Prediction of Compound-Protein Interactions based on Deep Learning

Masatoshi Hamanaka, Kei Taneishi, Hiroaki Iwata, and Yasushi Okuno
[1] Kyoto University, [2] RIKEN, [3] Foundation for Biomedical Research and Innovation

As the number of potential compound-protein interactions (CPIs) that could be assayed is essentially infinite, a brute-force experimental screening approach for CPIs is highly wasteful. Attention has thus been focused on CPI prediction models that can guide researchers to fast lanes for hit discovery.

Existing CPI prediction models have mostly used a curated database of interactions for building a single fixed model, with the Support Vector Machine (SVM) often used for model construction. On datasets of 100,000 CPIs, the SVM can train a model in less than one day. Yet the size of available datasets can be in the millions, and since SVMs require an exponential increase in resources, model construction on such large datasets is infeasible.

In light of this, we have investigated the ability of deep-layered neural networks, also known as Deep Learning, to handle such large volumes of CPIs that cannot be readily processed by SVMs. In this modelling process, many input and output layers are chained together before a final prediction output layer. Deep learning does not require learning on all input data at once as in the standard SVM, but rather the model is generatively tuned over the course of data input.

Research into deep learning for CPI modeling is still in its infancy, and in this presentation we share key results and insights obtained thus far. We model interactions of GPCRs using a 1974-dimensional vector split approximately equally between compound and protein descriptors. We test the effect of both the number of intermediate layers and units on prediction performance. Further, to generate expectations for learning from 4 million of CPIs, we evaluate the learning error rate as a function of the number of learning iterations, where each iteration uses a small subset of the total CPI space available.