m、X_k 和 X_m、X_l 和 X_m 的分隔集矛盾，即 $X_j \in sep\{X_i,X_m\}$，$X_j \in sep\{X_k,X_m\}$，$X_j \in sep\{X_l,X_m\}$。

利用 v-结构与汇点邻居集信息调整得到部分最佳节点序，由于部分无 v-结构信息的节点未被检测，仍存在错误节点序未被修正。受到组成贝叶斯网络结构的 3 个基本成分顺连、分连、汇连的启发，进一步提出了基于邻居集置换的序搜索策略。

情况 2 无 v-结构信息可用时，采用基于邻居集置换的序搜索策略。

定义 2（分连、顺连、汇连[9]） 节点 X_i 与节点 X_k 通过节点 X_j 相连，分为 3 种情况：

1）对于分连 $X_i \leftarrow X_j \to X_k$，若变量 X_j 已知，则信息传递被阻塞，此时 X_i 与 X_k 相互独立；若变量 X_j 未知，X_i 和 X_k 之间能进行信息传递，此时 X_i 与 X_k 相互依赖。

2）对于顺连 $X_i \to X_j \to X_k$，与分连情况相似，若变量 X_j 已知，则信息传递被阻塞，此时 X_i 与 X_k 相互独立；若变量 X_j 未知，X_i 和 X_k 之间能进行信息传递，此时 X_i 与 X_k 相互依赖。

3）对于汇连 $X_i \to X_j \leftarrow X_k$，若变量 X_j 未知，则信息传递被阻塞，此时 X_i 与 X_k 相互独立；若变量 X_j 已知，X_i 和 X_k 之间能进行信息传递，此时 X_i 与 X_k 相互依赖。

在调整节点位置时，该策略有效结合当前节点的邻居集，通过执行顺连、分连、汇连等基本结构的搜索，修正错误父节点与子节点的序位置，提高了学习效率和序质量。

具体而言，对于未确定位置的节点 X_i，如果存在 $Adj(X_i)=\{X_j,X_k\}$，那么根据定义 2，通过调整 X_i 关于邻居节点 X_j 和 X_k 的位置来搜索节点序列。

1）节点 X_i 位于其邻居节点 X_j 和 X_k 前方，作为节点 X_j 和 X_k 的父节点。

2）节点 X_i 位于其邻居节点 X_j 的后方、邻居节点 X_k 的前方，作为节点 X_j 的子节点、节点 X_k 的父节点。

3）节点 X_i 位于其邻居节点 X_j 和 X_k 后方，作为节点 X_j 和 X_k 的子节点。

执行示例如图 4 所示，节点优先序列为 (X_1,X_2,X_3,X_4,X_5)，邻居信息为 $Adj(X_1)=\{X_2,X_4\}$，利用邻居集置换的序搜索策略，生成节点 X_1 与其邻居节点 X_2 和 X_4 的分连、顺连和汇连 3 种形式，即新序列 (X_1,X_2,X_3,X_4,X_5)，(X_2,X_1,X_3,X_4,X_5)，(X_2,X_3,X_4,X_1,X_5)，采用 BD 评分函数对 3 个新序列对进行判定，选择最佳序列。在样本量较少或者节点依赖关系较弱时，可能存在某个节点没有邻居的情况，这时，对于无邻居的节点 X_i，根据 BD 评分判定，将其放在序列首位或者保持不动。判定过程中，更新评分最优的序列，直至将所有节点的邻居集遍历完成，得到最终的优质节点序。TSK2 算法如算法 1 所示。

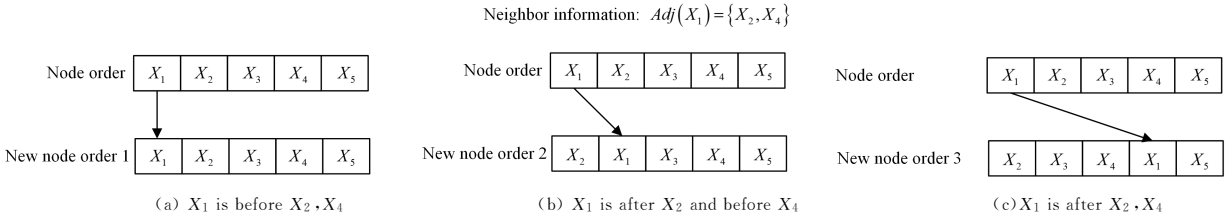
算法 1 TSK2 算法

输入：节点集 N，数据集 D，父节点上限数 u，v-结构集 v，邻居集 b

输出：一个贝叶斯网络

```
1. for i=1:n
2.     for j=1:n
3.         按照式(1)或式(2)计算每两条边的因果效应;
4.         if CE_{X_i→X_j} > CE_{X_j→X_i}
```

