

贝叶斯网络结构评分函数研究

李淑智¹, 徐光华^{1,2}, 刘弹¹, 张熠卓¹

1. 西安交通大学机械工程学院, 西安 710049

E-mail: lishuzhilsz@189.cn

2. 西安交通大学机械制造系统工程国家重点实验室, 西安 710049

E-mail: xugh@mail.xjtu.edu.cn

摘要: 为了从数据中学习贝叶斯网络结构, 人们提出了不同的结构评分函数用于贝叶斯网络结构的学习。本文通过对贝叶斯网络节点间关联强度的分析研究, 提出了一种新的网络结构评分函数, 即贝叶斯条件概率统计评分函数(BCPS), 并给出了 BCPS 评分函数的严密计算公式。利用 BCPS、BIC、BDe 评分函数和 K2 结构学习算法学习 Asia 网和 Alarm 网的网络结构, 学习结果表明, BCPS 评分函数能够正确评价数据和结构的拟合程度, 且 K2-BCPS 结构学习算法的稳定性优于 K2-BIC 结构学习算法, 精确度优于 K2-BDe 结构学习算法。

关键词: 贝叶斯网络, 结构学习, BCPS评分函数, BIC评分函数, BDe评分函数, K2算法。

Research on Bayesian Network Structure Score Function

Shuzhi Li¹, Guanghua Xu^{1,2}, Tan Liu¹, Yizhuo Zhang¹

1. School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

E-mail: lishuzhilsz@189.cn

2. State Key Laboratory for Manufacturing System Engineering, Xi'an Jiaotong University, Xi'an 710049, China

E-mail: xugh@mail.xjtu.edu.cn

Abstract: In order to induct a Bayesian network from data, the different structure score functions were proposed and applied to the structure learning of Bayesian network. This paper researches the inter-node correlation strength of network and presents a new score function, namely Bayesian conditional probability statistic (BCPS) score function. BCPS score function rigorous formula was given. The BCPS, BIC and BDe score function and K2 algorithm was applied to structure learning of Asia network and Alarm network. The results show that BCPS score function can estimate the fitting degree of data and structure rightly. The stability of K2-BCPS structure learning algorithm was better than K2-BIC structure learning algorithm, and the Accuracy is superior to K2-BDe structure learning algorithm.

Key Words: Bayesian network, structure learning, BCPS score function, BIC score function, BDe score function, K2 algorithm.

1 引言

贝叶斯网络是一种图形化的概率模型, 具有强大的表达不确定性知识和进行不确定性推理的能力。贝叶斯网络是在定性的图形结构中, 融入定量的概率数值, 表达随机变量之间潜在的联系及关联强度。

贝叶斯网络学习就是要寻找一个能够反映数据各变量相互之间的依赖关系的贝叶斯网络模型, 使给定数据能与网络模型最好的拟合。贝叶斯网络的学习包括结构建模和参数学习, 结构学习是贝叶斯网络学习的研究的核心。最具代表性的结构学习方法有两类: 基于搜索与评分的方法^[1]和基于约束测试的方法^[2]。

搜索与评分的方法是将网络结构学习视为最优化问题, 学习目标是搜索评分高的网络结构。影响贝叶斯

网络结构学习精确度的主要因素是评分函数和搜索算法, 选择合理的评分函数^[3-6]和优化搜索算法^[7-8]是贝叶斯网络结构学习的核心问题。比较常用的评分函数有贝叶斯信息标准 BIC 评分函数、MDL 评分函数、AIC 评分函数和基于贝叶斯统计的 BDe 评分函数等。BIC 评分函数是在样本满足独立同分布假设的前提下, 用对数似然度评价结构与数据的拟合程度, 并添加惩罚项以防止过拟合, MDL 和 AIC 与 BIC 的评价方式是一致的。BDe 评分函数结合了网络结构的先验知识, 然后在进行贝叶斯统计。几种评分函数均是在一定的假设前提下, 利用不同的统计方法进行统计计算, 实现结构与数据的拟合。

