Modeling of microarray time course data with dynamic Bayesian networks and within-time-point interaction

Brian Godsey and Peter Sykacek
Vienna Chair of Bioinformatics, Boku University, Vienna, Austria

Motivation

- Main contest data set: detect apoptosis regulators
- Main idea:

State space models detect temporal gene interactions.

But what about very fast interactions?

Speed of genetic interactions

- state space models detect temporal interactions
- this is limited by the time-series data
- very fast interactions are not detected

- clustering addresses this
- closely-interacting genes are in the same cluster
  - clusters can be thought of as processes
- many individuals within a process reduces noise

Strategy

- Affara, *et al*:
  - reduce dimensionality: looked at a subset of genes
  - infer interactions: dynamic Bayesian network (DBN) model
- We:
  - consider fast-interacting genes and reduce dimensionality: clustering genes
  - infer interactions: DBN model with hidden states, fit using variational Bayes (VB) methods

Using clustering to detect quick gene interactions

- k-means clustering on the time profiles
- number of clusters determined by Akaike's Information Criterion (AIC)

Results of clustering:

- 18,451 genes
- 273 clusters
- smallest cluster: 10 genes
- largest cluster: 225 genes

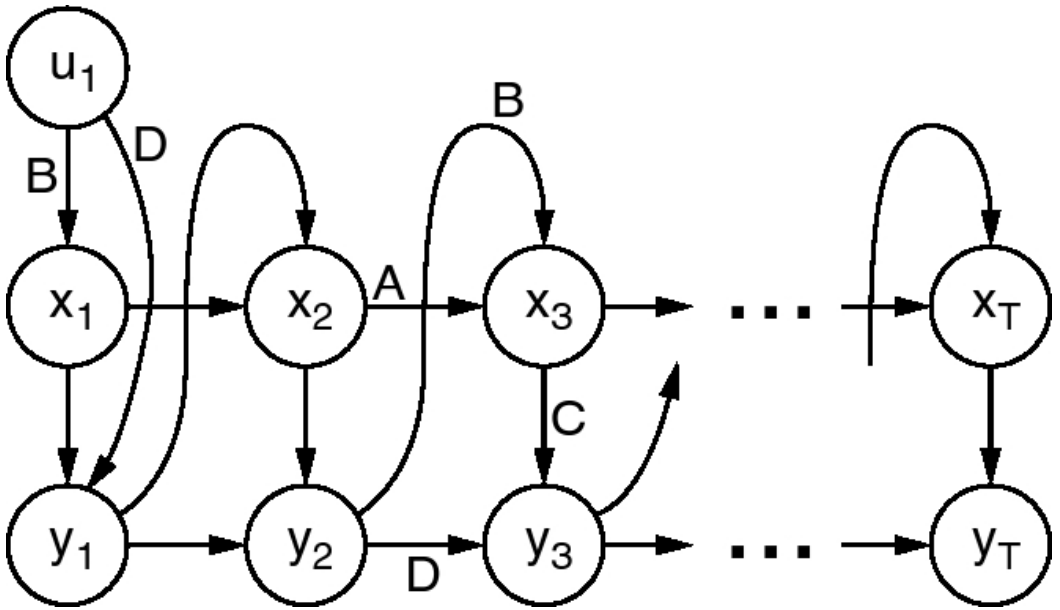
Cluster centers become data points

For fitting the model:

- we assume that the genes within each cluster act together as a process
- the cluster centers are nodes/data points

Detecting interactions between clusters

- DBN model: Beal, *et al* (2004)
  - linear DBN with hidden states, which represent unmeasurable quantities affecting the data
  - visible states are the cluster centers