5.　　$d_{X_i}=d_{X_i}+1$；
6.　　else $d_{X_j}=d_{X_j}+1$
7.　　end{if}；
8.　end{for}；
9. end{for}；
10. 将每个节点按照优先度降序排序，获得初始序列 order；
11. while $X_i,X_j,X_k\in v$ & $Adj(X_j)=\{X_i,X_k,X_l\}$
12.　　if $X_i\rightarrow X_j\leftarrow X_k$ = false
13.　　将 X_i,X_k 置于 X_j 前方；
14.　　end{if}；
15.　　将 X_l 置于 X_j 后方；
16. end{while}；
17. while $Adj(X_o)=\{X_m,X_n\}\in b$ & $X_o,X_m,X_n\notin v$
18.　根据调整策略得到评分 $Score_1,Score_2,Score_3$；
19.　$max=max(Score_1,Score_2,Score_3)$；
20. 更新最大评分 order；
21. end{while}；
22. 返回 order；
23. 将 order 作为 K2 算法的节点顺序；
24. for i=1:n
25.　　$\pi_i=\emptyset$；
26.　　计算出 X_i 的评分 P_{old}；
27.　　令 OKTOProceed 为真；
28.　　while OKTOProceed and $|\pi_i|<u$ do
29.　　找出排在 X_i 前面的变量 X_j，得到最大新评分 P_{new}；
30.　　if $P_{new}>P_{old}$ then
31.　　$P_{old}=P_{new}$；
32.　　$\pi_i=\pi_i\cup\{j\}$；
33.　　else OKTOProceed 为假
34.　　end{if}；
35.　end{while}；
36. end{for}；



(a) X_1 is before X_2, X_4　(b) X_1 is after X_2 and before X_4　(c) X_1 is after X_2, X_4

图 4　基于邻居集置换的序搜索策略

Fig. 4　Sequence search strategy based on neighbor set permutation

3.3　复杂度分析

3.3.1　计算复杂度

提议的 TSK2 算法的计算复杂度包含两部分：变量序搜索和贝网结构学习。假设 n 表示节点数目，v 表示 v-结构数目，p 表示未定位的节点数目。任意两个节点间的因果效应计算复杂度为 $O(n^2)$；得到初始序后，基于 v-结构与汇点邻居集的序搜索策略需要完成寻找节点和移动节点两个过程，计算复杂度为 $O(n^2v^2)$；基于邻居集置换的序搜索策略需要对其余节点进行置换操作，计算复杂度为 $O(n^2p^2)$；获得高质量序后，使用 K2 算法学习最佳贝叶斯网络结构，其计算复杂度为 $O(2^{\frac{n(n-1)}{2}})$。因此，TSK2 算法的计算复杂度为 $O(n^2)+O(n^2v^2)+O(n^2p^2)+O(2^{\frac{n(n-1)}{2}})$。

显然，与 K2 算法的计算复杂度 $O(2^{\frac{n(n-1)}{2}})$ 相比，TSK2 算法的计算复杂度较高，但其能够自动从已知数据中学习到高质量的先验序列，显著提升了 K2 算法的学习精度。

3.3.2　空间复杂度

算法在开始执行时，需要存储节点间因果效应以获得初始序，其空间复杂度为 $O(n^2)$，获得先验序后，该空间被释放；基于 v-结构与汇点邻居集的序搜索策略和基于邻居集置换的序搜索策略主要完成变量在初始序中的移动，其空间复杂度为 O(1)；获得高质量序后，调用 K2 算法。该过程需要存储每个变量的最佳父集，其空间复杂度为 $O(nk)$，其中 k 表示最大父节点数，最坏情况下即 $k=n-1$，空间复杂度为 $O(n^2)$。这与 K2 算法的空间复杂度相同。

4　实验结果

为了验证算法的优质性能，使用标准 Asia 网络和 Alarm 网络进行仿真实验。其中，Asia 网络主要应用于医疗诊断系统，利用式(1)学习节点初始序列；Alarm 网络主要应用于医疗监控系统，利用式(2)学习节点初始序列。标准网络的信息如表 1 所列。实验运行环境为：Windows 7 操作系统，处理器 Intel(R) CoreTM i5-4210M CPU，内存 8 GB，Matlab R2016a 软件平台。

表 1　标准贝叶斯网络参数

Table 1　Standard Bayesian network parameters

	Asia	Alarm
Number of variables	8	37
Number of edges	8	46
Variable state	2	2,3,4
Max in-degree	2	4
Max out-degree	2	5

