MSc project: Using (hybrid) neural networks for scientific discovery in biological learning research

Federico Mancinelli, Dominik Bach, Holger Fröhlich

Hertz Chair for Artificial Intelligence and Neuroscience, University of Bonn; Fraunhofer Institute for Algorithms and Scientific Computing; Bonn-Aachen International Center for IT, University of Bonn

February 2024

Abstract

Biological learning research seeks to discover how biological agents learn about their environment. Learning is the core principle of clinical interventions in psychiatry (e.g. psychotherapy) and neurology (e.g. rehabilitation). A better understanding of biological learning is likely to have real-world implications in clinical settings.

Classically, biological learning theory is based on hand-crafted analytical models of putative cognitive processes, such as Bayesian inference models, modified Kalman filters, or ad-hoc difference equations. Recurrent neural networks provide a biologically inspired solution to the same learning problem, but are typically data-hungry and less interpretable.

In this project, you will train different neural network (NN) architectures for time series data (e.g. LSTMs, GRUs) on synthetic and real data from biological agents performing learning tasks, where the NN input is a symbolic representation of the task structure, and the output is the agent's behaviour. Our team has gathered several large data sets from human agents, as well as a large library of task settings and idealised behaviour from biological agents, both of which you will leverage in the project. The project will explore (a) the sample sizes required to train different NN architectures, depending on noise structure in the data, (b) generalisation of trained NNs to unseen problem settings, (c) the use of various approaches to rendering the NN interpretable, and thus suitable for the formulation of novel scientific hypotheses about the underlying biology.

In summary, you will gain a thorough understanding of NNs in scientific discovery, practical application of various NN architectures, and challenges of NNs in real-world application settings.

Example literature:

Dezfouli, Amir, et al. "Disentangled behavioural representations." Advances in neural information processing systems 32 (2019).

Eckstein, Maria K., et al. "Predictive and Interpretable: Combining Artificial Neural Networks and Classic Cognitive Models to Understand Human Learning and Decision Making." bioRxiv (2023): 2023-05.

Song, H. Francis, Guangyu R. Yang, and Xiao-Jing Wang. "Reward-based training of recurrent neural networks for cognitive and value-based tasks." Elife 6 (2017): e21492.