R6/Sanit Gupta/160100010

August 6, 2020

This is a paper about a potential application of reinforcement learning. The aim is to use reinforcement learning to develop policies, mappings from EEG signal to applied stimuli, that optimally deal with seizures while keeping the amount of simulation applied low (this is done in order to minimize risk of damage to brain tissues and also helps by increasing the battery life of the stimulator). Although there is nothing very novel about the algorithms/techniques used, I believe this is an important problem because if successful this would make the lives of many epilepsy patients easier. Their results are also very impressive as they are able to reduce the incidence of seizures by 25% and the amount of simulation applied is now one-tenth the older strategies. It needs to be noted though that these results weren't obtained on actual live patients but using cross validation on the available data.

One issue with solving this problem is that it can't be done online. So, batch reinforcement learning using data of previous patients is used. The state space is a summary of past EEG activity and they create a reward function according to their perception of cost of stimuli and seizures. They use Fitted Q Iteration, a popular batch reinforcement learning algorithm. Also, this algorithm needs a regression model to learn the Q_N functions. Here, Extremely Randomized Trees are used.

There are some difficulties with the evaluation of the learnt policy as it can't be done online. They solve this by just using the trajectory data they have and using rejection sampling.

A few questions I have regarding the paper:

- What happened when these learned policies were tried online?
- Was there any basis to their design of reward function and choice of discount factor? Shouldn't the discount factor ideally be very high?