Asia 网络实验使用 2000，4000，6000，8000，10000，12000，14000，16000，18000，20000 的样本量数据进行分析，Alarm 网络使用 2000，4000，6000，8000，10000 的样本量数据进行分析，检测算法对不同样本量的敏感程度，并与基于因果效应的方法(Causal Effect，CE)[15]、MMHC 算法(Max-Min Hill-Climbing，MMHC)[24]、MCMC 算法(Markov Chain Monte Carlo，MCMC)[19] 和爬山法(Hill Climbing，HC)[9] 相对应的项目平均值和标准差进行比较，比较的项目如下：

正确边(Correct side)：与标准网络相比，算法学习到的正确边数。

有余边(Extra edge):与标准网络相比,算法学习到的多余边数。

遗失边(Missing edge):与标准网络相比,算法未能学习到的边数。

相反边(Opposite edge):与标准网络相比,算法学习到的方向相反边数。

差错边(Wrong edge):与标准网络相比,算法学习到的所有不正确边数,即有余边、遗失边和相反边的数量之和。

实验分析结果如表2、表3和图5—图10所示。

表2 Asia网络实验分析

Table 2 Experiment analysis of Asia network

project	TSK2 $\mu\pm\sigma$	CE $\mu\pm\sigma$	MMHC $\mu\pm\sigma$	MCMC $\mu\pm\sigma$	HC $\mu\pm\sigma$
Correct side	7.8±0.40	7.0±0.94	6.3±0.82	3.9±2.18	3.75±1.75
Extra edge	0.2±0.40	0.7±1.06	0.6±0.84	4.0±1.70	1.68±1.31
Missing edge	0.2±0.40	0.1±0.32	0.5±0.71	0.5±0.71	0.96±0.65
Opposite edge	0.0±0.00	0.8±0.79	1.2±0.63	3.5±1.78	3.29±1.75
Wrong edge	0.4±0.49	1.5±1.51	2.3±1.26	8.0±3.50	5.93±2.81

表3 Alarm网络实验分析

Table 3 Experiment analysis of Alarm network

project	TSK2 $\mu\pm\sigma$	CE $\mu\pm\sigma$	MMHC $\mu\pm\sigma$	MCMC $\mu\pm\sigma$	HC $\mu\pm\sigma$
Correct side	43.2±1.83	41.2±3.08	34.1±6.05	20.0±3.56	21.0±4.24
Extra edge	5.2±3.66	5.5±2.32	4.1±1.45	44.5±8.03	12.0±2.82
Missing edge	2.4±1.20	0.5±0.53	0.3±0.48	3.5±2.11	2.5±0.71
Opposite edge	0.4±0.80	4.2±2.53	11.6±6.19	22.5±2.83	22.5±4.94
Wrong edge	8.0±5.48	10.2±4.80	16.0±6.93	70.5±10.54	37.0±7.07

