

基于演化序搜索的混合贝叶斯网络结构学习方法

李明嘉, 钱鸿, 周爱民

引用本文

李明嘉, 钱鸿, 周爱民. 基于演化序搜索的混合贝叶斯网络结构学习方法[J]. 计算机科学, 2023, 50(10): 230-238.

LI Mingjia, QIAN Hong, ZHOU Aimin. Hybrid Bayesian Network Structure Learning via Evolutionary Order Search [J]. Computer Science, 2023, 50(10): 230-238.

相似文章推荐 (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)

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>

基于深度强化学习与程序分析的OJ习题推荐模型

OJ Exercise Recommendation Model Based on Deep Reinforcement Learning and Program Analysis

计算机科学, 2023, 50(8): 58-67. <https://doi.org/10.11896/jsjcx.220600260>

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>

基于贝叶斯攻击图的动态网络安全分析

Dynamic Network Security Analysis Based on Bayesian Attack Graphs

计算机科学, 2022, 49(3): 62-69. <https://doi.org/10.11896/jsjcx.210800107>

基于改进哈希时间锁的区块链跨链资产交互协议

Novel Hash-time-lock-contract Based Cross-chain Token Swap Mechanism of Blockchain

计算机科学, 2022, 49(1): 336-344. <https://doi.org/10.11896/jsjcx.210600170>

基于演化序搜索的混合贝叶斯网络结构学习方法

李明嘉 钱 鸿 周爱民

华东师范大学计算机科学与技术学院 上海 200062

华东师范大学上海智能教育研究院 上海 200062

(51205901074@stu.ecnu.edu.cn)

摘 要 贝叶斯网络是一种不确定性知识表示与推理的有效工具,学习其结构是利用这一工具进行推理的基础。现有的贝叶斯网络结构学习算法,在智能教育等应用场景中往往面临着难以权衡有效性与高效性的问题。一方面,评分搜索类方法能搜索到高质量的解,但面临着算法复杂度高的挑战。另一方面,混合类方法效率高,但所找到的解的质量不尽如人意。针对上述问题,提出了一种基于演化序搜索的混合贝叶斯网络结构学习方法(EvOS)。该方法首先通过约束类算法构建无向图骨架,然后利用演化算法搜索最优节点序,最后使用该节点序指导贪婪搜索得到贝叶斯网络结构。基于常用基准数据集以及教育知识结构发现任务,验证了所提方法的有效性与高效性。实验结果表明,所提方法相较于评分搜索类方法,能够在保持相仿精度的情况下最高加速百倍,且有效性显著高于混合类方法。

关键词: 贝叶斯网络;结构学习;序搜索;演化优化;知识结构发现

中图法分类号 TP183

Hybrid Bayesian Network Structure Learning via Evolutionary Order Search

LI Mingjia, QIAN Hong and ZHOU Aimin

School of Computer Science and Technology, East China Normal University, Shanghai 200062, China

Shanghai Institute of AI for Education, East China Normal University, Shanghai 200062, China

Abstract Bayesian network is an effective tool for uncertainty knowledge representation and reasoning. Learning and discovering its structure is the basis of reasoning via this tool. Existing Bayesian network structure learning algorithms often encounter the dilemma of balancing effectiveness and efficiency in real-world applications such as intelligent education. On the one hand, score-and-search methods can find out the high-quality solutions, but they suffer from the high algorithmic complexity. On the other hand, hybrid methods are highly efficient but the quality of the found solutions is not satisfactory. To address the above dilemma, this paper proposes an evolutionary order search based hybrid Bayesian network structure learning method called EvOS. First, the proposed EvOS constructs an undirected graph skeleton through a constraint algorithm, and then applies an evolutionary algorithm to search for the optimal node order, and finally uses the found node order to guide the greedy search so as to obtain the Bayesian network structure. This paper conducts the empirical study to verify the effectiveness and efficiency of the proposed EvOS in the commonly-used benchmark datasets as well as the real-world task of educational knowledge structure discovery. Experimental results show that, compared with the score-and-search methods, EvOS is able to achieve up to 100 times speedup while maintaining the similar accuracy, and its effectiveness is significantly better than that of the hybrid methods.

Keywords Bayesian network, Structure learning, Order search, Evolutionary optimization, Knowledge structure discovery

1 引言

贝叶斯网络(Bayesian Networks, BN)是机器学习领域中

一种常用的概率图模型。贝叶斯网络由其结构(一个有向无环图, Directed Acyclic Graph, DAG)以及参数(对应的条件概率表, Conditional Probability Table, CPT)组成, 这二者分别

到稿日期:2022-10-09 返修日期:2023-03-12

基金项目:国家自然科学基金(62106076);上海市科学技术委员会科技创新行动计划项目(19511120601);上海市自然科学基金面上项目(21ZR1420300);上海市教育委员会与上海市教育发展基金会“晨光计划”项目(21CGA32);中央高校基本科研业务费专项资金

This work was supported by the National Natural Science Foundation of China(62106076), Science and Technology Commission of Shanghai Municipality(19511120601), Natural Science Foundation of Shanghai(21ZR1420300), “Chenguang Program” Sponsored by Shanghai Education Development Foundation and Shanghai Municipal Education Commission(21CGA32) and Fundamental Research Funds for the Central Universities.

通信作者:钱鸿(hqian@cs.ecnu.edu.cn)

定性定量地描述了变量间的关联关系。贝叶斯网络是实现不确定性知识表达与推理的有效工具,并且具备良好的可解释性,目前已在智能教育^[1]、网络安全^[2]、计算机视觉^[3]、自然语言处理^[4]等诸多领域取得了成功应用。

贝叶斯网络的结构学习任务旨在从观测数据中推测贝叶斯网络的结构。发现贝叶斯网络的结构是后续计算其参数和使用贝叶斯网络做推理的前提,因此,结构学习是贝叶斯网络研究中至关重要的一部分。现有的贝叶斯网络结构学习方法可以分为3类:基于约束的方法、基于评分搜索的方法和混合类方法。基于约束的方法主要使用条件独立性检验来学习变量间的独立性与相关性,并以此为依据推断贝叶斯网络结构。其中,条件独立性检验一般是基于统计学或信息论中的度量来实现。约束类方法中的代表性算法包括PC(Peter-Clark)算法^[5]、最大最小父子集检测(Max-Min Parent and Children, MMPC)算法^[6]等。这类方法的不足之处在于:首先,囿于贝叶斯网络的马尔可夫等价性,基于约束的方法只能辨识出贝叶斯网络中的一些特定结构,故输出的结果中通常存在一些无法定向的边,这也是本文对该类方法不加以详细讨论的主要原因;其次,条件独立性检验有时会给出互相矛盾的结果,处理这些矛盾是一项非常困难的任务,这会影响约束类方法找到的解的质量;最后,此类方法对数据的要求较高,如果数据含有噪声则会大大影响学得结构的精度。