贝叶斯网络中父节点与子节点的关联方式是由条件概率表达的, 其评分函数的实质是定量的度量父节点

与子节点间的关联强度。本文从条件概率统计的角度提出了一种新的评分函数，贝叶斯条件概率统计 (Bayesian conditional probability statistic) 评分函数，简称 BCPS。该评分函数直接对每个样本的条件概率进行统计，度量网络结构和数据的拟合程度，为了防止数据与结构的过渡拟合添加惩罚项，以实现变量间的关联强度和网络的复杂度之间的平衡。用 BIC、BDe、BCPS 评分函数和 K2 结构学习算法学习贝叶斯网络结构，实验结果表明，当数据集达到一定规模（能够准确表达节点的关联关系）时，K2-BCPS 能够学习到较优的结构，具有较高的学习精确度。

2 贝叶斯网络结构评分函数

设 $X=\{X_1, X_2, \dots, X_n\}$ 是一组变量， $D=\{D_1, D_2, \dots, D_m\}$ 关于这些变量的是样本集， S 是以 X_1, X_2, \dots, X_n 为节点的贝叶斯网络。假设数据集满足独立同分布假设，则 S 相对于数据 D 的优劣可以用评分函数度量。

2.1 贝叶斯信息标准BIC评分

Schwarz 提出了等价的贝叶斯信息标准 (Bayesian information criterion)，简称 BIC 评分。BIC 评分函数是在样本满足独立同分布假设的前提下，用对数似然度量结构与数据的拟合程度。BIC 评分函数如下：

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

其中 q_i 是变量 X_i 父节点取值组合的数目， r_i 是变量 X_i 的取值数目， m_{ijk} 表示 X_i 的父节点取 j 值， X_i 取 k 值时的样本数目。 $m_{ij} = \sum_{k=1}^{r_i} m_{ijk}$ ， $\theta_{ijk} = \frac{m_{ijk}}{m_{ij}}$ 表示似然条件概率，有 $0 \leq \theta_{ijk} \leq 1, \sum_k \theta_{ijk} = 1$ 。

BIC 评分的第一项是模型的对数似然度，度量结构与数据的拟合程度。第二项是模型复杂度的罚项，防止数据与结构的过渡拟合。

后来，Rissanen 提出最小描述长度 MDL (minimum description length) 的评分方法，即网络结构的描述长度和样本的编码长度之和最小。MDL 评分函数与 BIC 评分函数实质是一致的，均是在网络结构的复杂性与准确性之间选择平衡点。

2.2 贝叶斯评分BDe

Cooper 等提出了基于贝叶斯统计思想的 BDe (Bayesian Dirichlet) 评分函数，核心思想是利用先验知识和数据寻求后验概率最大的网络结构。贝叶斯评分函数如下：

$$BDe(S|D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \left[\log \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + m_{ij})} + \sum_{k=1}^{r_i} \log \frac{\Gamma(\alpha_{ijk} + m_{ijk})}{\Gamma(\alpha_{ijk})} \right] \quad (2)$$

式中： $\alpha_{ij} = \sum_{k=1}^{r_i} \alpha_{ijk}$ 。

α_{ijk} 是参数的先验分布， α_{ijk} 的选取并非易事，需要对每一个可能结构都提供参数的先验分布，而结构数目众多，无法一一罗列。实际中一般用 $\alpha_{ijk} = \alpha / (r_i q_i)$ 得到超参数 α_{ijk} 。

2.3 贝叶斯网络条件概率统计评分

有 n 节点的贝叶斯网络，变量 X_i 的条件概率为 θ_{ijk} ，父节点取第 j ($1 \leq j \leq q_i$) 个值时，若 $\max(\theta_{ijk})$ 的值等于或接近 1，表明在第 j 个取值时，父节点和子节点具有因果关系；若 $\max(\theta_{ijk})$ 的值等于或接近于 $1/r_i$ ，表明在第 j 个取值时，父节点和子节点不具有因果关系，故 θ_{ijk} 反映出了父节点和子节点的因果强度。

