Computational Systems Biology: lecture 22

Tommi Jaakkola
tommi at csail dot mit dot edu
Revised by David Gifford
gifford at mit dot edu
Recall: signaling networks

- Human T cells, small molecule interventions

Can we automate the selection of interventions?
Broader motivation

- We have some control over what data to generate
- Computational experiment design (a.k.a. active learning) exploits this degree of freedom to minimize the cost (e.g., human involvement) needed to obtain an accurate model
- Computational experiment design methods have been used across a number of areas
  - agriculture
  - manufacturing
  - machine learning
  - biology
  etc.
Biological experiment design

- Example questions
  - which proteins to inhibit so as to best recover a pathway?
  - which genes to delete / silence?
  - which time points to sample?
  - which transcription factors to profile?
  - which probes to use in an array design?
  - etc.

- The key problem is to anticipate how much we would gain from a particular experiment prior to actually carrying it out
A priori / sequential design

- Possible “designs” can be
  - a priori designs where the experiments to be carried out are laid out before any outcomes are known
  - sequential designs where the next experiment is chosen in response to experiments carried out so far

- The process for sequential design:
  1. Select the “best” experiment
  2. Carry out the experiment
  3. Refine the model based on the outcome
A priori / sequential design

- Possible “designs” can be
  - a priori designs where the experiments to be carried out are laid out before any outcomes are known
  - sequential designs where the next experiment is chosen in response to experiments carried out so far

- The process for sequential design:
  1. Select the “best” experiment
  2. Carry out the experiment
  3. Refine the model based on the outcome

- The selection of “best” experiment is entirely model driven, based on hypothetically carrying out and evaluating the experiments
Outline of problems to solve

- In order to rank possible experiments, we have to be able to
  - quantify what we are uncertain about in the current model
  - characterize what outcomes we would expect to see from any particular experiment
  - evaluate, for each possible outcome, how the model would be refined
  - score resulting changes to the model

- Each possible new experiment is tried out computationally (hypothetically)
Five problems to address

1. Specify the set of hypotheses (model class), uncertainty
   – Bayesian networks over $x_1, \ldots, x_n$, existing data $D$, uncertainty characterized by $P(G|D)$

2. Limit the set of possible experiments
   – gene deletions (silencing, protein inhibition, etc)

3. Predict outcomes for each pair (model, deletion)
   – values for the remaining variables given del($x_i$) are drawn from possible models $P(x|\text{del}(x_i), D, G)$

4. Refine the model in response to hypothetical data
   – posterior $P(G|D, \{x, \text{del}(x_i)\})$

5. Evaluate a score for each possible experiment
   – value of information calculation
I. Posterior over BNs

- Suppose we have initial data in the form of $m$ complete observations of the values of the variables $D = \{x_{1t}, \ldots, x_{nt}\}_{t=1,\ldots,m}$

- We can write the posterior distribution over the possible Bayesian networks as a product of local scores of selecting parents $x_{pa_i}$ for each variable $x_i$

\[
P(G|D) \propto P(D|G)P(G)
\]

\[
P(D|G) = \prod_{i=1}^{n} S(i|pa_i, D)
\]

This decomposition results from our assumptions about the prior over the parameters (e.g., parameter independence)
1. Posterior over BNs

- If we assume a Dirichlet prior over the parameters, and that the prior is independent across possible parent configurations, then

\[
S(i|pa_i, D) =
\]

\[
= \int_\theta P(\theta|G) \left[ \prod_{t=1}^{m} \theta_{x_{it}|x_{pa_i,t}} \right] d\theta
\]

\[
= \int_\theta P(\theta|G) \left[ \theta_{x_{i1}|x_{pa_i,t}} \right] d\theta \int_\theta P(\theta|D_1, G) \left[ \theta_{x_{i2}|x_{pa_i,2}} \right] d\theta \cdots
\]

\[
= \prod_{t=1}^{m} \frac{N^{t-1}(x_{it}, x_{pa_i,t}) + \alpha(x_{it}, x_{pa_i,t})}{N^{t-1}(x_{pa_i,t}) + \alpha(x_{pa_i,t})}
\]

where the counts \( N^{t-1}(\cdot) \) are evaluated on the basis of the first \( t - 1 \) observations.
2. Limiting the set of experiments