基于评分搜索的方法将贝叶斯网络的结构学习建模为一个优化问题。其首先定义一个评分函数,该评分函数刻画了输入图结构与观测数据之间的拟合程度。常用的评分函数有:BIC(Bayesian Information Criterion)^[7],AIC(Akaike Information Criterion)^[8],BDe(Bayesian Dirichlet equivalent)^[9]和BDeu(Bayesian Dirichlet equivalent uniform)^[10]等。在定义好评分函数后,再使用基于搜索与优化的方法找到最大化该评分的图结构。常用的搜索算法包括K2算法^[11]、K2GA(K2GeneticAlgorithm)算法^[12]、爬山法(Hill Climbing, HC)^[10]、贪婪等价类搜索(Greedy Equivalent Search, GES)算法^[13]等。相较于基于约束的方法,搜索评分的方法的优势在于可以避免处理条件独立性检验互相矛盾等问题。然而,在DAG空间中找最优解这一过程已被证明是一个NP难(Non-determinism Polynomial Hard)的组合优化问题^[14],因此,评分搜索类方法面临的主要瓶颈是搜索算法的复杂度较高,在时间受限的情形下其搜索到的结果可能不是全局最优。

混合类方法试图融合约束类与评分搜索类两种方法来弥补它们各自的缺陷与不足。其主要思想为:先用条件独立性检验降低搜索空间的大小,再用评分搜索的办法来进行局部搜索或对无向边进行定向。代表算法包括最大最小爬山法(Max-Min Hill-Climbing, MMHC)^[15]和H2PC(Hybrid HPC)算法^[16]等。然而,囿于其搜索部分采用的方法比较简单,这类方法目前在有效性上的表现还不够令人满意。

综上所述,目前的算法应用往往面临着在有效性和高效性两端难以权衡的难题。以MMHC为代表的混合类方法能够很快地给出解,然而解的质量并不十分理想;而以K2GA为代表的评分搜索方法往往能够搜索到高质量的解,但它在效率上的短板阻碍了其进一步应用。实际应用场景中,现有

的贝叶斯结构学习算法在有效性与高效性两端进行取舍,一般会导致其在另一端的表现不尽如人意。以智能教育领域中知识结构发现任务为例,机器需要自动发现一张试卷中每类题目的知识点结构关系图,帮助老师进行学情分析,这对算法的有效性提出了较高要求。如果使用现有的评分搜索类方法,可能很长时间才能学到一类题目的知识结构图,若对所有类型的题目都进行自动知识结构发现,总时间代价难以让人接受。

鉴于此,本文从实际需求出发,提出了一种基于演化序搜索的混合贝叶斯网络结构学习方法(Evolutionary Order Search Based Hybrid Bayesian Network Structure Learning Method, EvOS),旨在为教育等实际应用场景中的贝叶斯网络结构学习任务提供一种能够较好权衡“快”和“好”的方法。该方法受混合类方法与演化评分搜索类方法的启发,将前者“快”的关键(即用约束的方法限制搜索空间)与后者“好”的关键(即用节点序指导贪婪搜索)统一在一个框架之下,由节点序指导贪婪搜索来保证解的质量,同时利用约束的方法加速节点序的适应度评估。具体而言,EvOS首先利用基于约束的方法确定一个无向图骨架,然后利用该无向图参与节点序的适应度评估,并在此基础上使用遗传算法^[17](Genetic Algorithms, GA)搜索最优节点序,最后根据搜索到的节点序指导贪婪搜索得到最终的贝叶斯网络结构。

本文的主要贡献包括以下几个方面:

(1)提出了一种基于演化序搜索的混合贝叶斯网络结构学习方法EvOS,其能够较好地权衡算法的有效性和高效性。

(2)EvOS将贝叶斯网络结构学习“好”的关键与“快”的关键有机统一到一个框架之下,使用演化优化方法搜索高质量节点序,利用约束的方法加速适应度评估的过程,最后基于高质量节点序来指导贪婪搜索以保证解的质量。

(3)在常用基准数据集与教育领域中知识结构学习任务上进行的实验结果表明:相比评分搜索类方法,EvOS能够在保持相当的精度的情况下最高加速百倍,并且精度显著高于混合类方法。

本文第2节简要介绍评分搜索类方法中搜索算法的研究情况以及EvOS和这些方法的区别与联系;第3节详细描述EvOS的整体框架以及关键步骤;第4节与第5节分别展示EvOS在常用基准数据集与教育领域中的真实任务上的实验结果与分析;最后总结全文并对未来工作进行展望。

2 相关工作

本章主要介绍评分搜索类方法中搜索算法的相关工作。

在定义好评分函数后,求解贝叶斯网络结构学习这一图搜索问题往往是困难的,尤其是当节点个数较多时,从庞大的搜索空间中找到最优的结构无疑是一个挑战。求解这一NP难的组合优化问题时,研究者们往往采用启发式优化算法。常见的搜索算法大致可分为贪婪搜索和基于演化优化的方法两类。

贪婪搜索通过贪婪地加、减、反转边等操作来使评分上升,代表算法有K2^[11],HC^[10]和GES^[13]等。其中,K2算法是由节点序指导的贪婪搜索,其需要给定一个节点序作为输入,

在输入正确节点序的情况下往往能够搜索到非常接近真实图结构的解。然而在很多场景下,节点序这一先验知识是未知的,这也极大地限制了 K2 的应用。HC 与 GES 则不需要节点序作为输入,它们从一个空图开始,贪婪地寻求最大化评分的加、减、逆转边等操作,直到上述操作不能使评分增加。HC 与 GES 的问题是容易陷入局部最优而停滞,导致最终找到的结构与全局最优差别较大。EvOS 采用带节点序指导的贪婪搜索来得到最终输出的网络结构,而节点序的获取则依赖序空间上的演化优化的技术来实现。

演化计算是人工智能领域用于求解复杂优化问题的一类方法,如组合优化问题,其在工业设计^[18]、强化学习^[19]等领域被广泛应用。在贝叶斯网络结构学习中,主要是将评分当作解的适应度函数。常用的演化类方法又可以根据解的表示分为在 DAG 空间搜索和在节点序空间搜索两类。

第一类方法将贝叶斯网络表示为邻接矩阵,这是一种很自然的想法。Larranaga 等^[20]首先提出了基于遗传算法的贝叶斯网络结构学习方法,将解编码为邻接矩阵,使用评分作为适应度函数进行求解。为保证搜索到的结构无环,其使用修复算子对解进行修正。为了避免使用修复算子,Etzeberria 等^[21]使用了 fuse-DAGs^[22]来确保重组算子不会破坏局部最优结构。这类方法主要的难点在于如何保证解的合法性,即得到的图无环。而基于序搜索的方法则可以有效避开这一问题。

