Towards self-learning agents in era of high-throughput omics

Presenter: Ameen Etemadi
Principal Investigator: Prof. Ilias Tagkopoulos

UC DAVIS
UNIVERSITY OF CALIFORNIA
I use Blue Waters to:

<table>
<thead>
<tr>
<th>1. Design artificial neural networks for gene expression prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Thermodynamic simulations</td>
</tr>
<tr>
<td>• Deep Learning</td>
</tr>
<tr>
<td>• Extensive evaluations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Determine optimal strategies to identify next set of experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Synthetic data generation</td>
</tr>
<tr>
<td>• RNA-Seq data processing</td>
</tr>
<tr>
<td>• Gaussian Processes</td>
</tr>
<tr>
<td>• Extensive evaluations</td>
</tr>
</tbody>
</table>
We have entered a new era ...

Knowledge Discovery

- High-throughput OMICS
- High-performance Computing
- Artificial Intelligence
- Robotic Equipment
Goal: Efficient Knowledge Discovery

Maximize knowledge about *E. coli* Bacteria With Minimum Cost

Applications

- Medicine
- Food Safety
- Basic Science
The Cycle of Knowledge Discovery

- Learn
- Experiment
- Add Observed Data
- Data
- Inteligent Agent
- E. coli
- Lab
Learn dynamic program of a cell

Learn dynamics of gene expression in a cell

Key Features:
- Captures Regulatory Relationships
- Models Transcription Factor Dynamics

Published at:
Bioinformatics Journal, 2018
Learn dynamics of gene expression in *E. coli*
Genetic Neural Network (GNN) is 40% more accurate (for chemotaxis genes)

Dataset size

Mean Absolute Error

Gene Expression (GE) Prediction Error

10 40 70 100

0.15

0.10

Synthetic Data Used
Genetic Neural Network (GNN) is 40% more accurate (for networks with 10-1000 genes)
The Cycle of Knowledge Discovery

Experiment → Learn

Intelligent Agent

E. coli

AddObservedData

Lab → Data
Optimal Experimental Design for Gene Expression Prediction

**Biocides**
- None
- Benzalkonium chloride
- Chlorhexidine
- Chlorophene
- Ethanol
- Glutaraldehyde
- H2O2
- Isopropanol
- Peracetic acid
- Providone iodine
- Sodium hypochlorite

**Antibiotics**
- Ampicillin
- Kanamycin
- Norfloxacine
- Rifampin

**Observed**
- None
- Benzalkonium chloride
- Chlorhexidine
- Chlorophene
- Ethanol
- Glutaraldehyde
- H2O2
- Isopropanol
- Peracetic acid
- Providone iodine
- Sodium hypochlorite

**Observed in Last Iteration**
- None
- Benzalkonium chloride
- Chlorhexidine
- Chlorophene
- Ethanol
- Glutaraldehyde
- H2O2
- Isopropanol
- Peracetic acid
- Providone iodine
- Sodium hypochlorite

**Iteration 2**
- Sampled Conditions (Iteration #1)

**Iterations 3 - 30**
- Sampled Conditions (Iteration #2)
- Sampled Conditions (Iteration #30)

**1. Build Predictive Model**
**2. Calculate Utility Scores**
**3. Select Optimal Conditions**
**4. Perform Experiments**

28 Rounds of OED
Accelerated Knowledge Discovery

Average Gene Expression Prediction Errors
We have entered a new era ...

Knowledge Discovery

- RNA-Seq
- High-throughput OMICS
- Microarrays
- Blue Waters
- High-performance Computing
- Genetic Neural Networks
- Artificial Intelligence
- Optimal Experimental Design
- Robotic Equipment
Blue Water Experience

Our Workload:
• Extremely parallel
• Independent small jobs

Advantages:
• Extremely reliable
• High availability
• Comprehensive documentation

BW Customer Support:
• Fast response
• High quality

Thank You!
Acknowledgements

• Ilias Tagkopoulos, PhD (Principal Investigator)
• Xiaokang Wang, PhD Candidate
• Navneet Rai, PhD
• Beatriz Merchel Piovesan Pereira, PhD Candidate

Funding: