GRAPH-BASED AND SPARSE VARIATIONAL ANALYSIS OF BIOLOGICAL NETWORKS, APPLICATION TO TRANSCRIPTOMICS

Aurélie Pirayre, Camille Couprie, Laurent Duval

IFP Energies nouvelles
1 et 4 avenue de Bois-Préau
92852 Rueil-Malmaison, France

Jean-Christophe Pesquet

Université Paris-Est
LIGM, UMR CNRS 8049,
Champs-sur-Marne, France

ABSTRACT

At IFPEN, biologists work on various micro-organisms in an energetic context: bio-ethanol production. In this particular case, they aim at optimizing fungi strains to improve their production enzymes, used as catalyst(s) in the bio-ethanol production process. For this purpose, enzyme production mechanisms require a better understanding. This kind of information may by encoded as a graph structure whose nodes and edges are respectively derived from genes and their correlations or regulations. The resulting network is called a Gene Regulatory Network (GRN). Indeed, inferring GRNs from gene expression data is especially useful for sketching transcriptional regulatory pathways and helps to understand phenotype variations. However, these graphs, involving thousands of genes, are difficult to visualize or analyze, especially incorporating either experimental uncertainties or additional information retrieved from similar organisms.

1. OVERVIEW

Despite the large number of available GRN inference methods, the problem remains a challenging one due to the underdetermination in the space of possible solutions.

We propose to address this problem with recently developed techniques pertaining to graph (nodes and edges) optimization with sparsity priors, leading to variational analysis based on sparsity-inducing metrics. Those choices are driven by an apparent simplicity of certain basic biological mechanisms in cells, and the ability to operate with conditions allowing variations in the studied enzyme production mechanisms. Given all pairwise gene regulation information available, we propose to determine the presence of edges in the final gene regulatory network by adopting an energy minimization formulation. To refine inference results by restricting the space of possible solutions, additional constraints are incorporated into our models. Some constraints reflecting biological (information about gene interactions) and structural (information about node connectivity) a priori have been developed. Different priors lead to different mathematical properties of the cost function, for which various optimization strategies can be applied. Optimization strategies are inspired by recent graph optimization works in image processing and computer vision, where pixels and their connectivity are used to interpret images at a higher level.

2. REFERENCES

This thesis has produced the following publications

- Journal papers:
  - Biotechnology for Biofuels, December 2014, *Kinetic transcriptome analysis reveals an essentially intact induction system in a cellulase hyper-producer* Trichoderma reesei *strain*
  - submitted in May 2015, under review, *BRANE Cut: Biologically-Related A priori Network Enhancement with Graph cuts for Gene Regulatory Network Inference*
- International conferences with proceedings:
  - presented to BASP (International Biomedical and Astronomical Signal Processing Workshop), Switzerland, January 2015, *Discrete vs Continuous Optimization for Gene Regulatory Network Inference*
  - presented to ICASSP (International Conference on Acoustics, Speech and Signal Processing), Australia, April 2015, *Fast Convex Optimization for Connectivity Enforcement in Gene Regulatory Network Inference*
  - accepted to EUSIPCO (European Signal Processing Conference), France, September. 2015, *Graph Inference enhancement with clustering: application to Gene Regulatory Network reconstruction*
- International conference without proceedings:
  - presented at the workshop satellite of the ECCB (European Conference on Computational Biology), France, September 2014, *Incorporating Structural a priori in Gene Regulatory Network Inference using Graph cuts*

978-0-9928626-3-3/15/$31.00 ©2015 IEEE