

Pot-luck causality challenge: FACT SHEET (for a task solved)
Title: *Reverse Engineering of Asynchronous Boolean Networks
via Minimum Explanatory Set and Maximum Likelihood*

Cheng Zheng

zzhengccheng@pku.edu.cn

Zhi Geng

zgeng@math.pku.edu.cn

School of Mathematical Sciences

Peking University

Beijing 100871, China

Task(s) solved: **SIGNET-- Abscisic Acid Signaling Network**

Reference:

Zheng et al., *Reverse Engineering of Asynchronous Boolean Network
via Minimum Explanatory Set and Maximum Likelihood*

Method:

Summarize the algorithms you used in a way that those skilled in the art should understand what to do. Profile of your methods as follows:

- Preprocessing
First, we combine the state vectors of each adjacent time point to one vector. Then we define whether a variable is different between every two new vector we create. Then for each variable, for every pair of vector (i, j) in which this variable has changed, we get a set of variables that are different between those two vectors $v(i, j)$
- Possible Explanatory Set finding & Boolean Rule Finding
Here we assume that the five root nodes are known. And considering the computational complexity, we limit the parent of each variable no more than 4. If it is not, we may only get that ABA is a root since its value never change. For each variable that are not root, first, we search from the small explanatory set V that can be the cover set of all $v(i, j)$ and then we find the Boolean Rule given the parent of that variable is V using the principle of maximum likelihood. If we find the maximum likelihood is not zero, then we stop search and regard the Boolean Rule as the true Boolean Rule. If the likelihood of all Boolean Rule associate with V , then we search for another V and repeat above process.
- Post Process
If all V with no more than 4 elements can not explain the data, then we select the V and a Boolean Rule associate with it as the true Boolean Rule if the number of vector in the data that contradicts this rule is the least.

Results: The reader should also know from reading the fact sheet what the strength of the method is. To that end, provide a result table:

Table 1: Result table.

The matlab code downloaded from your website can't run on my computer.

Comment about the following:

- quantitative advantages (e.g. compact feature subset, simplicity, computational advantages)

Our algorithm is simple to implement.

- qualitative advantages (e.g. compute posterior probabilities, theoretically motivated, has some elements of novelty).

The finding of Boolean Rule when given a node's parent is based on maximum likelihood which is theoretically consistent. And by computing the conditional probability of each node, we reduce the complexity of enumerating the random update rule for calculating the real likelihood. When the network is not too complex (the parents of each node is less than 5), our method can still be improved by computing the maximum likelihood for all V or making a chi-square test to test whether a Boolean Rule is sufficient to explain the observed data.

Briefly explain your implementation. Provide a URL for the code (if available). Precise whether it is a push-button application that can be run on benchmark data to reproduce the results, or resources such as modules or libraries.

Keywords: Put at *least one keyword in each category*. Try some of the following keywords and add your own:

- Preprocessing: different variable set
- Possible Explanatory Set & Boolean Rule Finding: maximum likelihood