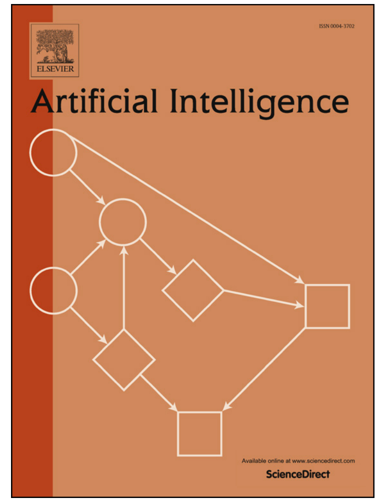
Accepted Manuscript



Entropy-based Pruning for Learning Bayesian Networks using BIC

Cassio P. de Campos, Mauro Scanagatta, Giorgio Corani, Marco Zaffalon

PII: S0004-3702(18)30167-X
DOI: https://doi.org/10.1016/j.artint.2018.04.002
Reference: ARTINT 3068

To appear in: *Artificial Intelligence*

Received date: 21 July 2017
Revised date: 19 March 2018
Accepted date: 13 April 2018

Please cite this article in press as: C.P. de Campos et al., Entropy-based Pruning for Learning Bayesian Networks using BIC, *Artif. Intell.* (2018), https://doi.org/10.1016/j.artint.2018.04.002

Entropy-based Pruning for Learning Bayesian Networks using BIC

Cassio P. de Campos[a,b,*], Mauro Scanagatta[c], Giorgio Corani[c], Marco Zaffalon[c]

[a] *Utrecht University, The Netherlands*
[b] *Queen's University Belfast, United Kingdom*
[c] *Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland*

Abstract

For decomposable score-based structure learning of Bayesian networks, existing approaches first compute a collection of candidate parent sets for each variable and then optimize over this collection by choosing one parent set for each variable without creating directed cycles while maximizing the total score. We target the task of constructing the collection of candidate parent sets when the score of choice is the Bayesian Information Criterion (BIC). We provide new non-trivial results that can be used to prune the search space of candidate parent sets of each node. We analyze how these new results relate to previous ideas in the literature both theoretically and empirically. We show in experiments with UCI data sets that gains can be significant. Since the new pruning rules are easy to implement and have low computational costs, they can be promptly integrated into all state-of-the-art methods for structure learning of Bayesian networks.
Keywords: Structure learning; Bayesian networks; BIC; Parent set pruning.

1. Introduction

A Bayesian network [1] is a well-known probabilistic graphical model with applications in a variety of fields. It is composed of (i) an acyclic directed graph (DAG) where each node is associated to a random variable and arcs represent dependencies between the variables entailing the *Markov* condition: every variable is conditionally independent of its non-descendant variables given its parents; and (ii) a set of conditional probability mass functions defined for each variable given its parents in the graph. Their graphical nature makes Bayesian networks excellent models for representing the complex probabilistic relationships existing in many real problems ranging from bioinformatics to law, from image processing to economic risk analysis.

Learning the structure (that is, the graph) of a Bayesian network from complete data is an NP-hard task [2]. We are interested in score-based learning, namely finding the structure which maximizes a score that depends on the data [3]. A typical first step of methods for this purpose is to build a list of suitable candidate parent sets for each one of the n variables of the domain. Later an optimization is run to find one element from each such list in a way that maximizes the total score and does not create directed cycles. This work concerns pruning ideas in order to build those lists. The problem is unlikely to admit a polynomial-time (in n) algorithm, since it is proven to be LOGSNP-hard [4]. Because of that, usually one forces a maximum in-degree (number of parents per node) k and then simply computes the score of all parent sets that contain up to k parents. A worth-mention exception is the greedy search of the K2 algorithm [5].

A high in-degree implies a large search space for the optimization and thus increases the possibility of finding better structures. On the other hand, it requires higher computational time, since there are $\Theta(n^k)$ candidate parent sets for a bound of k if an exhaustive search is performed. Our contribution is to provide new rules for pruning sub-optimal parent sets when dealing with the *Bayesian Information Criterion* score [6], one of the most used score functions in the literature. We

*Corresponding author