R01/Sarthak Mittal/200050129

January 7, 2023

In recent years, the medical community has shown interest in treatment strategies relying on automated prediction methods. However, due to lack of appropriate **sequential data** (key to applying reinforcement learning), the idea of applying reinforcement learning to optimize treatment strategies is relatively novel. This paper examines the problem of applying **reinforcement learning** to optimize control strategies for *deep-brain electrical stimulation* in the treatment of **epilepsy**. To this end, they begin by investigating the use of **batch** reinforcement learning to learn from *in-vitro* studies of stimulation.

In many medical domains, it is **not possible** for the agent to **interact** with the environment dynamically and update its policy after each action. Due to this, it is preferable to utilize a *batch* mode that uses a series of previously recorded (s,a,r) trajectories. They make use of **Fitted-Q Iteration** algorithm because it has been shown to make efficient use of training data. For the supervised regression algorithm required to learn the $\hat{Q}_N$ function, they use **Extremely Randomized (Extra)** trees due to it exhibiting excellent performance relative to other regression trees.

They study the field potential recordings of seizure-like activity recorded in *perirhinal cortex* of rat brains. To validate their optimization method, they rely on simple **empirical indicators** calculated using a hold-out test set, because the *in-vitro* testing is labour-intensive and there are no good generative models of temporal-lobe epilepsy for an *in-silico* test. To counter the fact that the **target policy** cannot be applied on a test set collected under a *given policy*, they use a form of **rejection sampling** to select only those segments of the test set that are consistent with target policy. They use **4 scores** to quantify the performance - an estimated proportion of seizure states when following a particular strategy, the number of actual electrical stimulation events used over the test trace, the expected immediate rewards and the expected return. They contrasted their **tree-based RL (TBRL)** approach with a **neural-net-based RL (NNRL)** approach using these evaluation metrics.

The results of their experiments indicate reduction of the total electrical stimulation to the brain by **10 times**, while reducing incidence of seizures by **25%** compared to current best simulation strategies. If the results carry over to the human model, this could mean a longer neuro-simulator battery life (installing a new one requires *surgery*) and better quality of life of the patient.