BIC 评分函数中 $\sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \theta_{ijk}$ 是变量 X_i 的家族分值，实质是对每个样本的条件概率取对数，然后求和。父节点和子节点的关联强度不仅与因果强度 θ_{ijk} 有关，同时与样本个数 m_{ijk} 有关。 θ_{ijk} 取对数的目的是为了减少计算量^[9]，使乘法运算变为加法运算，并且对数函数 $\log \theta_{ijk}$ 具有单调递增的性质，基于以上分析，本文提出了一种新的贝叶斯网络家族评分函数，表达式为 $\sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \theta_{ijk}$ ，其本质是直接对每个样本的条件概率进行求和计算，分值表示父节点和子节点的关联强度，分值越大，父节点和子节点的关联强度就越强。度量贝叶斯网络结构与数据拟合程度的评分函数则为所有家族分值之和，本文定义这种新的贝叶斯网络结构评分函数为贝叶斯条件概率统计评分，即 BCPS 评分函数，其表达式为：

$$BCPS(S|D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \theta_{ijk}$$

由于结构学习总是试图寻找分值最高的网络结构，利用 BCPS 评分函数进行结构学习其结果是一个完全的有向无环图，出现数据与结构的过渡拟合。对于同一个网络，样本数目 m 越大，网络结构的分值就越大，分值与 m 呈单调递增关系，则惩罚值与样本数目和结构复杂度也应该呈单调递增关系，BCPS 评分函数的惩罚项为 $\lambda \sum_{i=1}^n q_i (r_i - 1) m$ ，其中 $\sum_{i=1}^n q_i (r_i - 1)$ 为网络结构的复杂度， λ 为惩罚系数。

BCPS 评分函数改写为：

$$BCPS(S|D) = \sum_{i=1}^n \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \theta_{ijk} - \lambda \sum_{i=1}^n q_i (r_i - 1) m \quad (3)$$

BCPS 评分函数的分值越大说明变量之间的关联强度越大，数据和网络结构拟合的越好。

3 贝叶斯网络评分函数的分解

对贝叶斯网络结构进行评分，实质是对家族进行局部结构评分。局部结构评分的前提是评分函数具有可分解性，即网络结构的评分函数能被分解成家族评分函数。BIC 评分函数和 BDe 评分函数都是可以分解的^[9]。

由式(1)可以得出 BIC 的家族评分函数为：

$$BIC(\langle X_i, \pi(X_i) \rangle | D) = \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \log \theta_{ijk} - \frac{1}{2} q_i (r_i - 1) \log m \quad (4)$$

其中 $\pi(X_i)$ 为 X_i 的父节点集合。

由式(2)可以得出 BDe 的家族评分函数为：

$$BDe(\langle X_i, \pi(X_i) \rangle | D) = \sum_{j=1}^{q_i} \left[\log \frac{\Gamma(\alpha'_{ij})}{\Gamma(\alpha'_{ij} + m_{ij})} + \sum_{k=1}^{r_i} \log \frac{\Gamma(\alpha'_{ijk} + m_{ijk})}{\Gamma(\alpha'_{ijk})} \right] \quad (5)$$

BCPS 评分函数的实质是先对家族进行评分，然后计算贝叶斯网络的分值。因此 BCPS 评分函数的家族评分函数由式(3)可以得出：

$$BCPS(\langle X_i, \pi(X_i) \rangle | D) = \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} m_{ijk} \theta_{ijk} - \lambda q_i (r_i - 1) m \quad (6)$$

4 贝叶斯网络结构学习算法

从数据中学习贝叶斯网络结构的方法有约束测试方法、搜索与评分方法。约束测试的方法比较直观，贴近贝叶斯网络的语义，不足之处是对条件独立性测试产生的误差非常敏感，对于大的条件集，条件独立性测试难于完成。搜索与评分的方法是将网络结构学习视为最优化问题，学习目标是搜索评分高的网络结构，目前该方法应用范围很广。评分搜索的结构学习算法分为两步：首先确定网络结构评分函数，然后确定网络结构学习算法。学习算法有穷举法和启发式搜索算法，由于网络结构数目庞大，穷举法在实际中是不可行的，启发式结构学习算法主要有：K2 算法^[8,9]、爬山法、马尔可夫链蒙特卡洛(MCMC)搜索等。

K2算法是Cooper等提出的，K2算法出于复杂度的考虑，不是寻找评分最高的模型，而是寻找在一定条件下的最优模型。限制条件有两个，一是变量排序 p 是一个拓扑结构，二是任意一个变量的父节点个数不超过 u ，