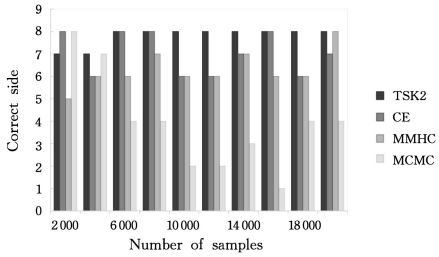


图5 Asia网络正确边数对比

Fig. 5 Comparison of correct sides in Asia network

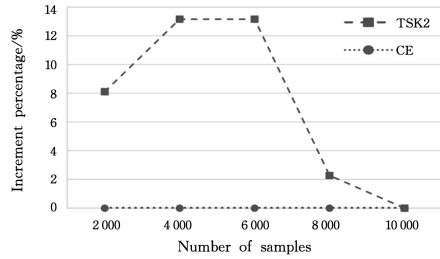


图8 Alarm网络正确边增量对比

Fig. 8 Comparison of increment in Alarm network

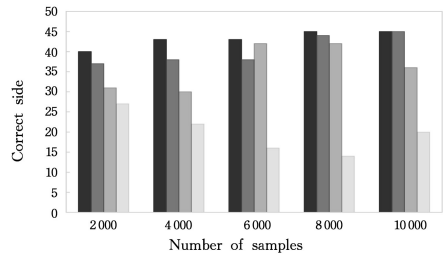


图6 Alarm网络正确边数对比

Fig. 6 Comparison of correct sides in Alarm network

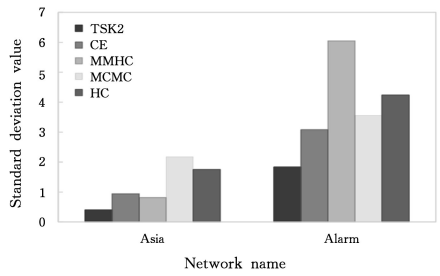


图9 不同算法标准差对比

Fig. 9 Comparison of SD of different algorithms

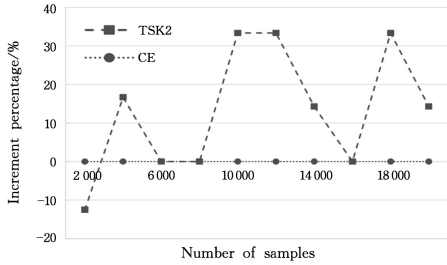


图7 Asia网络正确边增量对比

Fig. 7 Comparison of increment in Asia network

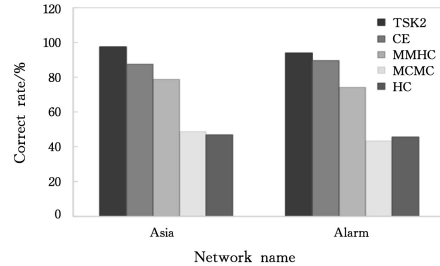


图10 不同算法正确率对比

Fig. 10 Comparison of accuracy of different algorithms

由表 2、表 3 可知，在 Asia 网络的 10 组样本实验中，TSK2 算法平均可以学习到 7.8 条正确边数，而基于因果效应的方法平均只能学习到 7.0 条正确边数；在 Alarm 网络的 5 组样本实验中，TSK2 算法平均学习到的正确边数目达到 43.2 条，而基于因果效应方法仅为 41.2 条。TSK2 算法较基于因果效应方法的正确边数平均提升了 8%，并且在相反边和有余边方面的性能也有优势，虽然缺失边略多于基于因果效应的方法，但总体性能表现在对比算法中最优。原因在于学习到的 v-结构与汇点邻居集信息可以准确修正错误汇点邻居顺序，再配合邻居集置换策略更新剩余不正确顺序，有效保证了父节点在其子节点之前，使得产生错误边的概率减小。

由图 5、图 6 可知，在 Asia 网络的学习中，TSK2 算法最少也可以学习到 7 条正确边，仅比标准网络结构少一条，并且在 6000 个样本时达到稳定，之后均可以学习到 8 条正确边。在 Alarm 网络的学习中，TSK2 算法通过 2000 个样本学习到 40 条正确边，在样本量较小时能够获得比其他算法更准确的网络结构，主要原因是 v-结构和邻居集信息的两步序搜索策略的结合使用，使其更容易获得全局最优的节点序列，从而学习到高质量的网络结构。

由图 7、图 8 可知，与对比算法中最优的基于因果效应的方法相比，无论是在 Asia 网络还是 Alarm 网络，TSK2 算法增量都很明显。相比基于因果效应方法采用的先学习网络，再通过反向调节和删边策略修正非法结构的方式，TSK2 算法直接对作为源头的先验节点序列进行更新的方式更灵活。

由图 9 可知，Asia 作为小型网络，需要学习的边数少，所以算法的正确边标准差波动不大；Alarm 作为复杂网络，需要学习的边数较多，容易陷入局部最优，所以算法的正确边标准差波动较大。从图中可以看出 TSK2 算法正确边标准差在两个网络的学习中都具有明显优势。

由图 10 可知，TSK2 算法准确率最高。TSK2 算法学习 Asia 网络的准确率可以达到 97.5%，学习 Alarm 网络的准确率可以达到 93.9%，均领先对比算法。

实验结果表明，在标准数据集 Asia 和 Alarm 中，TSK2 算法表现最好，在相同样本量下学习到的正确边最多，产生错误边的概率最小，并且正确边标准差较小，明显优于基于因果效应的方法、MMHC 算法、MCMC 算法和爬山法，可以获得更准确的网络结构。

结束语　K2 算法需要给定准确的先验节点序才能学习准确的贝叶斯网络，对此，本文提出了一种基于变量邻居集和 v-结构信息的两步序搜索策略学习方法 TSK2。该算法在 v-结构信息可用时，利用 v-结构与汇点邻居集的序搜索策略获得部分最优序列；在无 v-结构信息可用时，搜索邻居节点间顺连、分连、汇连 3 种结构形式，对节点序列进行有效调整，降低节点序列陷入较坏情况的概率，提升算法变量序列寻优的性能，最后配合 K2 算法构建优质的结构模型。实验结果证明，与其他算法相比，TSK2 算法具备更高的准确性和稳定性，为贝叶斯网络的研究提供了新的技术路线。

下一步研究内容是面对非稳态的数据场景，利用 K2 算法构建拟合数据较好的自适应贝叶斯网络模型。

参 考 文 献

[1] LOU C Y, LI X S, ATOUI 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.

