ABSTRACT FORM

SUPERVISED NETZORK INFERENCE

Jean-Philippe Vert and Yoshihiro Yamanishi
Ecole des Mines de Paris, France
Kyoto University, Japan

The reconstruction of biological networks such as protein interaction, gene regulatory or metabolic networks from large-scale heterogeneous genomic data generated by high-throughput technologies, including gene sequencing and microarrays, is one of the main challenges of current computational biology. Various approaches have been proposed to attack this problem, including for example probabilistic modeling of genomic data with graphical models or dynamical systems modeling through Boolean or Petri networks.
In this contribution we first argue that in many cases, part of the network to be reconstructed is known, and that the problem can therefore be formulated as a supervised inference problem: given a set of known edges, reconstruct other edges.
Second, as opposed to classical supervised learning algorithm not adapted to this context, we propose to cast the supervised network reconstruction problem as a problem of distance metric learning: given the known edges, how to embed the genes in a Euclidean space in such a way that the resulting Euclidean metric fits the known edges. Once such embedding is performed, the reconstruction of new edges simply amounts to selecting pairs of genes as close to each other as possible.
Third, we propose a new method to perform this embedding when genomic data are represented through a positive definite kernel between genes. This setting has two important advantages: on the one hand, many such kernels have been developed recently and have been shown to be relevant representation of genomic data for various supervised learning problems; on the other hand, such a representation provides convenient ways to integrate heterogeneous data with operations on kernels, such as convex combination.
We illustrate the approach by providing encouraging experimental results on the reconstruction of the metabolic network of the budding yeast.

References:

B. Schölkopf, K. Tsuda and J.-P. Vert, *Kernel Methods in Computational Biology*, MIT Press, 2004.
J.-P. Vert and Y. Yamanishi, *Supervised Graph Inference*, Advances in Neural Information Processing Systems, 2005 (to appeau).