第二类方法则将搜索对象从 DAG 变为节点序。对一个贝叶斯网络来说,节点序刻画了其中变量的先后关系,在结构学习过程中可以说是最为重要的一方面^[23]。同时,基于节点序的搜索还有很多其他好处,如极大地限制了搜索空间,天然确保了解的无环性等。Larranaga 等^[12]提出的 K2GA 以 K2 算法输出的贝叶斯网络评分作为节点序的适应度函数,并借鉴了旅行商问题中重组与变异算子的设计,使用 GA 来搜索最优的节点序。Aouay 等^[24]在 K2GA 的基础上将演化的部分替换成了粒子群优化算法(Particle Swarm Optimization, PSO),提出了 K2PSO。与此类似的工作,还有 Wu 等^[25]提出的基于蚁群算法(Ant Colony Optimization)的 K2ACO。此类方法往往能给出质量较高的搜索结果,但其问题也很明显:每次评估节点序适应度时都需要调用 K2 算法,这一点极大地限制了这类方法的效率。EvOS 同样采用序空间上的演化优化来搜索最优节点序,不同之处在于 EvOS 采用的演化算法中适应度评估这一步不需要调用 K2 算法,而是用节点序在由前置步骤得到的无向图骨架上定向得到的有向图的评分作为该序的适应度值,这样做的好处是可大幅提升适应度评估这一过程的效率。

3 EvOS 算法

本节主要介绍 EvOS 算法的框架,其主要包含 3 个关键步骤:骨架学习、节点序搜索以及贝叶斯网络构建。第一步是无向图骨架学习,在这个阶段中 EvOS 需要学习一个刻画变量间依赖关系的无向图结构。无向图骨架学习的目的是限制每个变量可能父节点的范围,相当于变相地约减了搜索空间,从而加速了后续节点序的适应度评估。第二步是节点序

搜索,用节点序为第一步输出的无向图定向后所得的有向图评分作为该节点序的适应度值,并在设计好节点序空间上的演化算子后,使用遗传算法来搜索到最优节点序。最后是贝叶斯网络构建,利用第二步学到的节点序指导贪婪搜索,得到最终的贝叶斯网络结构。EvOS 的主要流程如图 1 所示。

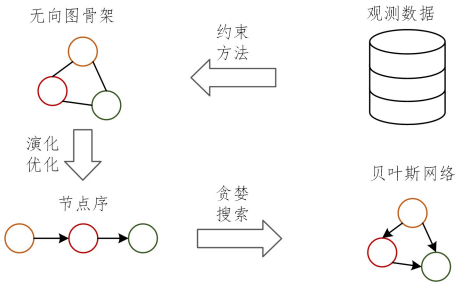


图 1 EvOS 流程示意图

Fig. 1 Sketch map of EvOS

3.1 问题定义与假设

令 $BN=(G,\theta)$ 是一个以变量集合 $V=\{V_1,V_2,\cdots,V_n\}$ 为节点的贝叶斯网络,其中 $G=(V,E)$ 是表示节点关系的有向无环图,也被称为贝叶斯网络的结构; $\theta=\{\theta_1,\theta_2,\cdots,\theta_n\}$ 表示每个节点 V_i 在给定它的父节点集 $Pa(V_i)$ 时的条件概率表,也被称为贝叶斯网络的参数。贝叶斯网络的结构学习问题被定义如下:给定关于变量集 V 上的 m 组观测数据矩阵 $D\in R^{m\times n}$,学习出变量 $V_1,V_2,\cdots,V_n$ 间的关联关系,以有向无环图 G 表示,使得 V 上的联合概率分布 $P(V)$ 能够被写成如下乘积分解的形式:

$$P(V)=\prod_{i=1}^n P(V_i|Pa(V_i))$$
 (1)

其中, $Pa(V_i)$ 表示 G 中 V_i 的父节点集。

在问题的定义之外,还会涉及如下假设。

假设 1(充分性假设)^[26] 设 V_i 为一变量, D 为观测数据矩阵,如果 D 中所有的样本都不包含 V_i 的取值,则称 V_i 为隐变量。如果变量集 V 中任意两个变量的直接原因变量都不是隐变量,那么变量集 V 被认为是满足充分性的。

假设 2(马尔可夫条件假设)^[5] 对于满足充分性的变量集 V 而言,假设有向无环图 $G=(V,E)$, V 的联合概率分布为 $P(V)$,如果满足对任意的 $V_i\in V$,给定其父节点集合 $Pa(V_i)$ 为条件集合时条件独立于其他所有非后裔节点,即称 $G=(V,E)$ 满足马尔可夫条件。

假设 3(忠诚性假设)^[5] 如果在给定变量集 V 的前提下,变量 V_i 和 V_j 互相独立或条件独立,那么在有向图 G 中, V_i 和 V_j 之间的所有路径被变量集 V 中合适的变量 d -分离(d -separation)^[27],则称变量集 V 的联合概率分布 P 与图 G 是互相忠诚的。

直观上讲,充分性假设确保了没有未被观测的共同因变量存在。事实上,在大多数贝叶斯网络结构学习任务中,也是不考虑未能被观测的隐变量的存在的。马尔可夫条件假设则是将所有变量间的关系都假设为其对应的条件概率分布依赖关系,这是对现实情况的一种简化。忠诚性假设的隐含意义是在结构学习的过程中变量不会出现额外的条件独立关系,这其实是一个较强的假设。Spirtes 和 Meek^[28]的工作已经

指出是存在一些分布 P 没有对应忠诚的有向无环图 G (尽管这样的分布非常少见)的,并且对于同一个分布 P ,可能存在多个忠诚的 G 。

此外,本文提出的贝叶斯网络结构学习方法是基于节点序搜索的,这里的节点序 ρ 是一个定义在变量集 V 上的排列,显然,每个节点序都与一个完全 DAG (Complete DAG, 变量两两之间都存在一条有向边)存在一一对应的关系,因此对于一个有向无环图 $G=(V,E)$,定义其对应节点序集合如下。

定义 1(节点序集合)

