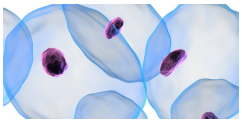
Laboratory of Systems and Synthetic Biology
www.wur.nl\ssb
www.systemsbiology.nl



Leveraging machine learning to speed up construction and analysis of dynamic metabolic models

Supervisors: Rob Smith

Contacts: robert1.smith@wur.nl

Type of thesis: Computational

Required competences: Being able to understand and analyse ODE models will be essential, as taught in SSB30806 and SSB31806. Coding in Python will also be required (e.g. INF22306, SSB30806 or BIF30806). An understanding of the principles of machine learning would be useful but not essential.

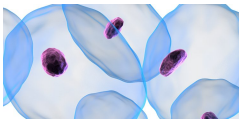
Acquired competences: Use of machine learning for applied tasks. Data-fitting of mathematical models, analysis of their output, and experimental design.

Date: 29-09-22

Description

Mathematical models allow us to test our current knowledge and design new experiments. To build a useful model we require: translation of our current knowledge about a biological system into equations, estimates for reaction rates that cannot be directly measured, and data to compare or test our model. In the model building process, estimating unmeasured reaction rates from datasets is often the most time-consuming step. Current methods require a lot of computational power and the testing of 1000000s of combinations of reaction rates for thorough analysis. Thus, finding ever more efficient ways to test and analyse parameter combinations would be highly desirable.

Recent work has leveraged machine learning to speed up this process. Choudhury et al. have used machine learning to find maximum velocities and equilibrium constants of metabolic systems. Yazdani at al. have used machine learning architectures to find optimal reaction rates to describe time-dependent data. Whilst, Rackauckas et al. have used machine learning to find optimal time-dependent reaction rates. All of these show promise for finding optimal reaction rates for a model, given time-dependent data.



However, these methods have not sped up the post-hoc analysis of mathematical models and experimental design.

In this project, I wish to test whether we can use machine learning to find optimal reaction rates of a mathematical model *and* perform model analysis at the same time, leading to the design of new experimental hypotheses. I would also like to push this idea to its limit – what is the least amount of information needed for us to approximate "true" biological systems using machine learning algorithms? We will start our testing on a small synthetic network and, if the machine learning parameter analysis is successful, we will apply the methodology to a larger model of *E. coli* metabolism.

References

Rackauckas et al. (2020) `Universal differential equations for scientific machine learning', *arXiv*: 2001.04385

Yazdani et al. (2020) 'Systems biology informed deep learning for inferring parameters and hidden dynamics', *PLoS Computational Biology,* doi: 10.1371/journal.pcbi.1007575

Choudhury et al. (2022) `Reconstructing kinetic models for dynamical studies of metabolism using generative adversarial networks', *Nature Machine Intelligence*, doi: 10.1038/s42256-022-00519-y