



HAL
open science

Learning a forest of Hierarchical Bayesian Networks to model dependencies between genetic markers

Raphaël Mourad, Christine Sinoquet, Philippe Leray

► **To cite this version:**

Raphaël Mourad, Christine Sinoquet, Philippe Leray. Learning a forest of Hierarchical Bayesian Networks to model dependencies between genetic markers. 2010. hal-00444087v1

HAL Id: hal-00444087

<https://hal.science/hal-00444087v1>

Preprint submitted on 5 Jan 2010 (v1), last revised 19 Jan 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning a forest of Hierarchical Bayesian Networks to model dependencies between genetic markers

Raphaël Mourad[†], Christine Sinoquet[‡], Philippe Leray[†]

[†]LINA, UMR CNRS 6241, Ecole Polytechnique de l'Université de Nantes, rue Christian Pauc, BP 50609,
44306 Nantes Cedex 3, France,

[‡]LINA, UMR CNRS 6241, Université de Nantes, 2 rue de la Houssinière, BP 92208, 44322 Nantes
Cedex, France

— *KnOwledge and Decision (KOD)* —



RESEARCH REPORT

N^o hal-xxxxxxx

January 2010



Raphaël Mourad†, Christine Sinoquet‡, Philippe Leray†

Learning a forest of Hierarchical Bayesian Networks to model dependencies between genetic markers

?? p.

Les rapports de recherche du Laboratoire d'Informatique de Nantes-Atlantique sont disponibles aux formats PostScript® et PDF® à l'URL :

<http://www.sciences.univ-nantes.fr/lina/Vie/RR/rapports.html>

Research reports from the Laboratoire d'Informatique de Nantes-Atlantique are available in PostScript® and PDF® formats at the URL:

<http://www.sciences.univ-nantes.fr/lina/Vie/RR/rapports.html>

© January 2010 by **Raphaël Mourad†, Christine Sinoquet‡, Philippe Leray†**

Learning a forest of Hierarchical Bayesian Networks to model dependencies between genetic markers

Raphaël Mourad[†], Christine Sinoquet[‡], Philippe Leray[†]

raphael.mourad,christine.sinoquet,philippe.leray@univ-nantes.fr

Abstract

We propose a novel probabilistic graphical model dedicated to represent the statistical dependencies between genetic markers, in the Human genome. Our proposal relies on building a forest of hierarchical latent class models. It is able to account for both local and higher-order dependencies between markers. Our motivation is to reduce the dimension of the data to be further submitted to statistical association tests with respect to diseased/non diseased status. A generic algorithm, CFHLC, has been designed to tackle the learning of both forest structure and probability distributions. A first implementation of CFHLC has been shown to be highly-scalable: benchmarks describing 10^5 variables for 2000 individuals were parsed in about 45 hours on a standard personal computer.

Learning a forest of Hierarchical Bayesian Networks to model dependencies between genetic markers

Raphaël Mourad[†], Christine Sinoquet[‡], Philippe Leray[†]

Abstract

We propose a novel probabilistic graphical model dedicated to represent the statistical dependencies between genetic markers, in the Human genome. Our proposal relies on building a forest of hierarchical latent class models. It is able to account for both local and higher-order dependencies between markers. Our motivation is to reduce the dimension of the data to be further submitted to statistical association tests with respect to diseased/non diseased status. A generic algorithm, CFHLC, has been designed to tackle the learning of both forest structure and probability distributions. A first implementation of CFHLC has been shown to be highly-scalable: benchmarks describing 10^5 variables for 2000 individuals were parsed in about 45 hours on a standard personal computer.