[2] TAGHI-MOLLA A, RABBANI M, GAVARESHKI M, 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.

[3] ZHANG W J, JIANG L X, ZHANG H, et al. A two-layer Bayesian model: random forest naive Bayes[J]. Computer Research and Development, 2021, 58(9): 2040-2051.

[4] MIGLIORINI F, DRIESSEN A, QUACK V, et al. Comparison between intra-articular infiltrations of placebo, steroids, hyaluronic and PRP for knee osteoarthritis: a Bayesian network meta-analysis[J]. Archives of Orthopaedic and Trauma Surgery, 2021, 141(9): 1473-1490.

[5] HU X G, LIU F, BU C Y. Research progress of cognitive tracking model in educational big data [J]. Journal of Computer Research and Development, 2020, 57(12): 2523-2546.

[6] CHICKERING D M. Learning Bayesian networks is NP-complete[J]. Networks, 1996, 112(2): 121-130.

[7] HECKERMAN B D, GEIGER D, CHICKERING D M. Learning Bayesian networks: the combination of knowledge and statistical data[J]. Machine Learning, 1995, 20(3): 197-243.

[8] SPIRTES P, GLYMOUR C. An algorithm for fast recovery of sparse causal graphs[J]. Social Science Computer Review, 1991, 9(1): 62-72.

[9] ZHANG L W, GUO H P. Introduction to Bayesian networks [M]. Beijing: Science Press, 2006: 66-70, 186-188.

[10] LIU X Q, LIU X S. Structure learning of Bayesian networks by continuous particle swarm optimization algorithms[J]. Journal of Statistical Computation and Simulation, 2018, 88(9): 1-29.

[11] PINTO P C, NAGELE A, DEJORI M, et al. Using a local discovery ant algorithm for Bayesian network structure learning [J]. IEEE Transactions on Evolutionary Computation, 2009, 13(4): 767-779.

[12] CHEN H Y, ZHANG N. Bayesian network structure learning based on hybrid improved bird swarm algorithm[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(1): 85-91, 98.

[13] COOPER G F, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data[J]. Machine Learning, 1992, 9(4): 309-347.

[14] LERAY P, FRANCOIS O. BNT structure learning package: Documentation and experiments[R]. Technical Report, Labora-

toire PSI, 2006.
[15] AN N, TENG Y, YANG J Y, et al. Bayesian network structure learning method based on causal effect[J]. Computer Application Research, 2018, 35(12): 3609-3613.
[16] WU Y H, MCCALL J, CORNE D. Two novel ant colony optimization approaches for Bayesian network structure learning[C]// Proceedings of IEEE Congress on Evolutionary Computation. Barcelona, Spain: IEEE, 2010: 1-7.
[17] BENMOHAMED E, LTIFI H, AYED M B. ITNO-K2PC: An improved K2 algorithm with information-theory-centered node ordering for structure learning[J/OL]. Journal of King Saud University-computer and Information Science. https://doi.org/10.1016/j.jksuci.2020.06.004.
[18] GAO F, HUANG D. A node sorting method for K2 algorithm in Bayesian network structure learning[C]// Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications. IEEE Press, 2020: 106-110.
[19] ANDRIEU C, FREITAS N D, DOUCET A, et al. An introduction to MCMC for machine learning[J]. Machine Learning, 2003, 50(1): 5-43.
[20] ZHOU M Y, LIU Y A, XIAO Y. Improvement of K2 algorithm in Bayesian network structure learning[J]. Journal of Nanjing University of Technology, 2020, 44(3): 320-324.
[21] AOUAY S, JAMOUSSI S, AYED Y B. Particle swarm optimization based method for Bayesian Network structure learning [C]// Proceedings of the 5th International Conference on Modeling, Simulation and Applied Optimization. Hammamet, Tunisia: IEEE, 2013: 1-6.
[22] BEHJATI S, BEIGY H. Improved K2 algorithm for Bayesian network structure learning[J]. Engineering Applications of Artificial Intelligence, 2020, 91(5): 1-12.
[23] YU K, LI J Y, LIU L. A review on algorithms for constraint-based causal discovery[EB/OL]. (2016-11-12)[2021-12-06]. https://arxiv.org/abs/1611.03977.
[24] 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.

XU Miao, born in 1996, master. Her main research interest is probability graph model learning.

QI Xiaolong, born in 1981, Ph.D, associate professor. His main research interests include probability graph modeling and machine learning.

（责任编辑：何杨）