The M.Sc thesis by Teemu Rintala presents a new machine learning algorithm that combines the effectiveness of deep learning and the interpretability of logic programming to learn complex behaviours in the form of logical statements like: if A and B are active while C is inactive then D will activate. Such logical models are crude analogies to actual biochemical systems, but have historically been useful providing holistic insight into biological systems.

Traditionally researchers would design experiments to identify specific interactions and build complete network models more or less by hand. Statistical methods can be used to establish correlations, while kinetic models informed by correlations can be fitted using regression. Otherwise expression patterns can be predicted using various machine learning methods. However there is a downside: the predictions are like an oracle, for unseen circumstances it is difficult to know if the predictions are correct and the reasoning is often obscure.

On the other hand logical models have the benefit of being relatively easy to understand, but they have issues with large networks due to the exponential number of possible explanations. Deep learning or differentiable programming is able to explore mostly good approximate explanations by working in a relaxed continuous model -- a neural network -- rather than clunky logic. Neural networks are accurate, but very opaque oracles. However, differentiable neural networks can be demystified somewhat with the help of sensitivity analysis. This informs us about which elements are important in determining predictions. Sometimes this is enough but when the interactions between different elements are important, we can go further and form a new transparent model based purely on predictions given by the oracle. Given binary inputs and outputs we can form a truth-table by repeatedly querying the oracle for every possible combination of the inputs. The truth-tables can often be simplified into shorter Boolean formulae, which can be read by humans in the form of statements.

The presented algorithm works by learning a neural network from binary time-series data and extracting relatively small truth-tables based on predictions made by the neural network. The choice of variables for the truth-tables is controlled by sensitivity analysis. This makes the algorithm faster than most exact algorithms for learning logic based models.

More information:

Teemu Joel Rintala, Aalto University School of Science
teemu.rintala.a at gmail.com