限制条件使搜索空间大大缩小。在正确指定节点次序的情况下，K2算法具有执行效率高、精度高的特点，故结构学习算法选用K2算法。

5 实验与分析

根据 <http://www.norsys.com> 提供的 Asia 网（8 个变量，8 条边）的结构和概率参数表分别生成 200、1000、2000、5000 个样本的随机模拟数据集作为 Asia 网的标准数据集。根据 <http://www.norsys.com> 提供的 Alarm 网（37 个变量，46 条边）的结构和概率参数表分别生成 1000、2000、5000、10000 个样本的随机模拟数据集作为 Alarm 网的标准数据集。

K2 算法中 Asia 网和 Alarm 网的节点次序指定为网络的原始节点顺序，Asia 网的任意一个变量的父节点个数不超过 2，Alarm 网的任意一个变量的父节点个数不超过 5，评分函数分别为 BIC、BDe 和 BCPS，结构学习算法简称为 K2-BIC、K2-BDe、K2-BCPS 算法。

5.1 BCPS惩罚系数 λ 对结构精确度的影响

Asia 网和 Alarm 网均使用 5000 个样本的数据集，评分函数 BCPS 的惩罚系数 λ 取值分别为 0.01、0.001、0.0001，利用 K2-BCPS 结构学习算法学习贝叶斯网络结构，学习结果如表 1。表中 A 表示增加边，De 表示缺失边，R 表示正确边。

表1 K2-BCPS学习算法实验结果

λ	Asia			Alarm		
	A	De	R	A	De	R
0.01	0	3	5	0	21	25
0.001	1	1	7	1	1	45
0.0001	2	0	8	16	1	45

从表 1 的实验结果可以看出， $\lambda=0.001$ 时，Asia 网和 Alarm 网能够学习到较优的网络结构； $\lambda=0.01$ 时，网络结构缺失边较多，没有增加边，表明惩罚项太重，出现数据和结构的欠拟合； $\lambda=0.0001$ 时，网络结构增加边较多，表明惩罚项太轻，出现数据和结构的过拟合。实验结果表明 λ 直接影响结构学习的精确度， $\lambda=0.001$ 是一个合理取值。

5.2 评分函数和数据集规模对结构精确度的影响

Asia 网和 Alarm 网均使用各自的随机模拟数据集，结构学习算法为 K2-BIC、K2-BDe、K2-BCPS($\lambda=0.001$)。Asia 网和 Alarm 网的结构学习实验结果如表 2、3。

从表 2 的实验结果可以看出，Asia 网在 200 个样本时，由于数据集太小，不能表达节点的关联关系，故三种结构学习算法的学习精确度很差，均不能学习到原始结构；在 500 个样本时，K2-BDe 学习到原始结构，

在 5000 个样本时，K2-BIC 学习到原始结构；在 2000、5000 个样本时 K2-BCPS 学习到缺失一条边的较优结构。实验结果表明：对于节点较少的 Asia 网，数据集的规模大于 500 时，三种结构学习算法均能学习到较优或最优的结构。

从表 3 的实验结果可以看出，Alarm 网在 1000 个样本时，K2-BIC、K2-BDe、K2-BCPS 三种结构学习算法的学习精确度均非常差。其原因是：对于一个具有 37 个变量的 Alarm 网，1000 个样本相对于网络复杂度而言数据集规模太小，故数据和结构不能正确拟合。

表 3 中 K2-BIC 实验结果表明，在 5000 个样本时，Alarm 网学习到的模型结构有 45 条正确边，仅有一条边

(12→23)由于关联强度很弱而缺失（这条边在所有的学习结果中均缺失），故该结构是一个最优的结构；在 10000 个样本时，由于惩罚值相对于总分值过小，即惩罚较轻，出现了数据和结构的过拟合；在 2000 个样本时，由于惩罚值相对于结构分值过大，即惩罚较重，导致缺失边增加。实验结果表明，BIC 评分函数的惩罚项对学习结果的精确度有较大影响。

表 3 中 K2-BDe 实验结果表明，随着样本个数的增加，可以学习到更多的正确边，模型结构更接近于实际结构。由于先验分布被假设为均匀分布，和实际先验分布差别较大，故通过结构学习得到的结构与实际结构差别较大，不能学习到较优的结构。

