Combining Autosmooth and Alternating Regression for the Estimation of S-System Parameters

Wolfgang Beyer[1], Ana Teresa Freitas[1,2], João Miranda Lemos[1,2], Marco Vilela[3], Jonas S Almeida[3], Susana Vinga[1,4]
[1]INESC-ID R Alves Redol 9, 1000 Lisboa, PORTUGAL, [2]IST, Lisboa, PORTUGAL, [3]Univ Texas - MDAnderson Cancer Center, USA, [4]FCM-UNL, Lisboa, PORTUGAL

Systems Biology is an emerging field that uses a global and integrative perspective to capture the behavior of complex living organisms. An important challenge in this area is the dynamic modeling and optimization of metabolic networks. This task relies on new and innovative approaches such as the availability of accurate time course data of metabolite concentrations in living cells generated by methods of modern molecular biology such as in vivo nuclear magnetic resonance (NMR) measurements.

One of the main tasks is to identify the structure of the network and to estimate the model parameters from the available data. By using the S-systems modeling framework within Biochemical Systems Theory the complex structure identification task can be reduced to a parameter estimation task. S-systems have proven to be flexible enough to be able to handle the most dissimilar metabolic networks and have the advantage that every single parameter can be directly interpreted biochemically, offering insight into the topological structure of the network and into the kinetic orders of the chemical reactions involved. The parameter estimation task corresponds essentially to the numerical solution of a system of nonlinear differential equations, which is both difficult and very time-consuming. This parameter estimation step still poses a significant challenge and is the bottleneck in the whole modeling process.

In this work, two novel methods that aim to simplify the parameter estimation task for S-systems are combined. First, the raw time series data are smoothed using the autosmooth (AS) algorithm [1], which also generates the corresponding derivatives. The smoothed time series as well as their derivatives are then used as input data for alternating regression (AR) [2], which estimates the corresponding S-system parameters. Various scenarios using synthetic data were analyzed in order to judge the quality of the results generated by the combination of AS and AR and in order to find out whether one can expect this combination of methods to be suited to handle real-life experimental data.

The results show that this combination of methods performs very well in a noise-free environment, but when tested in a more realistic case that includes noise, the algorithms exhibit some difficulties. AS and AR are fast compared to other methods which directly estimate systems of differential equations and also deliver results that allow one to reconstruct the behavior of the biochemical system, but when trying to predict the system's behavior under untested conditions, the quality of these predictions is often limited.

Combining AS and AR is a good way to estimate the parameters of an S-system, but is still far from being a perfect solution. A lot of effort in optimization algorithms is still needed in order to automatically model networks from metabolic time series data.

[1] Vilela M, Borges C, Vinga S, Vasconcelos AT, Santos H, Voit E, Almeida JS: Automated smoother for numerical decoupling of dynamic models. BMC Bioinformatics. 2007. In press.
[2] Chou I, Martens H, Voit EO: Parameter estimation in biochemical systems models with alternating regression. Theor Biol Med Model 2006, 3:25.