Bayesian Networks

The Sparse Candidate Algorithm

Yvonne Pöschl

March 29, 2006

Outline

Outline

1. Introduction
2. Bayesian Networks
3. Sparse Candidate Algorithm
4. Data and Problems

Introduction

- Regulatory networks
- Metabolic pathways
- Biology, pharmacology, and medicine
- (causal) Bayesian networks
- Reconstruction of underlying system

Outline

What are Bayesian networks?

- Graphical models
- Represent probability distribution on $\mathbf{X} = \{X_1, \ldots, X_l\}$ in a dataset $\mathbf{D}$
- Consists of 2 components
  - Structure S: directed acyclic graph
    - Nodes
    - Edges
  - Parameters θ: conditional probability distributions

Learning Bayesian networks

- Find the network maximizing some objective function (score)
- Bayesian approach
- MAP approach
  $\log P(S|\mathbf{D}) = \log P(S) + \log P(\mathbf{D}|S) - \log P(\mathbf{D})$
- Evidence $P(\mathbf{D}|S)$
  $P(\mathbf{D}|S) = \int_\theta P(\mathbf{D}|S,\theta)P(\theta|S)\mathrm{d}\theta$
- Priors

Outline

1. Introduction
2. Bayesian Networks
3. **Sparse Candidate Algorithm**
4. Data and Problems

Reconstructing the structure S

- Search is NP-hard
- Restriction of the search space
- Heuristics - Sparse Candidate Algorithm (SCA)
- Local search (local sub-problems)
- Decomposable score

SCA - Restriction

- Iterative algorithm
- The SCA consists of 2 parts
  - **Restrict** phase
  - **Maximize** phase
- Restriction of the parents for each node to a small candidate set
- Instead of *l*-1 parents, only *c* possible parents for each node, where $c << l$
- Sub-optima possible

SCA - Maximize

- Consider local sub-problems
- $\mathit{Score}(S : \mathbf{D}) = \sum_i \mathit{ScoreContribution}(X_i, \mathit{Pa}(X_i) : \mathbf{D})$
- Possible edge changes

	before			after		
add	B		A	B	$\rightarrow$	A
remove	B	$\rightarrow$	A	B		A
reverse	B	$\rightarrow$	A	B	$\leftarrow$	A

- Greedy hill climbing, simulated annealing
- Sub-optima possible

Outline

Data and Problems

- Data
  - Microarray data
  - LCMS/GCMS data
- Problems
  - Small datasets
  - Direction of an edge - causality
  - Normalization and discretization