表2 Asia网结构学习实验结果

algorithm	m=200			m=500			m=2000			m=5000		
	A	De	R	A	De	R	A	De	R	A	De	R
K2-BIC	1	4	4	1	1	7	0	1	7	0	0	8
K2-BDe	3	3	5	0	0	8	1	0	8	1	0	8
K2-BCPS	5	4	4	1	1	7	0	1	7	0	1	7

表3 Alarm网结构学习实验结果

algorithm	m=1000			m=2000			m=5000			m=10000		
	A	De	R	A	De	R	A	De	R	A	De	R
K2-BIC	18	22	24	0	3	43	0	1	45	1	1	45
K2-BDe	28	15	31	5	8	38	11	5	41	8	3	43
K2-BCPS	18	17	29	1	1	45	1	1	45	1	1	45

表 3 中 K2-BCPS 实验结果表明，在 2000、5000、10000 个样本时，学习得到的结构均有 45 条正确边，一条边缺失和一条增加边，结构完全一致的，没有随着数据集的增大而发生变化。实验结果表明 BCPS 评分函数的惩罚项是合理的，能够很好的平衡网络结构的复杂性与准确性。当数据集为 m=2000 时，数据集能够准确表达节点的关联关系，m≥2000 时，K2-BCPS 能够学习到较优的网络结构。

通过对表 2、表 3 的实验结果的分析，BCPS 评分函数能够敏感到能够准确表达节点关联关系的数据集，数据集能够准确表达节点关联关系时，K2-BCPS 能够学习得到最优的结构。BIC 评分函数不能敏感到准确表达节点关联关系的数据集，样本个数过多或过少时 K2-BIC 均不能学习到最优结构，BCPS 评分函数稳定性明显优于 BIC 评分函数。BCPS 评分函数的评价性能明显优于 BDe 评分函数的评价性能，BDe 评分函数的评价性能是最差的。

6 结论

本文分析了贝叶斯网络父节点与子节点关联强度的定量表达方式，提出了一种新的 BCPS 评分函数，并

给出了严密的计算公式。实验结果表明了 BCPS 评分函数能够正确评价数据和结构的拟合程度，并能够敏感到能够准确表达节点关联关系的数据集，其稳定性优于 BIC 评分函数，精确度优于 BDe 评分函数，为贝叶斯网络结构提供了一种新的评分函数。

致谢

本文得到了 863 国家高技术研究发展计划 (2007AA04Z432)资助。

参考文献

[1] YU Jing, SMITH V A, WANG P P. Advances to Bayesian network inference for generating causal network form observational biological data. *Bioinformatics*, 20(18):3594-3603, 2004.

[2] Cheng Jie, Bell D, Liu Wei-ru. Learning Bayesian networks from data: an efficient approach based on information theory. *Artificial Intelligence*, 137(1, 2): 43-90, 2002.

[3] Gurovskii M Z, Bidyuk P I, Terent'ev A N. Methods of constructing Bayesian network based on scoring functions. *Cybernetics and System Analysis*, 44(2):219-224, 2008.

[4] Shulin Yang, Kuo-Chu Chang. Comparison of Score Metrics for Bayesian Network Learning. *IEEE Trans. MAN AND*

- CYBERNETICS-PART A: SYSTEMS AND HUMANS*, 32(3): 419-428, 2002.
- [5] Schwarz G. Estimating the dimension of a model. *Annals of Statistic*, 7(2):461-464, 1978.
- [6] Cooper G F. Herskovits E.A. Bayesian method for the induction of probabilistic network form data. *Machine learning*, 9(4):309-347, 1992.
- [7] 范敏, 黄席樾, 石为人等. 一种改进的贝叶斯网络结构学习算法. *系统仿真学报*, 20(17): 4613-4617, 2008.
- [8] 王双成, 冷翠平, 李小琳. 小数据集的贝叶斯网络结构学习. *自动化学报*, 35(8): 1063-1070, 2009.
- [9] 张连文, 郭海鹏. 贝叶斯网引论, 第一版, 科学出版社, 2006.