

P.O.Box 8718, Beijing 100080, China	Journal of Software, July. 2004,15(7):1042-1048
E-mail: jos@iscas.ac.cn	ISSN 1000-9825, CODEN RUXUEW, CN 11-2560/TP
http://www.jos.org.cn	Copyright © 2004 by The Editorial Department of Journal of Software

具有丢失数据的贝叶斯网络结构学习研究

王双成, 苑森淼

[Full-Text PDF](#) [Submission](#) [Back](#)

王双成, 苑森淼 (吉林大学 计算机科学与技术学院, 吉林 长春 130025)

作者简介: 王双成(1958—),男,吉林长春人,博士,副教授,主要研究领域为数据库,数据挖掘;苑森淼(1943—),男,教授,博士生导师,主要研究领域为数据库,人工智能,计算机网络系统.

联系人: 王双成 Phn: +86-431-5701777, E-mail: wangsc@nenu.edu.cn, <http://www.jlu.edu.cn>

Received 2003-01-20; Accepted 2003-09-15

Abstract

At present, the method of learning Bayesian network structure with missing data is mainly based on the search and scoring method combined with EM algorithm. The algorithm has low efficiency and easily gets into local optimal structure. In this paper, a new method of learning Bayesian network structure with missing data is presented. First, unobserved data are randomly initialized. As a result, a complete data set is got. Based on the complete data set, the maximum likelihood tree is built as an initialization Bayesian network structure. Second, unobserved data are reassigned by combining Bayesian network with Gibbs sampling. Third, on the basis of the new complete data set, the Bayesian network structure is regulated based on the basic dependency relationship between variables and dependency analysis method. Finally, the second and third steps are iterated until the structure goes stable. This method can avoid the exponential complexity of standard Gibbs sampling and the main problems in the existing algorithm. It provides an effective and applicable method for uncertain knowledge representation, inference, and reasoning with missing data.

Wang SC, Yuan SM. Research on learning Bayesian networks structure with missing data. *Journal of Software*, 2004,15(7):1042~1048.

<http://www.jos.org.cn/1000-9825/15/1042.htm>

摘要

目前主要基于EM算法和打分-搜索方法进行具有丢失数据的贝叶斯网络结构学习,算法效率较低,而且易于陷入局部最优结构.针对这些问题,建立了一种新的具有丢失数据的贝叶斯网络结构学习方法.首先随机初始化未观察到的数据,得到完整的数据集,并利用完整数据集建立最大似然树作为初始贝叶斯网络结构,然后进行迭代学习.在每一次迭代中,结合贝叶斯网络结构和Gibbs sampling修正未观察到的数据,在新的完整数据集的基础上,基于变量之间的基本依赖关系和依赖分析思想调整贝叶斯网络结构,直到结构趋于稳定.该方法既解决了标准Gibbs sampling指数复杂性,又避免了现有学习方法所存在的主要问题,为具有不完整数据的不确定性知识表示、推断和推理提供了有效和可行的方法.

基金项目: Supported by the National Natural Science Foundation of China under Grant No.60275026 (国家自然科学基金); the Natural Science Foundation of Jilin Province of China under Grant No.20030517-1 (吉林省自然科学基金)

References:

- [1] Heckerman D. Bayesian networks for data mining. Technical Report, MSR-TR-97-02, Microsoft Research, Redmond, 1997.
- [2] Chickering DM. Learning Bayesian networks is NP-Hard. Technical Report, MSR-TR-94-17, Microsoft Research, Redmond, 1994.
- [3] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977,39(1):1~38.

- [4] Binder J, Koller D, Russell S, Kanazawa K. Adaptive probabilistic networks with hidden variables. *Machine Learning*, 1997,29 (2-3):213~244.
- [5] Chickering DM, Heckerman D. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning*, 1997,29(2-3):181~212.
- [6] Liu DY, Wang F, Lu YN, Xue WX, Wang SX. Research on learning Bayesian network structure based on genetic algorithm. *Journal of Computer Research and Development*, 2001,38(8):916~922 (in Chinese with English abstract).
- [7] Friedman N. Learning belief networks in the presence of missing values and hidden variables. In: *Proc. of the 14th Int'l Conf. on Machine Learning*. San Francisco: Morgan Kaufmann Publishers, 1997. 125~133.
- [8] Friedman N. The Bayesian structural EM algorithm. In: *Proc. of the 14th Int'l Conf. on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann Publishers, 1998. 129~138.
- [9] Geman S, Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1984,6(6):721~742.
- [10] Mao SS, Wang JL, Pu XL. *Advanced Mathematical Statistics*. Beijing: Higher Education Press, Berlin: Springer-Verlag, 1998. 401~459 (in Chinese).
- [11] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Machine Learning*, 1997,29(2-3):131~161.
- [12] Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*, 1997,29(2-3):103~130.
- [13] Cheng J, Bell D, Liu WR. Learning Bayesian networks from data: An efficient approach based on information-theory. *Artificial Intelligence*, 2002,137(1-2):43~90.
- [14] Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufmann Publishers, 1988. 117~133.
- [15] Chickering DM. Learning equivalence classes of Bayesian network structures. *Machine Learning*, 2002,2(2):445~498.
- [16] Lam W, Bacchus F. Learning Bayesian belief networks: An approach based on the MDL principle. *Computational Intelligence*, 1994,10(4):269~293.

附中文参考文献:

- [6] 刘大有,王飞,卢奕南,薛万欣,王松昕.基于遗传算法的Bayesian网结构学习研究. *计算机研究与发展*,2001,38(8):916~922.
- [10] 茆诗松,王静龙,濮晓龙. *高等数理统计*.北京:高等教育出版社,柏林:施普林格出版社,1998.401~459.