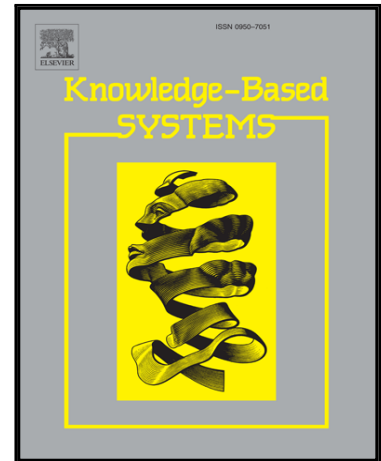


Accepted Manuscript

A Parallel Algorithm for Bayesian Network Structure Learning from Large Data Sets

Anders L. Madsen, Frank Jensen, Antonio Salmerón, Helge Langseth, Thomas D. Nielsen

PII: S0950-7051(16)30246-5
DOI: [10.1016/j.knosys.2016.07.031](https://doi.org/10.1016/j.knosys.2016.07.031)
Reference: KNOSYS 3619



To appear in: *Knowledge-Based Systems*

Received date: 28 February 2016
Revised date: 8 July 2016
Accepted date: 23 July 2016

Please cite this article as: Anders L. Madsen, Frank Jensen, Antonio Salmerón, Helge Langseth, Thomas D. Nielsen, A Parallel Algorithm for Bayesian Network Structure Learning from Large Data Sets, *Knowledge-Based Systems* (2016), doi: [10.1016/j.knosys.2016.07.031](https://doi.org/10.1016/j.knosys.2016.07.031)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Parallel Algorithm for Bayesian Network Structure Learning from Large Data Sets

Anders L. Madsen^{a,b,*}, Frank Jensen^a, Antonio Salmerón^d, Helge Langseth^c,
Thomas D. Nielsen^b

^a*HUGIN EXPERT A/S, DK-9000 Aalborg, Denmark*

^b*Aalborg University, DK-9220 Aalborg, Denmark*

^c*Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*

^d*University of Almería, ES-04120 Almería, Spain*

Abstract

This paper considers a parallel algorithm for Bayesian network structure learning from large data sets. The parallel algorithm is a variant of the well known PC algorithm. The PC algorithm is a constraint-based algorithm consisting of five steps where the first step is to perform a set of (conditional) independence tests while the remaining four steps relate to identifying the structure of the Bayesian network using the results of the (conditional) independence tests. In this paper, we describe a new approach to parallelization of the (conditional) independence testing as experiments illustrate that this is by far the most time consuming step. The proposed parallel PC algorithm is evaluated on data sets generated at random from five different real-world Bayesian networks. The algorithm is also compared empirically with a process-based approach where each process manages a subset of the

*Corresponding author

Email addresses: alm@hugin.com (Anders L. Madsen), fj@hugin.com (Frank Jensen), antonio.salmeron@ual.es (Antonio Salmerón), helgel@idi.ntnu.no (Helge Langseth), tdn@cs.aau.dk (Thomas D. Nielsen)

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات