Learning Temporal Sequence of Biological Networks

Le Song

Carnegie Mellon University

Joint work with Eric Xing

December 12, 2008
Temporal Biological Process

Drosophila Life Cycle

Egg → Larva → Pupa → Larva → Egg

Le Song
Learning Temporal Sequence of Biological Networks
Gene Regulatory Networks

Le Song

Learning Temporal Sequence of Biological Networks
Time Series of Gene Expression

Three clusters of genes during the development of *Drosophila melanogaster*
Existing Approaches

- Static networks (ignore temporal evolution)

- Dynamic Bayesian networks (have only one time series)
Our Approaches

- Assumption: genes and their interactions vary *smoothly*.
- Model statistical dependency rather than causality.
- Borrow information across time.
Modeling Statistical Dependency

Gene expression: \( X^t \in \{-1, 1\}^p \) upregulated vs. downregulated.

Dependency graph: \( G^t = (V, E^t) \), \( \theta_{uv} \) for each pair of genes.

Markov random fields (MRFs):

\[
\mathbb{P}_{\theta^t}(X) = \exp\left( \sum_{(u,v) \in E^t} \theta_{uv}^{t} X_u X_v - g(\theta^t) \right)
\]
Models are close between adjacent time points

\[ \| \theta^t - \theta^{t-1} \|_1 \]

Extreme multi-task learning
Temporally fused maximum likelihood estimation

\[ \{\theta_t\}_{t=1}^T = \arg\min \sum_{t=1}^T \left( g(\theta^t) - \sum_{(u,v) \in E^t} \theta_{uv}^t x_u x_v \right) \text{ likelihood} \]

\[ + \lambda_1 \left( \sum_{t=2}^T \|\theta^t - \theta^{t-1}\|_1 \right)^2 \text{ fusion} \]

\[ + \lambda_2 \left( \sum_{t=1}^T \|\theta^t\|_1 \right)^2 \text{ sparsity} \]
Efficient Optimization

- Approximate likelihood using pseudo-likelihood
  \[ P_{\theta_t}(x) \approx \prod_{v \in V} P_{\theta_t}(x_v | x_{\bar{v}}) \]

- Approximate fusion and sparsity term using variational presentation
  \[
  \left( \sum_{t=1}^{T} \| \theta^t \|_1 \right)^2 = \min \sum_{t=1}^{T} \sum_{(u,v) \in E^t} \frac{\theta^t_{uv}}{a^t_{uv}}^2,
  \sum_{t=1}^{T} \sum_{(u,v) \in E^t} a^t_{uv} = 1, \quad a^t_{uv} > 0
  \]
Gene Ontology Groups

Le Song
Learning Temporal Sequence of Biological Networks
Computational and biological questions:

- What is the theoretical guarantee for the learned networks?
- How to incorporate static information (ChIP-chip data, sequence data)?
- How does gene regulation program evolve?
- How does interaction between gene clusters evolve?
- ...
Questions?