$\Omega^G = \{\rho; \rho \text{ 对应的完全 DAG 是 } G \text{ 的超图}\}$

其中,超图指该图的边集是 G 的边集的超集。值得注意的是,对于一个不完全的有向无环图 G ,其对应的节点序不是唯一的,即 $|\Omega^G| > 1$ 。

3.2 基于约束类方法的骨架学习

EvOS 算法的第一步是学习出变量间的依赖关系骨架,以一个无向图表示。这一步可以看作是对每个变量选择其可能的邻居节点集合,这将极大地缩减搜索空间的规模,为后续节点序搜索时解的评估加速。在贝叶斯网络学习领域,学习这样的无向图骨架可使用约束类方法完成,这类方法采用条件独立性检验来判断两个变量间是否可能存在边。然而,从理论上来说,想要得到完全精准的结果(即不多也不漏地找出无向边),需要遍历所有可能的条件集合,也就是对于任意两个变量 V_i 与 V_j ,需要给出所有 $IT(V_i, V_j | Z)$ 的结果。这里 IT 表示条件独立性检验 (Independence Test),如果两个变量条件独立,其返回结果为 True;反之则为 False; Z 是集合 $V \setminus (V_i \cup V_j)$ 所有可能的子集,也就是对于 n 个变量的问题,需要做 $2^n - 2$ 次条件独立性检验,而这仅仅只确定了两个变量间边存在的情况,显然这样的测试代价是不能承受的。因此,现有的约束类方法往往采用启发式的方法来进行条件独立性测试,如 MMPC 算法的核心启发式信息就是对于变量集中的每一个变量,找到最大化最小关联值的变量组成的集合,这个集合就是算法要找的父子节点集。其背后的直觉在于,最小关联值描述了两个变量是否可能条件独立,若大于 0 则有可能,那么与一个变量最小关联值最大的那个变量最有可能是其直接的父亲或儿子节点。然而在本文中,EvOS 算法第一步所需要的仅仅只是对搜索空间进行一定的约减,对无向图框架的精度要求并没有那么高,具体来说,EvOS 的骨架学习允许学到的无向图骨架中存在多出的边。因此本文提出了一种更简约的基于条件独立性检验的无向图骨架学习方法,并给出了该方法能够保障所有正确边都包含在该骨架中的理论证明。首先介绍邻居集合的定义。

定义 2(邻居集合)

$Nei(V_i) = \{V_j; V_j \in V \setminus V_i, IT(V_i, V_j | S) = \text{False}\}$ (2)

其中, $S = V \setminus (V_i \cup V_j)$,即 V 中除了 V_i 和 V_j 剩下变量组成的集合。邻居集合对于每个变量 V_i ,找出那些给定其他所有变量作为条件集合下与 V_i 不条件独立的节点组成的集合。显然,如果 $V_j \in Nei(V_i)$,也一定有 $V_i \in Nei(V_j)$,因此可以根据所有节点的邻居集合来构建无向图,下面的定理保证了这样做的合理性。

定理 1 给定一个贝叶斯网络结构图 $G=(V,E)$,对于

每一个节点 V_i ,有 $Pa(V_i) \subseteq Nei(V_i)$,即 V_i 在 G 中的父亲节点集合包含在其邻居集合中。

证明:对任意的 $V^* \in Pa(V_i)$,根据式(1)的乘积分解形式有 $IT(V_i, V^* | S) = \text{False}$,由定义 2 可得 $V^* \in Nei(V_i)$,因此 $Pa(V_i) \subseteq Nei(V_i)$ 。

值得注意的是,定理 1 只保证了根据邻居集合构建的无向图不会少边,但还是会存在一些多出的边。然而这里的无向图骨架只是用于后续节点序的适应度评估,与最终输出的网络结构无直接关联,在后续的章节中我们会看到 EvOS 是如何使用该骨架为节点序搜索提供快速评估的。

3.3 节点序搜索

上一节中介绍了如何通过基于约束的方法得到变量间的无向图骨架,接下来将介绍 EvOS 最核心的部分,即如何利用该无向图与相应的启发式算法搜索到最优的节点序。

首先,给定一个无向图 G_{Und} 与一个节点序 ρ ,可以根据 ρ 中节点的排序为 G_{Und} 中的边进行定向,其中节点序在前的节点成为父节点,反之成为子节点,得到有向图 G 。显然, G 中没有环路,是一个有向无环图。图 2 给出了一个该过程的示例,在该示例中,考虑序 $(1, 2, 3, 4)$ 与 $(4, 3, 1, 2)$ 分别为 4 个节点的完全无向图定向得到的有向无环图如图 2 所示。

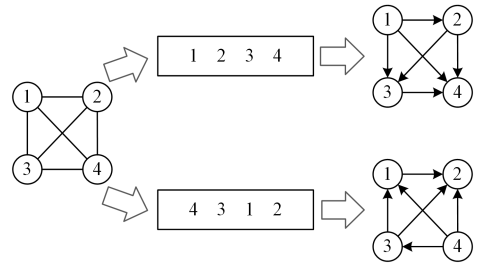


图 2 节点序定向过程实例

Fig. 2 Example of node order orientation process

得到有向无环图 G 后,将其当作贝叶斯网络的结构,根据给定的观测数据 D ,可以得到其 BDeu 评分如下:

$$f_{\text{BDeu}}(G, D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \left(\log \frac{\Gamma(\alpha/q_i)}{\Gamma(m_{ij} + \alpha/q_i)} + \sum_{k=1}^{r_i} \log \frac{\Gamma(m_{ijk} + \alpha/r_i q_i)}{\Gamma(\alpha/r_i q_i)} \right) \quad (3)$$

其中, n 是变量个数, r_i 为节点 V_i 的状态个数, q_i 为节点 V_i 的父节点状态组合总数, Γ 是 Gamma 函数, α 是超参数, m_{ijk} 是观测数据 D 中节点 V_i 为状态 k 且父节点状态组合为 j 的样本数量, $m_{ij} = \sum_{k=1}^{r_i} m_{ijk}$ 。从式(3)也能看出为什么需要做第一步的无向图骨架学习,这个无向图相当于把每个节点可能的父亲限制在了一定范围内,而 BDeu 评分中 m_{ij} 与 q_i 这两项的计算开销都是与父节点状态个数成正比的,这意味着多出一个父节点哪怕其只是一个二值变量,父节点的状态排列总数也要翻一倍。因此如果不使用无向图骨架对父节点个数加以限制,对节点序的搜索相当于对一个完全有向图的搜索,此时计算 BDeu 评分的代价是根本无法接受的。

接下来,这个评分值被当作该节点序的适应度值,然后根据这样的适应度值定义,使用遗传算法来搜索最优节点序,

如算法 1 所示。

算法 1 节点序搜索遗传算法

输入: 无向图骨架 G_{Und} , 观测数据 D , 搜索代数 $\max_epoch$, 评分函数 $f_{BDeu}(G, D)$
输出: 最优节点序 ρ^*

1. 随机初始化节点序种群 pop
2. 根据 f_{BDeu} 计算个体 ρ 为 G_{Und} 定向后得到的有向无环图 G 的评分作为 ρ 的适应度值并保存最优 ρ^*
3. for $epoch \leftarrow 1$ to $\max_epoch$ do
4. 使用重组、变异算子产生子代
5. 计算适应度值并根据选择算子选出新一代种群 pop'
6. 更新 ρ^*
7. end for
8. 输出 ρ^*

