

Recurrent Convolutional Neural Network in Genetics

Lecturer: Sebastian Proft

Maximum number of participants: 3

Period/preliminary appointment: 19.04.2021 - 11.06.2021

Location: Charité/BIH - Bioinformatics & Translational Genetics, Invalidenstr. 97

Short description:

In this project, your task would be to familiarise yourself with a Convolutional Neural Network (CNN) that is used to predict the binding sites of transcription factors binding sites (TFBS). CNN's are one of the most widely used machine learning methods in the current age. CNN's use the mathematical methods of convolution to have a NN learn more generalised information about the training data. In the example of TFBS, the task of the convolutional network would be to learn the binding site motifs that the transcription factors bind to. After getting acclimatised with the code you will design a Recurrent Block that should help improve the predictive power of the NN (RCNN). What has been shown to work nicely for genomic data are Long short-term memory (LSTM) architectures. These architectures allow the Neural Network to have feedback connections instead of only feedforward connections. These connections allow the network to better adapt to time series data like video, speech or sequences.

Biological topics:

- Transcription factors and their binding sites
- Prediction of the effect of variants in the binding sites on binding

Informatics learning goals:

- Using Tensorflow for designing artificial neural networks
- Analysing what the networks has learned
- Python programming skills

Quantitative allocation (in %):

Practical programming work: 80%

Soft Skills: 20%

Programming language(s) used: Python

Difficulty level:

A Programming	*****
B Biology/Chemistry	**
C Projekt management	*

Required previous knowledge:

- Interest in Neural Networks
- Interest in genetics
- Some experience in Python

Contact address, website / link: sebastian.proft@charite.de