Inferring Time-Delayed Gene Regulatory Networks

Presented by: Mina Moradi
Advisor: Dr. Abdollah Homaifar

North Carolina A&T State University
Dept. of Electrical & Computer Engineering
mmoradik@aggies.ncat.edu
http://acitcenter.ncat.edu
February, 17, 2017
Outline

1. Introduction and Motivation
2. Literature Review
3. Objective of the Work
4. Proposed Method
5. Simulations and Results
6. Conclusion
Introduction and Motivation

- Your genes are part of what makes you the person you are. You are different from everyone alive now and everyone who has ever lived.
- Genes are not independent. Regulate each other and act collectively.
- Gene regulatory network (GRN) is an abstract mapping of gene regulations in living cells.
- GRNs identify the specific functional roles of individual genes in cellular systems and can open up a window on the disease progression and drug development.

Therapy: This time it’s personal, Lauren Gravitz, Nature (2014)
Introduction and Motivation

- DNA microarrays and RNA sequencing technologies measure the expression levels of thousands of genes inside cells in response to specific environmental conditions [1].
- GRN is usually represented by a directed graph, with nodes representing the genes and links representing the regulatory relationships.

![Diagram of GRN](image.png)
Reverse engineering of GRNs is a challenging problem due to:

- The stochastic characteristics of biological phenomena, the inherent noise of measured gene expression data, and high dimensionality [2].
- There is strong non-linearity on temporal patterns of regulatory genes [3].
- Genetic interactions among different genes can have different time delays due to the time required for regulatory genes to express their protein products and etc. [4, 5]
Literature Review

- Boolean networks: Based upon binary outcomes (on and off) for gene expression and therefore lack adequate dynamic resolution [6].
- Bayesian networks: Represent probabilistic relationships among genes, the inherent noise and stochasticity of gene expression [7].
- Ordinary differential equations: Deterministic models, where interactions among genes represent causal interactions rather than statistical dependencies [8].
- TD-ARACNE [9]: The time-delayed dependencies between the genes in terms of mutual information by assuming a stationary Markov Random Field as its underlying probabilistic model.
- HCC-CLINDE [10]: Infer a time-delayed GRN in the presence of hidden common causes based on either a correlation test or mutual information test.
Objective of the Work

- Inferring a time-delayed GRN which takes into account the non-linearity of gene interaction and the noise of measurements.
- RNNs are computational tools for temporal data processing, approximating nonlinear patterns and tolerating noise in measurements.
- RNNs are usually considered “black box” models. The internal structure and learned parameters are not interpretable [11].
Proposed a hierarchical RNN (HRNN) that surmounts the interpretation difficulties of the RNNs for modeling of GRN.

Time-delayed regulations can be captured through hierarchical paths between leaf nodes (regulatory genes) and a target node (regulated gene) in the HRNN.

\[ x_1, \ldots, x_C \] are context nodes.
\[ x_{C+1}, \ldots, x_{C+P} \] are genes.
Hierarchical Recurrent Neural Network

- A population of candidate HRNNs are randomly generated.
- The network with $c$ context nodes has $c + 1$ neurons.
- In a network with $c \leq C$ context nodes, the first $c$ context nodes and genes (excluded the target gene) are potential inputs of the neurons.
- The target gene is the output of the first neuron. The context node $c_i$ is the input of $\text{neuron}_i$ and output of $\text{neuron}_{i+1}$.
- If input of the neuron is a context node, the weight is positive.

$$x_i(t + 1) = f\left(\sum_j w_{k,j} x_j(t)\right) \quad (1)$$
Hierarchical Recurrent Neural Network

The corresponding hierarchical model which shows the direct regulation of $x_8$ by gene $x_4$, and time-delayed regulations of $x_7$ by genes $x_4, x_5, x_7$. 
Candidate networks in the GA are represented by their number of neurons \(N_n\), number of inputs to each neuron \(N_{in}\), indices of the input nodes \(In\), weights of the input connections \(W\) and the decay rate of the target gene’s expression level \(\mu\) if it exists.