其中, 重组、变异算子是遗传算法中产生新个体所用到的的一种操作; 而选择算子是从父代子代中选出新一代个体的策略, 其重点在于兼顾新种群的质量与多样性。算法 1 具体采用的 3 种算子分别是顺序交叉、置换变异算子以及锦标赛选择算子。

(1) 顺序交叉算子^[12]。顺序交叉算子通过选择一个父母个体的一段子序列并保留其在另一个父母个体中的顺序来产生子代。例如两个父母个体分别是节点序 (1, 2, 3, 4, 5, 6) 和 (3, 1, 4, 6, 5, 2), 现在在第二个父母个体中随机选择 3 个点位, 如第 2, 4, 5 这 3 个位置, 其对应的节点分别是 1, 6, 5, 保留除了这 3 个点位之外的其他位置并复制到子代, 得到 (3, x, 4, x, x, 2), 子代剩下的部分, 即为 x 的那些节点, 按照第一个父母个体中 (1, 6, 5) 这 3 个节点的顺序进行排列, 得到第一个子代个体 (3, 1, 4, 5, 6, 2)。同理, 选择同样点位, 互换一二父母个体位置, 得到第二个子代个体 (1, 4, 3, 5, 2, 6)。顺序交叉算子的一个操作实例如图 3 所示。

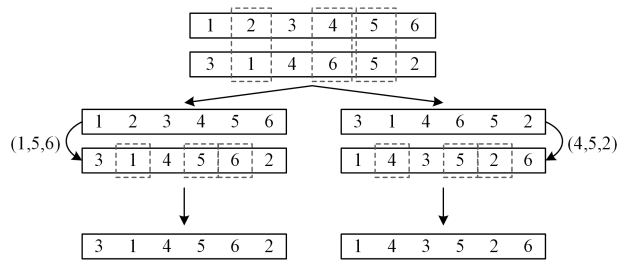


图 3 顺序交叉实例
Fig. 3 Example of order crossover operator

(2) 置换变异算子^[12]。置换变异算子选择个体的一段子片段, 将其随机移动到另一个地方形成新的个体。如节点序 (1, 2, 3, 4, 5, 6), 随机选择一个片段, 如 (2, 3), 将其移动到节点 5 后面, 形成新个体 (1, 4, 5, 2, 3, 6)。置换变异算子的一个操作实例如图 4 所示。

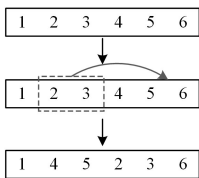


图 4 置换变异算子实例
Fig. 4 Example of displacement mutation operator

(3) 锦标赛选择算子^[12]。锦标赛选择算子是演化类算法中较为常用的一种选择算子, 其主要步骤为: 每次从待选择种群中选出一定数量的个体, 将其中适应度最高的放入下一代种群中, 重复该操作, 直到下一代种群规模达到预定大小。

3.4 贝叶斯网络构建

本节介绍 EvOS 的最后一步, 即结合 K2 算法并使用 ρ^* 作为其输入, 构建最终的贝叶斯网络结构。K2 算法是一种由节点序指导的贪婪搜索, 其要求输入一个节点序 ρ 与一个正整数 u , 输出为一个有向无环图, 即找到的贝叶斯网络结构。K2 算法的特点是效果高度依赖所输入节点序的质量, 如果所输入的节点序质量很高, 则能找到准确度非常高的网络结构, 反之输出的结构可能与真实结构相差甚远。K2 算法从一个包含所有节点的空图出发, 按照节点序 ρ 的顺序逐个考察节点, 并通过为该节点选择父节点的方式增加边, 其中父亲节点的选择遵循两个准则: (1) 其在节点序 ρ 中的位置在该节点之前; (2) 选择其为父节点能最大地增加整个网络的评分。正整数 u 是一个超参数, 被用来限制最大父节点的个数。在 EvOS 算法中, 通过之前的两步已经找到了我们所认为正确的节点序, 最后一步可使用 K2 来得到最终的贝叶斯网络结构。EvOS 的全流程如图 5 所示。

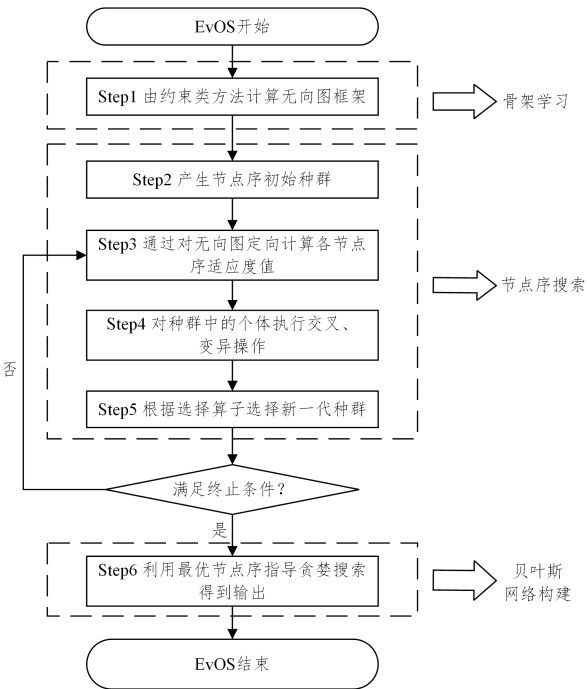


图 5 EvOS 全流程图
Fig. 5 Flow chart of EvOS

4 EvOS 在基准数据集上的实验

4.1 实验设置

为了验证 EvOS 算法的性能, 选取了多个常用的基准测试贝叶斯网络进行仿真实验, 分别是: Asia^[29], Sachs^[30], Child^[31], Insurance^[32], Alarm^[33] 以及 Hailfinder^[34]。各基准贝叶斯网络的属性如表 1 所列。

表 1 基准贝叶斯网络参数

Table 1 Parameters of benchmark Bayesian networks

网络	节点数	边数	参数数量
Asia	8	8	18
Sachs	11	17	178
Child	20	25	230
Insurance	27	52	984
Alarm	37	46	509
Hailfinder	56	66	2656

在以上贝叶斯网络中,分别通过采样产生样本容量大小为 5000 的仿真数据作为实验数据。为降低采样带来的随机性对结果的影响,每个网络分别产生 30 组样本数据,研究各算法在这 30 组采样中结构学习任务上的表现,结果以均值与标准差的形式呈现。

本文选取贝叶斯网络结构学习混合类方法 MMHC^[15]、贪婪评分搜索类方法 GES^[13]、演化评分搜索类方法 K2GA^[12],以及约束类方法 PC-stable^[35](相比原始 PC 算法,性能更加稳定)作为对比算法,每个算法分别在以上每个网络的每组采样数据中运行 10 次。

