Towards deep mechanistic neural networks for multi-omits modeling.

Ameen Eetemadi
Tagkopoulos Lab, UC Davis
Outline

Introduction
  - Motivation
  - Recent work

Method (TRNN)
  - Approach
  - Architecture
  - Optimization

Results

Conclusion
GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.
GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Image credit: Erik Jacobsen, Covert Lab
GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Synthetic biology
Test gene circuit in host cell without experiment

Genome-scale model

Metabolic engineering
Find environment that maximizes engineered products

Image credit: Erik Jacobsen, Covert Lab
Motivation

GOAL: Predict cellular state and behavior from environmental settings and genetic background of the cell.

Synthetic biology
Test gene circuit in host cell without experiment

Genome-scale model

Predictive medicine
Optimization of antibiotic effect from ~100 existing antibiotics

Metabolic engineering
Find environment that maximizes engineered products

Image credit: Erik Jacobsen, Covert Lab
Motivation

**GOAL:** Predict **cellular state** and **behavior** from **environmental settings** and **genetic background** of the cell.

<table>
<thead>
<tr>
<th>Synthetic biology</th>
<th>Genome-scale model</th>
<th>Metabolic engineering</th>
<th>Predictive medicine</th>
<th>Guided experimentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test gene circuit in host cell without experiment</td>
<td>Predict cellular state and behavior from environmental settings and genetic background of the cell.</td>
<td>Find environment that maximizes engineered products</td>
<td>Optimization of antibiotic effect from ~100 existing antibiotics</td>
<td>Find the conditions with highest uncertainty</td>
</tr>
</tbody>
</table>

Image credit: Erik Jacobsen, Covert Lab
Recent work (ECOMICS) - Data

Most comprehensive compendium for an organism

Recent work (ECOMICS) – Predictive Model

Recent work (ECOMICS) – Predictive Model

How to Integrate:
1) Transcription regulatory network
2) Transcription Thermodynamics
3) Transcription profiles
Approach – Transcription Regulatory Neural Network (TRNN)

How to Integrate:
1) Transcription regulatory network
2) Transcription Thermodynamics
3) Transcription profiles
How to Integrate:
1) Transcription regulatory network
2) Transcription Thermodynamics
3) Transcription profiles
Architecture – TRNN Activation Function (*Alphabet*)

1. Concentration of TFs
2. Computational Unit
3. mRNA concentration (initial prediction)
4. Perturbation
5. Decision Unit
6. mRNA concentration (adjusted prediction)

**Example:**

\[ f(x_1, x_2) = e^{p_1 x_1} e^{p_2 x_2} \]

\[ z = p \times y \]

\[ p \in \{0, 1\} \]
Architecture – TRNN Gene Module (Word)

Words

Diagram showing the architecture of TRNN Gene Module with modules $M_0$, $M_1$, and $M_2$. The diagram includes arrows indicating the flow of information and variables such as $f(x)$, $x$, $y$, $x_1$, and $x_2$. The text includes symbols and notation related to concentration and perturbation.

Example:

\[ f(x, y) \]

\[ x \in [0, 1] \]
Architecture – TRNN Gene Circuit (Sentence)

Sentences

(a) \[ x \rightarrow M_1 \rightarrow M_1 \rightarrow y \]

(b) \[ x \rightarrow M_1 \rightarrow M_2 \rightarrow z \]

(c) \[ x_1 \rightarrow M_2 \rightarrow z_1 \]

(d) \[ M_0 \rightarrow M_1 \rightarrow M_1 \rightarrow z_1 \]
Training Strategy

**Issue:** backpropagation is very sensitive to initial values

**Strategy:**
- **Step 1:** train individual modules separately through a customized LP based method
- **Step 2:** train full network together

Use full-batch, conjugate gradient
Training Individual Modules

Algorithm 1: Fit $f_\theta$, given $C$

Input: $C = \{x_1, x_2, y\}$
Output: $\theta^* = \{t_0^*, t_1^*, t_2^*, b_1^*, b_2^*, p_1^*, p_2^*\}$
1 initialize $p^* = [p_1^*, p_2^*]$
2 while loss$_\theta^*(C)$ is not converged do
3 \[ w^* \leftarrow \text{ArgMin}_w ||A.w - y||^1, \text{ s.t. } b_i \geq 0; \]
4 \[ \theta^* \leftarrow \begin{bmatrix} w^* & p^* \end{bmatrix}; \]
5 for $j \leftarrow 1:2$ do
6 \[ \Delta_{p_j} \leftarrow \frac{\partial}{\partial p_j} \text{loss}_\theta^*(C); \]
7 \[ p_j^* \leftarrow p_j^* - \alpha.\Delta_{p_j}; \]
8 end
9 \[ \theta^* \leftarrow \begin{bmatrix} w^* & p^* \end{bmatrix}; \]
10 end
11 return $\theta^*$

\[
\text{loss}_\theta(C) = \sum_{i=1}^{m}[f_\theta(x_1^i, x_2^i) - y^i]^2
\]

\[ h_1 = e^{p_1 x_1^i}, h_2 = e^{p_2 x_2^i} \]

\[ A = \begin{bmatrix} 1 & h_1 & (-h_1 \odot y) & h_2 & (-h_2 \odot y) \end{bmatrix} \]

$\odot$: entry-wise vector multiplication,
Dataset

• 46 different datasets:
  • Multiple candidate networks extracted from DREAM4 challenge
  • Number of genes for each network range from 3 to 9.
  • Synthetic data generated thermodynamic simulation
    • Gene Knockouts
    • Parameter perturbations (e.g. transcription rate, binding affinity of TFs, etc.)
    • Different noise levels

• Validation
  • Mean Squared Error
  • 5-Fold cross validation
  • Comparison with common ANN architectures
Results – Overall MSE Comparison

Average difference of MSE from best performing method in each dataset
Conclusion

• TRNN, is a deep learning based framework for prediction of steady state mRNA concentration levels under perturbation.
• TRNN, requires specialized training due to it’s architecture.
• Evaluation of TRNN on real data is key next step.
• For scaling up TRNN to genome scale modelling, we will make use of highly parallel computation provided by the cluster.