

GRNPar: Parallel Inference of Gene Regulatory Networks Using Boolean Network Models

Anushka Gupta (ag4351), William Das (whd2108)

November 28, 2022

1 Introduction

Gene regulatory networks (GRNs) are graph-like representations of genes prone to activating or inhibiting the expression of other genes in a localized network. Boolean networks are often used to model GRNs—these networks consist of a series of nodes with connections between genes represented as boolean functions. Each node is a boolean variable that encodes its state (0 or 1). Each node’s state is determined by a boolean function that takes as input the states of a subset of other nodes in the network. Researchers have developed algorithms for inferring these boolean functions to model gene regulation activity.

For our project, we aim to implement a parallelized algorithm for inferring gene regulatory networks from gene expression time-series data using the boolean network model. We plan to build off of a simplified brute force algorithm put forth by Akustu and Kuuhara (1999), and optimize key steps in the algorithm, particularly in the computation of possible boolean functions that satisfy the various gene expression interactions in the network.

1.1 Boolean Networks

A boolean network $G(V, F)$ consists of nodes $V = \{v_1, \dots, v_n\}$ and a list of boolean functions $F = \{f_1, \dots, f_n\}$ corresponding to each node. A boolean function $f_i(v_i, \dots, v_k)$ for each node v_i takes as input the states of k other nodes in the network $\{v_i, \dots, v_k\}$.

Each node’s state corresponds to whether or not a gene is expressed—0 encodes a gene that is not expressed, and 1 encodes a gene that is expressed. Each boolean function encodes whether a certain gene activates or inhibits another in the network. Due to their simplicity, these models are particularly useful in modelling the combinatorial logic underlying gene activity.

2 Inference Algorithm

In “Identification of Genetic Networks From a Small Number of Gene Expression Patterns Under the Boolean Network Model,” the authors describe a simplified GRN inference algorithm that takes as input a time-series representation of genes in a network, and determines if there is a boolean network model that is consistent with each pairwise expression in the time-series in the *consistency problem*.

Examples						
	v_1	v_2	v_3	v'_1	v'_2	v'_3
I_1	1	0	0	0	0	1
I_2	0	1	0	0	1	1
I_3	0	1	1	1	0	0

G_1	G_2
$v'_1 = v_3$ $v'_2 = v_2 \text{ AND } (\text{NOT } v_3)$ $v'_3 = \text{NOT } v_3$	$v'_1 = v_3$ $v'_2 = v_2 \text{ XOR } v_3$ $v'_3 = v_1 \text{ OR } v_3$

Figure 1: State values for $(I_1, O_1), \dots, (I_3, O_3)$ and two Boolean Networks G_1 and G_2 . G_1 satisfies all of the pairs of (I_j, O_j) as input whereas G_2 is not consistent with all of the pairs of (I_j, O_j) .

2.1 Consistency Problem

Let n be the number of total nodes, and (I_j, O_j) be a set of corresponding states for $\{v_1, v_2, \dots, v_n\}$ where I_j corresponds to the values of $\{v_1, v_2, \dots, v_n\}$ at time t and O_j corresponds to the values of $\{v'_1, v'_2, \dots, v'_n\}$ at time $t + 1$.

Each node $v_i \in I$ has a Boolean function $f_i \in F$ which results in $v'_i \in O$ where $F = (f_1, \dots, f_n)$. As an example, we have shown a set of three gene expression pairs $\{(I_1, O_1), (I_2, O_2), (I_3, O_3)\}$ in Figure 1. The goal is to determine a set of Boolean functions that satisfy all of the pairs (I_j, O_j) given as input.

The algorithm is as follows. Let $EX = \{(I_1, O_1), (I_2, O_2), \dots, (I_m, O_m)\}$:

Algorithm 1 Inference algorithm

```

procedure INFER( $n, EX$ )
  for each node  $v_i \in V$  do
    for each pair of nodes  $(v_k, v_n)$  where  $k < n$  do
      for each Boolean function  $f$  between  $v_k$  and  $v_n$  do
        if  $O_j(v_i) = f(I_j(v_k), I_j(v_n))$  for all  $j$  where  $1 < j < m$  then
          Assign  $f$  to be the Boolean function for  $v_i$  and output  $v_k$  and  $v_n$  as
input nodes to  $v_i$ 

```

3 Parallelization of Algorithm

We plan to implement Algorithm 1 on a dataset of gene expression time-series data. To parallelize this algorithm, we plan to optimize steps 4-6, which consists of the exhaustive search for boolean functions between two input nodes v_k and v_h .

4 References

Akutsu, Tatsuya, and Satoru Kuhara. "Identification of Genetic Networks From a Small Number of Gene Expression Patterns Under the Boolean Network Model" Pacific Symposium on Biocomputing, 1999. [https://psb.stanford.edu/psb-online/proceedings /psb99/Akutsu.pdf](https://psb.stanford.edu/psb-online/proceedings/psb99/Akutsu.pdf).