为评价算法性能,选取的评估指标有:F1 值、结构汉明距离(Structure Hamming Distance,SHD)与算法运行时间。其中,F1 是一个 0 到 1 之间的实数,衡量了预测结果与真实网络相比精确率与召回率的综合情况,其值越大,说明预测结果越接近真实。SHD 则衡量了两者相差边的数量,其是一个正整数,越小表示误差越小。F1 与 SHD 值由下式计算得到:

$$F1=2\times\frac{precision\times recall}{precision+recall}$$
 (4)

$$precision=\frac{TP}{TP+FP}$$
 (5)

$$recall=\frac{TP}{TP+FN}$$
 (6)

$$SHD=FN+FP$$
 (7)

其中,TP(True Positive)表示预测正确的边数,FP(False Positive)表示在预测结果中但不在真实网络中的边数,FN(False Negative)表示在真实网络中但没有被预测到的边数。

4.2 实验结果与分析

4.2.1 算法的有效性

EvOS 与 4 个对比算法在上述 6 个数据集上 30 组采样数据中运行的 F1 与 SHD 均值与标准差情况分别如图 6 与图 7 所示。

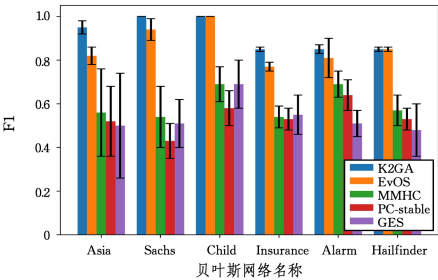


图 6 EvOS 与对比算法在 6 个基准数据集上的 F1 值对比 (F1 值越大越好)

Fig. 6 F1 comparison of EvOS and baselines on six benchmark data sets(the higher F1 score the better)

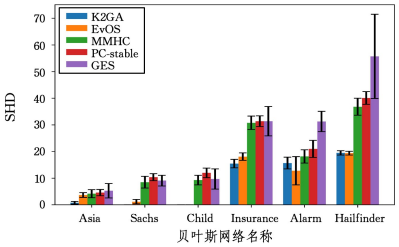


图 7 EvOS 与对比算法在 6 个数据集上的 SHD 值对比 (SHD 值越小越好)

Fig. 7 SHD comparison of EvOS and baselines on six benchmark data sets(the lower SHD the better)

由图 6 可见,若 F1 作为评价指标,EvOS 在有效性上与 K2GA 总体差距不大,在 6 个问题中有 4 个问题表现略差于 K2GA,另外 2 个问题上表现几乎一样。6 个问题上的 F1 性能损失百分比分别为:13.7%,6.0%,0.0%,9.4%,4.7%与 0.0%。相比其他 3 个对比算法,EvOS 在 6 个问题中 F1 这一指标上均明显占优,其中 EvOS 比 MMHC 在 6 个问题中 F1 值高出 17.4%~74.1%,比 PC-stable 高出 26.6%~118.6%,比 GES 高出 40.0%~84.3%。以上是从均值方面得到的分析结果。从标准差上看,K2GA 与 EvOS 的标准差较其他 3 个算法都偏小,说明这两种方法更加稳定,采样偏差所带来的不稳定性对其影响更小。值得一提的是,在 Sachs, Alarm 与 Hailfinder 中虽然 EvOS 的均值较 K2GA 偏小,但如果按均值加上标准差作为上限来看,在这 3 个问题中 EvOS 的上限与 K2GA 是非常接近的。

由图 7 可见,若 SHD 作为评价指标,EvOS 同样与 K2GA 在有效性上相差甚微。在 Sachs 上 K2GA 的平均 SHD 为 0,因此其对应的柱状条几乎不可见,在 Child 上同理。在 6 个基准问题上计算出 K2GA 和 EvOS 的平均 SHD 之差占真实网络结构总边数的百分比分别为:-36.3%,-7.1%,0.0%,-5.0%,6.3%与 0.2%。由此可见,在 Alarm 和 Hailfinder 两个基准问题上,EvOS 的平均 SHD 值比 K2GA 的更低。在 6 个基准问题上计算出 MMHC 和 EvOS 的平均 SHD 之差占真实网络结构总边数百分比最低为 6.3%,最高可达 42.9%;同理,PC-stable 和 GES 的百分比最低为 11.3%和 20.0%,最高可达 54.1%和 55.0%。这表明相较于 MMHC,PC-stable 和 GES 这 3 个算法,EvOS 在 SHD 这一指标上占优更加明显。

综上,从以 F1 与 SHD 两个指标衡量的有效性上看,EvOS 与 K2GA 差不多,同时显著优于 MMHC,PC-stable 和 GES。下面研究 EvOS 与 K2GA 在算法时间复杂度上的对比表现。

4.2.2 算法的高效性

在算法时间高效性方面,图 8 展示了 EvOS 与 K2GA 在 6 个贝叶斯网络中运行平均速度的对比,纵坐标的含义是 EvOS 相比 K2GA 的加速百分比。由图 8 可以看出,在以上 6 个数据集中,EvOS 在速度上相比 K2GA 均有数量级上的提升,最多可至 170 倍加速,最少也有 4 倍多加速。

从有效性与高效性两个方面综合来看,相比以 K2GA 为代表的演化评分搜索类方法,EvOS 在有效性相仿的情况下,大幅提升了时间效率,同时有效性显著好于以 MMHC,PC-stable 与 GES 为代表的混合类、约束类以及贪婪评分搜索类方法。

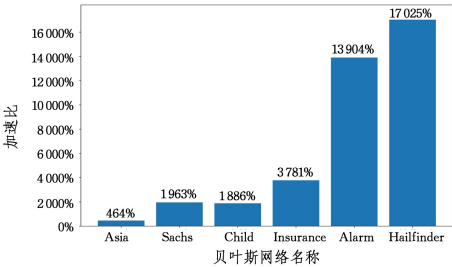


图 8 EvOS 与 K2GA 在 6 个数据集上的运行时间对比

Fig. 8 Comparison of running time of EvOS and K2GA on six benchmark data sets

5 EvOS 在知识结构发现任务上的应用

目前人工智能技术已经在教育领域扎根并推动着教育信息化的快速发展,智能教育从教学质量、教学效率以及教育公平性等多个方面创造了传统教育所缺失的价值,对于学生、学校、政府相关部门等都具有重大意义。本文结合贝叶斯网络结构学习技术,旨在通过对知识结构图的学习与推理为教师和学生提供具有启发性的建议。

