Master Project

Autonomously Learning Dynamical Systems

Peyman Mohajerin Esfahani, Delft Center for Systems and Control, TU Delft
P.MohajerinEsfahani@tudelft.nl

Martijn Wisse, Cognitive Robotics, TU Delft
M.Wisse@tudelft.nl

Context

In attempt to understand the most complex organ known to mankind, the human brain, neuroscientists have long been searching for a unifying theory. It has become clear during the last decades that the brain relies heavily upon predictive capabilities. Recently, a new theory with unparalleled unifying strength and explanatory power about the brain's functioning has emerged: Hierarchical predictive coding.

The main notion within the theory postulates that the brain constantly minimizes prediction error between sensory input and predictions about that input. The predictions are generated in top-down neural connections and compared with bottom-up errors, see Figure 1. Concisely, minimizing the prediction errors at all levels enables the brain to recognize the causes of the input.

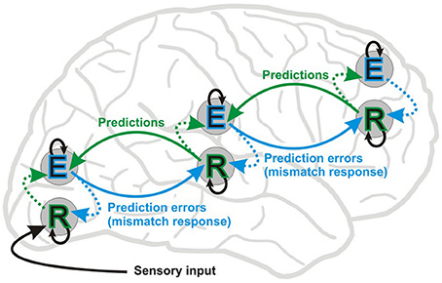

Figure 1: Schematic representation of the prediction error mechanism in human brain

Besides perception, also action can be integrated into the framework, creating active inference. Precisions of predictions can be improved through action, helping to minimize prediction errors further. This creates a system with both in- and output, rendering active inference suitable for control applications.

As biological mechanisms often provide a basis for efficient engineering implementations, it is worth attempting to translate the hierarchical predictive coding framework into a new

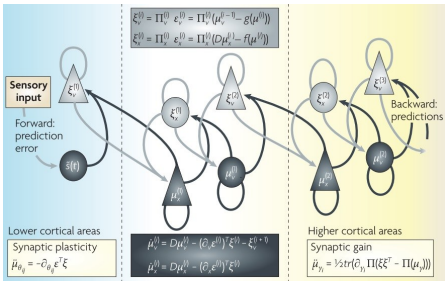

Figure 2: The prediction error is minimized at each level, providing the basis of perception and action.

control or learning algorithm which uses prediction errors as fundamental variable. One of the founders of the active inference framework, Karl Fristion, has mentioned that active inference might have potential to become the new artificial intelligence (clearly it is aptly named).

Project tasks

The goal of the research is to develop an unsupervised learning algorithm based on prediction error minimization. The algorithm is to be implemented on robotic systems (initially cars). We hypothesize that the algorithm will learn to control complex systems better than currently existing approaches such as reinforcement learning or optimal control.

This thesis is part of the initial studies attempting to use the predictive brain theory as a bio-inspired approach for new control and/or learning algorithms. The aim of this thesis to investigate feasibility of the prediction error minimization approach in robot learning. The goal is to create interesting behavior on some benchmark system as basis for further research.