Modelling Regulatory Pathways in *E.coli* From Time Series Expression Profiles

Irene Ong
Jeremy Glasner
David Page

University of Wisconsin-Madison

Synopsis

- Task: model *regulation* of genes expressed under different conditions over *time*
- Approaches:
  - Bayesian Network with hidden variables (incorporating prior knowledge)
  - Dynamic Bayesian Network (temporal information)
Motivation

- Understanding gene regulation
  - Discover unknown interactions
  - Novel drug targets
  - Early diagnosis
- Bayesian Network with hidden variables
  - Simplifies dependency structure
- Dynamic Bayesian Network
  - Regulatory events
  - Feedback loops and causality

Outline

- Gene regulation (prior knowledge)
- Bayesian Network (BN)
- Incorporate prior knowledge into Bayesian Network
- Use time in Bayesian Network
  - => Dynamic Bayesian Network
Prior knowledge: Operons

- Roughly 400 known operons in *E.coli*
- Estimated over 1000 operons
- Craven *et al.* (ISMB 2000) mapped *E.coli* genes into known and predicted operons

Bayesian Network (BN)

Data:

<table>
<thead>
<tr>
<th></th>
<th>geneA</th>
<th>geneB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt1</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Expt2</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Expt3</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Expt4</td>
<td>↓</td>
<td>↑</td>
</tr>
</tbody>
</table>

Friedman *et al.* (2000)
Learning BN graph structure

Using *prior knowledge*: Operon

Our approach: group genes into operons
Incorporating operons in BN

- Use known and predicted operons
- Alleviate noise (independent indicators)

BN Operon structure search
Dataset

- Experiments by Khodursky et al. (2000)
  - Vary tryptophan availability
  - Well studied regulatory pathway
- Data consists of
  - 5 rich, 9 starved
  - 169 genes (142 operons) selected by Khodursky et al. based on expression levels

BN operon structure search

- For each of 9 key *trp* operons
  - Considered all 141 other operons as possible parent
  - Log likelihood used as scoring function
  - Not guaranteed globally optimal
  - Select 20 most probable parents
Tryptophan pathway from literature

9 key operons

- **trpR** operon – precursor of *trp* operon
- **tryR** operon
- Tryptophan molecule

Results of BN operon model

- Identified dependencies with direct parents, children and siblings
**Excess tryptophan condition**

- All 9 key operons influenced by excess tryptophan
- Histidine operon was potential parent for 5 of 9 key operons

**Tryptophan starved condition**

- \textit{trpR} correlated with 6 operons
- In addition, 7 other operons were potential parents for all 9 key operons
Are the correlations meaningful?

- Khodursky *et al.* (2000) found correlations for 2 of the 7 but not *rplK*

- We found a correlation for *rplK*
  - Yang *et al.* (2001) identified that *rplK* is involved in regulating the response to starvation of amino acids

Using time information

- Build Dynamic Bayesian Network
- Compare two consecutive measurements (relative change)
- 3 sets of 4 time step experiments
  - 1 tryptophan *rich* (excess)
  - 2 tryptophan *starved*
Modelling time

- Build Dynamic Bayesian Network model
  - Use operon map

![Operon expression influence diagram]

 operon’s expression at time $i$ influences its expression at time $i+1$

time 0  time 1  time 2

Time related assumptions

- Changes in the world are caused by stationary processes (change is governed by laws that themselves do not change)

![Operon state diagram]

current state

previous state evidence variables

time 0  time 1  time 2
Detailed view: initial DBN structure

DBN structure learning

- For each of 9 key *trp* operons
  - One parent operon is fixed
  - Consider all 141 other operons from previous time step as possible second parent
  - Select highest likelihood score
  - Record best 20 second parents
DBN structure search

Dynamic Bayesian Net result

- 7 would correctly model causality (red lines)
- Probability of at least 1 of these 10 operons being in the top 20 best parents for each of the 9 operons: 0.059
Summary

- Incorporating operon information in BN
  - Reduces search space
  - Alleviates noise
  - Avoids learning arcs between genes in same operon
- Modelling time information using DBN
  - Captures sequence of regulatory events
  - Allows modelling of feedback loops

Future directions

- Use of larger data sets
- Include additional prior knowledge:
  - hidden or observed variables
  - partial pathway structure
- Improve efficiency of structure learning
Summary

- Incorporating operon information in BN
  - Reduces search space
- Modelling time information using DBN
  - Captures sequence of regulatory events
- Acknowledgements: NIH 5 T32 GM08349, NSF Grant 9987841
- Paper at http://www.cs.wisc.edu/~ong