本文在 Wu 等^[36]公布的数据集上进行实验,该数据集记录了某市近四千名高二学生在一次数学期末统考中的表现,包括这些学生在 16 道选择/填空题中的对错情况以及由教育学专家所标注的这些题目包含的知识点信息,这些知识点分别是“不等式性质”“数据采样方法”“等比数列”“函数与等式”“解三角形”“数据分析原理”“经典概率论”“线性规划”“算法定义”“算法逻辑”“等差数列”“空间想象能力”“抽象与总结能力”、“推理与论证能力”“计算能力”“数据处理能力”。这些知识点从某类非常具体的数学题(如解三角形)到某种数学概念(如函数与等式)再到学生某方面的能力(如计算能力),在抽象程度上层层递进。然而这是一整张试卷上包含的全部知识点,本文从中选取了一类题目,以该类题目包含的知识点为例进行知识结构学习,这类题目的一个示例如图 9 所示。

例:设△ABC 三内角:A、B、C 成等差数列,其正弦值:sinA、sinB、sinC 成等比数列,则这个三角形的形状是()

A. 直角三角形

B. 钝角三角形

C. 锐角三角形

D. 都有可能

图 9 知识结构学习例题

Fig. 9 Example for knowledge structure learning

由图 9 可知,该类题目并不是单独考查某个知识点,而是综合了三角函数与数列两个部分,这也是数学考试中难题、压轴题的一大特点。而对这样的题目作知识结构学习是非常有意义的,由知识结构学习得到的贝叶斯网络能帮助教师在教学过程中针对性地制定策略,以及考试后对学生错题归因分析。由专家标注的该类题目所包含的知识点间的知识结构如图 10 所示。

值得注意的是,图 10 中的箭头并不表示我们通常认知中的学习先后关系,而是表示知识点间的依赖关系。知识点 A 与知识点 B 间的箭头表示:掌握 A 的前提是掌握 B。如图 10 中“解三角形”→“函数与等式”这条边表示:掌握“解三角形”这一知识点依赖于学生掌握“函数与等式”的知识点,只有

掌握了后者才有可能掌握前者。可以看到,图中的每条路径体现的依赖都是在抽象程度上逐步提高,从对具体的某个考点的掌握到对某个数学概念、性质的理解,再到某方面的能力的高低,这也符合人们对于知识结构的直觉经验。

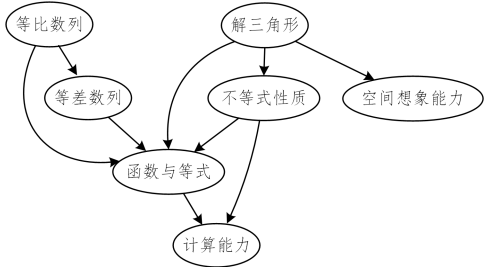


图 10 专家定义的知识点间贝叶斯网络结构图

Fig. 10 Expert defined Bayesian network structure between nodes of knowledge

本文根据 DINA 模型^[37]推理出这批学生在这些知识点上的掌握情况矩阵,以此为观测数据进行贝叶斯网络结构学习,并将得到的结构图与专家标注的真实结构图进行对比。EvOS 及其对比算法分别运行 30 次的性能均值与标准差的比较如表 2 所列。

表 2 EvOS 与对比算法在智能教育知识结构学习任务上的性能表现

Table 2 Performance of EvOS and comparison algorithms in knowledge structure learning task

方法	F1	SHD	运行时间/s
EvOS	0.58±0.04	6.6±1.1	199±14
K2GA	0.60±0.03	6.3±0.8	3964±129
MMHC	0.21±0.00	13.0±0.0	<1
GES	0.29±0.00	14.0±0.0	<1
PC-stable	0.22±0.00	13.0±0.0	<1

从表 2 可以看出,EvOS 与 K2GA 相较其他 3 个算法均在有效性上有显著优势,同时在这个 7 个节点 9 条边 4000 条观测数据的问题中,K2GA 的效率只有 EvOS 的 1/20,K2GA 平均要一个多小时才能得出知识结构图,而 EvOS 则只需要 3 分钟左右。这只是考虑一张试卷中一类题目的情况,如果是教师想要对某次考试的所有类题目知识点结构进行分析,EvOS 相比 K2GA 能够节省更多的时间。

如果综合考虑有效性与高效性,能得到如图 11 所示的性能帕累托(Pareto)图。

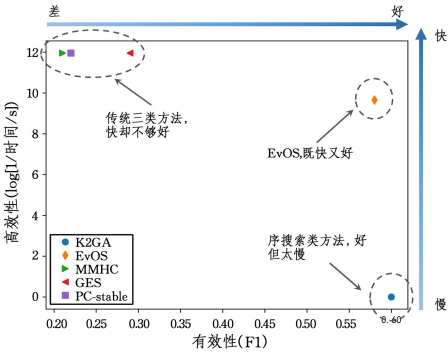


图 11 EvOS 与对比算法在知识结构学习任务中有效性与高效性的综合对比

Fig. 11 Comparison of effectiveness and efficiency of EvOS and benchmark algorithms in knowledge structure learning task

结束语 针对现有的贝叶斯网络结构学习算法在很多实际应用场景中难以权衡“好”和“快”的问题,本文提出了一种基于演化节点序搜索的混合贝叶斯网络结构学习方法EvOS。EvOS将贝叶斯网络结构学习“好”的关键与“快”的关键有机统一到一个框架之下,使用演化优化方法搜索高质量节点序;针对适应度评估过慢的问题,利用约束的方法进行加速;最后基于高质量节点序来指导贪婪搜索,以保证解的质量。在常用基准数据集与智能教育知识结构发现任务上的实验结果均表明:在众多对比方法中,EvOS是在高效性与有效性两端综合最优的方案。本文提出的EvOS可为贝叶斯网络结构学习在智能教育等实际场景中提供切实可行的方案。

对于未来的工作,我们的设想如下:首先,目前我们使用的技术是基于贝叶斯网络的,也就是对数据中关联关系的研究,后续希望能够将研究对象从贝叶斯网络变为因果网络,即从关联关系发现上升为因果关系发现,进而使我们发现的图结构具有因果语义,这样对我们的决策才能有指导意义;其次,目前EvOS能够处理的问题规模仍然不大,基本上能够做到数十个节点左右,后续希望能够将EvOS扩展到更高维度,如上百、数百个节点的问题中去;此外,虽然EvOS相比演化类方法快了许多,但其高效性相比传统方法仍处于弱势,因此后续希望能够在不损失有效性的前提下进一步提升算法的效率;最后,除了智能教育,许多其他应用场景也对贝叶斯网络结构发现的有效性和高效性权衡提出了类似需求,未来我们将在这些场景中应用本文提出的方法,进一步发挥其潜力。