\[
\begin{align*}
N_n &= c + 1 \\
N_{in}^{(1)} & \quad In^{(1)} & \quad W^{(1)} & \quad \ldots \\
N_{in}^{(c)} & \quad In^{(c)} & \quad W^{(c)} & \quad N_{in}^{(c+1)} & \quad In^{(c+1)} & \quad W^{(c+1)} & \quad \mu
\end{align*}
\]

\[
\text{nodes} = \{x_1, \ldots, x_c, \ldots, x_c, x_{c+1}, \ldots, x_{c+p}\}
\]

\[
In^{(1)} = \{x_1, \text{random}\}, Out^{1} = \text{target gene}
\]

\[
In^{(c)} = \{x_c, \text{random}\}, Out^{(c)} = x_{c-1}
\]

\[
In^{(c+1)} = \{\text{random genes}\}, Out^{(c+1)} = x_c
\]
Fitness of candidate networks

The performance of the candidate networks (fitness) is evaluated by measuring the trade-off between the goodness of fit and complexity of the model by using the Akaike information criterion (AIC) and the Akaike information criterion with correction (AICc).

\[
AIC = n \cdot \ln\left(\frac{1}{n} \sum_l \left( \sum_t \left( x_i^l(t) - \hat{x}_i^l(t) \right)^2 \right) \right) + 2k
\]  

(2)

\[
AICc = AIC + 2k(k + 1)/(n - k - 1)
\]  

(3)

\(k\) is the number of leaf nodes in the HRNN and \(n\) is the total number of temporal samples for gene expression. If \(n\) is small or \(k\) is large, the AICc is preferred rather than AIC. As \(n\) gets larger, AICc converges to AIC.
The Crossover Operator

**Figure: Parent 1**

**Figure: Parent 2**

**Figure: Child 1**

**Figure: Child 2**
The Mutation operator

For a mutation site $m_{site}$ in the network, the mutation works as below:

- If $m_{site}$ is on the number of inputs of a neuron ($N_{in}$), it is mutated to $N_{in} = N_{in} \pm 1$. Therefore, a new input and its corresponding weight are added or deleted.

- If $m_{site}$ is on an input connection of a neuron ($In$), the selected connection is rewired to another node in the network.

- If $m_{site}$ is on a connection weight of a neuron and input is a context node, the Gaussian mutation evolves the weight in the range of $[0, w_{max}]$; else, the weight is mutated in the range of $[w_{min}, w_{max}]$. 
Simulations and Results

The HRNN is evaluated on the GRN of *Saccharomyces cerevisiae* and nonlinear synthetic generated data for different sizes of networks and variances of noise. The results are compared with TD-ARACNE and HCC-CLINDE in terms of:

- **Links:** if and only if both the gene pair and the direction are correct
- **Delays:** if and only if both the link and the time delay are correct
- **Effects:** if and only if both the link and the sign of an effect are correct

For each term, \( \text{Recall} = \frac{TP}{TP+FN} \), \( \text{Precision} = \frac{TP}{TP+FP} \) and \( F\)-score \( = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \) metrics are computed.
The Effect of Network Size

**Figure: 5 genes**

**Figure: 10 genes**

**Figure: 20 genes**

**Figure: 30 genes**
The Effect of Noise Level

Figure: $\sigma^2 = 0.5$

Figure: $\sigma^2 = 1.0$

Figure: $\sigma^2 = 1.5$

Figure: $\sigma^2 = 2.0$
Saccharomyces cerevisiae

IRMA is a recent significant contribution to systems biology reported in [12] where the authors built a synthetic network of the yeast organism \textit{Saccharomyces cerevisiae}.

\textbf{Figure:} True regulations

\textbf{Figure:} TD-ARACNE

\textbf{Figure:} Proposed method

\textbf{Figure:} HCC-CLINDE
<table>
<thead>
<tr>
<th>Methods</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>0.667</td>
<td>0.75</td>
<td>0.706</td>
</tr>
<tr>
<td>TD-ARACNE</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>0.667</td>
<td>0.25</td>
<td>0.366</td>
</tr>
<tr>
<td>HCC-CLINDE</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>0.25</td>
<td>0.125</td>
<td>0.166</td>
</tr>
</tbody>
</table>

*Table*: Comparison of Results for GRN Reconstructions of IRMA.
In this study, we proposed a hierarchical recurrent neural network approach to identify time-delayed regulatory interactions of genes. The designed HRNN facilitates capturing the paths with different lengths from the leaf nodes in the network to the target node. Hierarchy in the network and possibility of recurrent connections in HRNN provide a capability for modeling the temporal patterns of gene expression. The proposed method outperformed TD-ARACNE and HCC-CLINDE in terms of non-linearity and high level of noise in measurements.
References


Acknowledgment

This work is partially supported by the National Science Foundation (NSF) under Cooperative Agreement No. CCF-1029731.

Thank you for your attention