Project Proposal

黄越 张卓豪

1. Introduction

The brain has a collection of high-order functions, guiding us to feel, act and learn. However, the brain is so complex that we know little about how brain learn new functions. Recently, *in vitro* biological neural network (BNN) was shown to be able to play Pong within a close-loop training system[1]. Previous to this work, scientists used BNNs to perform some easier tasks such as moving a cursor in a certain direction[2]. In these works, they all used multielectrode arrays to stimulate and record from the cells. Signals from recorded cells are used to manipulate a cursor or a paddle. Feedback from the task or pong game will be converted to electrical signal to stimulate the cells via electrodes. If the task is done well, "reward signal" would be delivered to the cells and otherwise "penalty signal". The neural network is believed to change its synaptic connectivity according to the feedback signal and perform better in later trials.

The choice of feedback signal is a little tricky. Some work[1] believed neural network would obey the free energy principle[3]. This theory indicates that the neural network would minimize the difference between its predictions about the environment and observed sensations (prediction error). Therefore, they used regular stimuli as "reward signal" and noise as "penalty signal". Others used "patterned training stimuli" as "penalty signal" to induce network plasticity and shuffled background stimulation as "reward signal" to stabilize accumulated plasticity[2]. The applied "patterned training stimuli" would induce network plasticity, but how much the plasticity is and the result from the plasticity is largely unknown and unpredictable.

According to different purposes, these two experimental studies were pre-designed with a mode of feedback signals respectively. However, it is not clear whether a more appropriate feedback mode exists. Thus, a better strategy to derive "reward signal" and "penalty signal" is needed.

Based on the experiments[2], computational models were derived and similar results were observed using spiking neural network with spike-timing dependent plasticity[4]. Here, we try to use the framework of reinforcement learning to find the optimal "reward signal" and "penalty signal" for a given spiking neural network.

2. About implementation

Here is the plan:

(1) Reproduce the model and core results in the paper[4].

(2) Apply the framework of reinforcement learning to find the best strategy for delivering reward and penalty.

For reinforcement learning

Observations: the position and velocity of cursor (which are derived from the signals recorded by electrodes), success or failure.

Reward: could be defined based on the difference between observation and the target, or based on success or failure.

Action: feedback stimulus delivered to the network (electrodes, frequency, temporal patterns, amplitude…).

Reference

1. Kagan, B. J. *et al.* In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron* (2022) doi:10.1016/j.neuron.2022.09.001.
2. Bakkum, D. J., Chao, Z. C. & Potter, S. M. Spatio-temporal electrical stimuli shape behavior of an embodied cortical network in a goal-directed learning task. *J. Neural Eng.* **5**, 310–323 (2008).
3. Friston, K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* **11**, 127–138 (2010).
4. Zenas C. Chao, Bakkum, D. J. & Potter, S. M. Shaping Embodied Neural Networks for Adaptive Goal-directed Behavior. *PLoS Comput Biol* **4**, e1000042 (2008).