参 考 文 献

- [1] TING C, CHEAH W, CHIUNG C. Student engagement modeling using bayesian networks[C]// Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. New York: IEEE, 2013: 2939-2944.
- [2] LI J R, LING X B, LI C X, et al. Dynamic Network Security Analysis Based on Bayesian Attack Graphs [J]. Computer Science, 2022, 49(3): 62-69.
- [3] WANG S F, HE S, WU Y, et al. Fusion of visible and thermal images for facial expression recognition [J]. Frontiers of Computer Science, 2014, 8(2): 232-242.
- [4] XU Y Y, CHAI Y M, WANG L M, et al. Emotional Sentence Classification Method Based on OCC Model and Bayesian Network [J]. Computer Science, 2020, 47(3): 222-230.
- [5] SPIRITES P, CLARK G, RICHARD S. Causation, prediction, and search [M]. Massachusetts: MIT Press, 2000: 221-234.
- [6] MARGARITIS D. Learning Bayesian network model structure from data [D]. Pittsburgh, Pennsylvania: Carnegie-Mellon Univ, Pittsburgh Pa School of Computer Science, 2003.
- [7] SCHWARZ G. Estimating the dimension of a model [J]. The Annals of Statistics, 1978, 6(2): 461-464.
- [8] AKAIKE H. A new look at the statistical model identification [J]. IEEE Transactions on Automatic Control, 1974, 19(6): 716-723.
- [9] BUNTINE W. Theory refinement on Bayesian networks[C]// Proceedings of the 7th Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann, 1991: 52-60.
- [10] HECKERMAN D, DAN G, DAVID M C. Learning Bayesian networks: The combination of knowledge and statistical data [J]. Machine Learning, 1995, 20(3): 197-243.
- [11] COOPER G F, HERSKOVITS E. A Bayesian method for the induction of probabilistic networks from data [J]. Machine Learning, 1992, 9(4): 309-347.
- [12] LARRANAGA P, KUIJPERS C M H, MURGA R H, et al. Learning Bayesian network structures by searching for the best ordering with genetic algorithms [J]. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 1996, 26: 487-493.
- [13] CHICKERING D M. Optimal structure identification with greedy search [J]. Journal of Machine Learning Research, 2002, 3: 507-554.
- [14] COOPER G F. The computational complexity of probabilistic inference using Bayesian belief networks [J]. Artificial Intelligence, 1990, 42(2/3): 393-405.
- [15] TSAMARDINOS I, BROWN L E, ALIFERIS C F. The maximum hill-climbing Bayesian network structure learning algorithm [J]. Machine Learning, 2006, 65(1): 31-78.
- [16] GASSE M, ALEX A, HAYTHAM E. A hybrid algorithm for Bayesian network structure learning with application to multi-label learning [J]. Expert Systems with Applications, 2014, 41(15): 6755-6772.
- [17] HOLLAND J. Adaptation in Natural and Artificial Systems [M]. Ann Arbor: University of Michigan Press, 1975: 174-186.
- [18] JIN Y C, SENDHOFF B. A systems approach to evolutionary multiobjective structural optimization and beyond [J]. IEEE Computational Intelligence Magazine, 2009, 4(3): 62-76.
- [19] QIAN H, YU Y. Derivative-free reinforcement learning: a review [J]. Frontiers of Computer Science, 2021, 15(6): 156336.
- [20] LARRANAGA P, POZA M, YURRAMENDI Y, et al. Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(9): 912-926.
- [21] ETXEBERRIA R, LARRANAGA P, JUAN M P. Analysis of the behaviour of genetic algorithms when learning Bayesian network structure from data [J]. Pattern Recognition Letters, 1997, 18(11/12/13): 1269-1273.
- [22] MATZKEVICH I, BRUCE A. The topological fusion of Bayes nets[C]// Proceedings of the 8th Annual Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann, 1992: 191-198.
- [23] LV Y L, MIAO J Z, LIANG J Y, et al. BIC-based node order learning for improving Bayesian network structure learning [J]. Frontiers of Computer Science, 2021, 15(6): 156337.
- [24] AOUAY S, JAMOSSI S, AYED Y B. Particle swarm optimization based method for Bayesian network structure learning [C]// Proceedings of 5th International Conference on Modeling, Simulation and Applied Optimization. New York: IEEE, 2013: 1-6.
- [25] WU Y H, JOHN M, DAVID C. Two novel ant colony optimization approaches for Bayesian network structure learning[C]//

Proceedings of the 2010 IEEE Congress on Evolutionary Computation. New York:IEEE,2010:1-7.

[26] SPIRTEs P,ZHANG K.Causal discovery and inference;concepts and recent methodological advances[C]// Proceedings of Applied Informatics. Springer Open,2016:1-28.

[27] PEARL J.Causality;models,reasoning and inference(Second Edition)[M]. Cambridge: Cambridge University Press, 2009: 123-145.

[28] SPIRTEs P,MEEK C.Learning Bayesian networks with discrete variables from data[C]// Proceedings of the 1st Annual Conference on Knowledge Discovery and Data Mining, Montreal, Menlo Park;Morgan Kaufmann,1995:294-299.

[29] LAURITZEN S L,DAVID J S.Local computations with probabilities on graphical structures and their application to expert systems [J]. Journal of the Royal Statistical Society; Series B (Methodological),1988,50(2):157-194.

[30] SACHS K,PEREZ O,PE’ER D,et al. Causal protein-signaling networks derived from multi parameter single-cell data[J]. Science,2005,308(5721):523-529.

[31] SPIEGELHALTER D J,DAWID A P,LAURITZENS L,et al. Bayesian analysis in expert systems [J]. Statistical Science, 1993,8(3):219-247.

[32] BINDER J,KOLLER D,RUSSELL S,et al. Adaptive probabilistic networks with hidden variables [J]. Machine Learning, 1997,29(2):213-244.

[33] BEINLICH I A,SUERMONDTH J,CHAVEZ R M,et al. The ALARM monitoring system:A case study with two probabilistic inference techniques for belief networks[C]// LNMI 38; Proceedings of the 2nd European Conference on Artificial Intelli-

gence in Medicine. Berlin,Heidelberg;Springer,1989:247-256.

[34] ABRAMSON B,BROWN J,EDWARDS W,et al. Hailfinder: A Bayesian system for forecasting severe weather [J]. International Journal of Forecasting,1996,12(1):57-71.

[35] COLOMBO D,MATHUIS M H. Order-Independent Constraint-Based Causal Structure Learning [J]. Journal of Machine Learning Research,2014,15:3921-3962.

[36] WU R Z,LIU Q,LIU Y P,et al. Cognitive modelling for predicting examinee performance[C]// Proceedings of the 24th International Joint Conference on Artificial Intelligence. Menlo Park; AAAI,2015:1017-1024.

[37] TORRE J. DINA model and parameter estimation : A didactic [J]. Journal of Educational and Behavioral Statistics, 2009, 34(1):115-130.



LI Mingjia, born in 1998, postgraduate. His main research interests include evolutionary machine learning, causal inference and intelligent education.



QIAN Hong, born in 1991, Ph.D, associate researcher. His main research interests include derivative-free optimization, evolutionary machine learning, and intelligent education.

(责任编辑:柯颖)