- Why should we not consider a large number of possible experiments rather than just deletions of individual (or a few) genes?
2. Limiting the set of experiments

- Why should we not consider a large number of possible experiments rather than just deletions of individual (or a few) genes?
- The experiments must have predictable effects
  - we have to be able to predict likely outcomes from any experiment given only structure $G$ and the parameters
  - this excludes experiments (e.g., temperature change) whose initial effect on the variables is more diffuse (the intervention is not well characterized)
- For simplicity, we will use single gene deletions
  \[
  \text{del}(x_i) \text{ means set}(x_i = 0)
  \]
3. Predict outcomes of deletions

- We have to be able to predict possible outcomes (values for variables) in response to any particular experiment we might carry out

- Given any experiment and a network $G$, we evaluate

$$P(x|\text{del}(x_i), D, G) = \int_{\theta} P(\theta|D, G) \left[ \prod_{j \neq i} \theta_{x_j|x_{pa_j}} \right] d\theta$$

(note that the conditional pertaining to $x_i$ is omitted)
3. Predict outcomes of deletions

- We have to be able to predict possible outcomes (values for variables) in response to any particular experiment we might carry out.

- Given any experiment and a network \( G \), we evaluate

\[
P(x|\text{del}(x_i), D, G) = \int_\theta P(\theta|D, G) \prod_{j \neq i} \theta_{x_j|x_{pa_j}} \, d\theta
\]

\[
= \prod_{j \neq i} \frac{N(x_j, x_{pa_j}) + \alpha(x_j, x_{pa_j})}{N(x_{pa_j}) + \alpha(x_{pa_j})}
\]
3. Predict outcomes of deletions

\[
P(x | \text{del}(x_i), D, G) = \int_{\theta} P(\theta | D, G) \left[ \prod_{j \neq i} \theta_{x_j | x_{pa_j}} \right] d\theta
\]

\[
= \prod_{j \neq i} \frac{N(x_j, x_{pa_j}) + \alpha(x_j, x_{pa_j})}{N(x_{pa_j}) + \alpha(x_{pa_j})}
\]

- Since we are also uncertain about which graph \( G \) is correct, we will have to average such predictions over \( P(G | D) \)

\[
P(x | \text{del}(x_i), D) = \sum_{G} P(G | D) P(x | \text{del}(x_i), D, G)
\]

- This gives our best estimate of the system response to an intervention \( \text{del}(x_i) \).
4. Model refinement

- Each possible outcome $x$, drawn in response to $\text{del}(x_i)$, causes us to update the distribution over graphs according to

$$P(G|D, x, \text{del}(x_i)) = \frac{P(G|D)P(x|\text{del}(x_i), D, G)}{P(x|\text{del}(x_i), D)}$$

- We know $x$ only for experiments actually carried out.
5. Score possible experiments

- The goal is to select an experiment that we expect will reduce our uncertainty about the graph $G$ (not the parameters) the most, i.e., has the highest information gain.

- We can evaluate the uncertainty prior to $\text{del}(x_i)$

$$H(G|D) = - \sum_{G} P(G|D) \log_2 P(G|D)$$
5. Score possible experiments

- The goal is to select an experiment that we expect will reduce our uncertainty about the graph $G$ (not the parameters) the most, i.e., has the highest information gain.