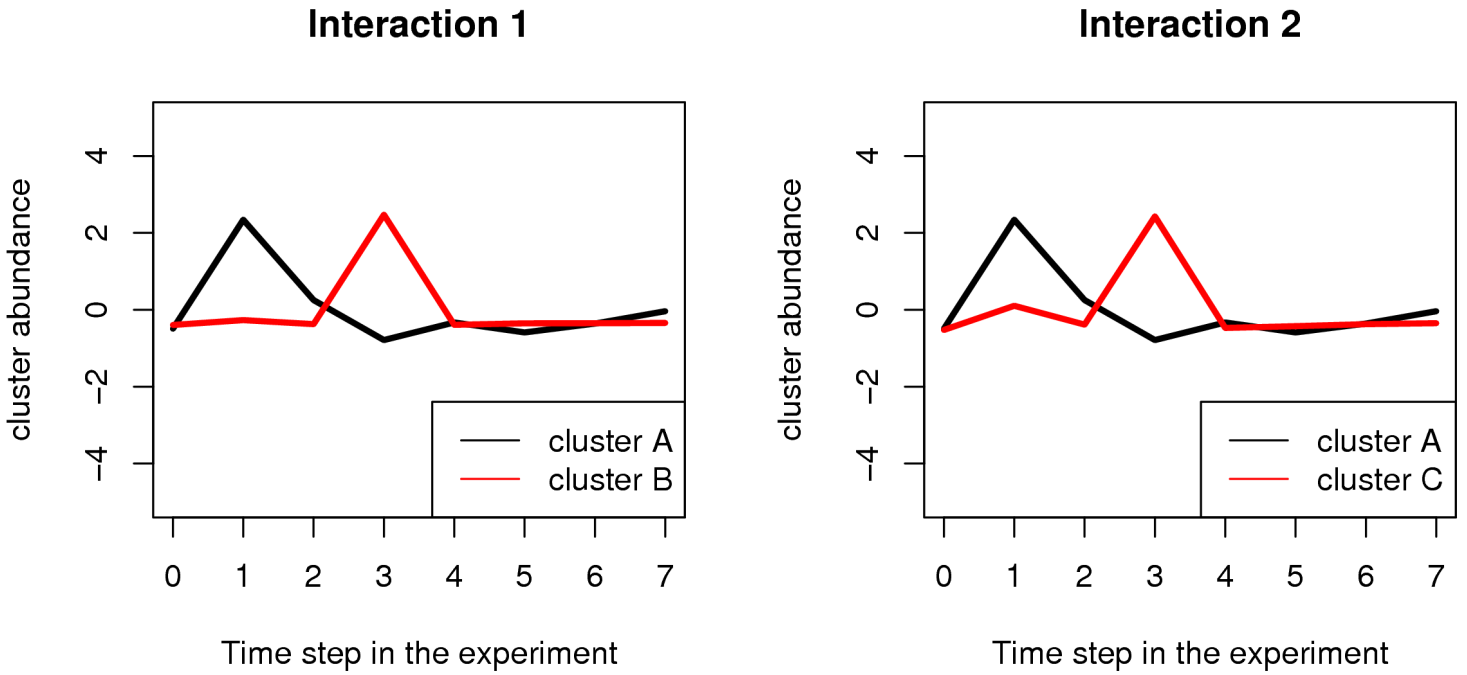
Variational Bayes model fitting

- Hyperparameters:
  - non-informative precisions on the transition matrices
  - this is the “zero prior” from Beal (2004)
- The VB algorithm provides lower bound on the marginal likelihood of the model
- The final results include descriptions of the distributions of these estimated parameters
- Estimating interaction significance:
  - calculate individual p-values from posterior distributions, and then use FDR correction

Results of model fitting

Top 3 interaction coefficients have:

- interaction 1, P<0.00016, FDR P<0.044
- interaction 2, P<0.00324, FDR P<0.443
- interaction 3, P<0.02336, FDR P<0.491



FatiGO Compare analysis

- compare the cluster members against genome
- In the most significant interaction, cluster B:

  top GO term, by P-value:

  apoptotic mitochondrial changes
  (GO:0008637), P<0.00087, FDR P<0.14

Conclusions

- Clustering genes by time profile
  - considers very fast gene interactions
  - reduces dimensionality
  - decreases noise in data for model fitting
- The DBN model
  - detects temporal cluster interactions
- Using both of them together
  - identifies important clusters in the time series

    ...and the genes that are in them