- We can evaluate the uncertainty prior to $\text{del}(x_i)$

$$H(G|D) = -\sum_{G} P(G|D) \log_2 P(G|D)$$

- If we knew the outcome $x$, the uncertainty after carrying out the experiment $\text{del}(x_i)$ would be

$$H(G|D, x, \text{del}(x_i)) = -\sum_{G} P(G|D, x, \text{del}(x_i)) \log_2 P(G|D, x, \text{del}(x_i))$$

- The gain is the difference between these two entropies.
5. Score possible experiments

- We don't know the outcome $x$ and instead have to entertain many possible outcomes, drawn from $P(x | \text{del}(x_i))$.

- We can evaluate the expected information gain (expected value of information) by averaging over possible outcomes:

$$
\text{Gain}(\text{del}(x_i)) = H(G|D) - \sum_x P(x | \text{del}(x_i)) \cdot H(G|D, x, \text{del}(x_i))
$$

- This specifies the average number of bits about the graph that we would get by carrying out experiment $\text{del}(x_i)$.
Example: protein phosphorylation

- Flow cell cytometry data: intensity measurements corresponding to phosphorylation states of 11 proteins across 853 individual cells

<table>
<thead>
<tr>
<th></th>
<th>pref</th>
<th>pmeK</th>
<th>pJNK</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell 1</td>
<td>26.4</td>
<td>13.2</td>
<td>...</td>
</tr>
<tr>
<td>cell 2</td>
<td>35.9</td>
<td>16.5</td>
<td>...</td>
</tr>
<tr>
<td>cell 3</td>
<td>59.4</td>
<td>44.1</td>
<td>...</td>
</tr>
<tr>
<td>cell 4</td>
<td>73.0</td>
<td>82.8</td>
<td>...</td>
</tr>
</tbody>
</table>

- Measurements here are “observational”... our goal is to determine the best intervention to carry out
Preprocessing

- We will need to discretize (log-transformed) intensities appropriately. This step can critically affect the Bayesian networks we would infer.

![Graph showing log-transformed histogram of PKC intensities and kernel density estimate.](image-url)
Preprocessing

- We will need to discretize (log-transformed) intensities appropriately. This step can critically affect the Bayesian networks we would infer.

![log-transformed histogram of PKC intensities](image1.png)  ![kernel density estimate with discretization](image2.png)
Preprocessing

- We will need to discretize (log-transformed) intensities appropriately. This step can critically affect the Bayesian networks we would infer.
Preprocessing

- We will need to discretize (log-transformed) intensities appropriately. This step can critically affect the Bayesian networks we would infer.
Preprocessing

- We will need to discretize (log-transformed) intensities appropriately. This step can critically affect the Bayesian networks we would infer.
Resulting Bayesian network

- Posterior average across Bayesian networks

(Sachs et al.)
Resulting Bayesian network

- Posterior average across Bayesian networks

- Interventions (small molecule inhibitions) by Sachs et al.
Best intervention

- We evaluate the information gain for structures given each possible intervention \( \text{del}(x_i) = \text{set}(x_i = 0) \), \( i = 1, \ldots, 11 \). (the gain written here differently)

\[
\text{Gain}(\text{del}(x_i)) = \sum_{G} P(G|D) \sum_{x} P(x|\text{del}(x_i), D, G) \log \frac{P(x|\text{del}(x_i), D, G)}{P(x|\text{del}(x_i), D)}
\]

where

\[
P(x|\text{del}(x_i), D) = \sum_{G} P(G|D) \sum_{x} P(x|\text{del}(x_i), D, G)
\]

- These are a bit harder (but not impossible) to evaluate exactly in our case. We can evaluate an approximate score by sampling.
Top three interventions

- Posterior average across Bayesian networks
Top three interventions

- Posterior average across Bayesian networks

![Diagram of signaling pathways]

**Inhibition**
- Akt
- PKA
- PKC

**Activation**
- Mek
- Raf

Activators:
1. α-CD3
2. α-CD28
3. ICAM-2
4. PMA
5. cAMP

Inhibitors:
6. G69976
7. AKT inh
8. Psitect
9. U0126
10. LYS94002