

Simulating Hyperactivity in ADHD using Reinforcement Learning

Patrick Connor¹, Laurent Mattina², and Thomas Trappenberg (P)¹

¹ Faculty of Computer Science, Dalhousie University, Halifax, Canada

²ENSTA-Bretagne, Brest, France

E-mail: patrick.connor@dal.ca, laurent.mattina@ensta-bretagne.fr, tt@cs.dal.ca

Abstract—Hyperactivity is a key symptom in those diagnosed with Attention Deficit Hyperactivity Disorder. In the present work, we model hyperactivity in terms of a two-armed bandit task from Reinforcement Learning, where initial state-values are set abnormally high. Extinction of these state-values when neither action is very rewarding induces repetitive switching between actions over a series of trials with a frequency that is proportional to the initial state-value. Here we propose that although setting initial values may be a useful exploration strategy, switching can become overly frequent or “hyperactive” when they are set too high.

Keywords—ADHD, Hyperactivity, Reinforcement Learning, Exploration Strategy

1 Introduction

Attention Deficit Hyperactivity Disorder, or ADHD, is a developmental disorder which, as the name suggests, involves the presence of an attention deficit and hyperactivity. In this paper we decided to focus on the latter symptom. A classic example is a young student in class. Several actions can be exercised, such as “listening to the teacher”, “chatting with classmates”, “playing with a pen”, etc. While every child (and adult) struggles to maintain focus and sometimes switches between possible actions, hyperactivity could be characterized by more frequent switching. We can define this scenario in terms of a Markov Decision Process (MDP), which describes how an agent (i.e. someone or something) acts in an environment. Our agent begins in the “seated in class” state and, from here, explores its environment by moving to neighbouring states, which requires taking corresponding actions (i.e. “listening”, “chatting”, “playing”). After taking an action, the MDP agent learns something about the rewarding nature of that move and is returned to its initial state, which allows it to choose again, this time more informed about the value of potential actions. Remaining focused is analogous to the repeated selection of the same action, while hyperactivity could be characterized by a high rate of switching between actions.

2 The Optimistic Initial Value Exploration Strategy

Common to many Reinforcement Learning (RL) algorithms is the question of exploration strategy. Exploring is simply trying a variety of actions from each state to help the agent discover rewarding paths. Some RL approaches involve a probabilistic policy, meaning that there is randomness in the choice of actions [1]. In addition, RL approaches often use probabilistic outcomes, where the actual outcome of taking an action is not always the expected one. For example, if the selected action, or policy, for a certain state is to “move left”, the agent will sometimes end up to

the right of its present position. Although these approaches can be effective and reasonable for exploration (we do not always choose rationally or arrive where desired), we consider an alternative. In an MDP problem formulation, the initial value of taking a specific action, or the action-value, is usually set to zero or some small negative amount. Instead, if the action-values are given equal optimistic (positive) initial values [1], an interesting exploration strategy naturally emerges. Consider an agent in an initial state, having to choose its next action. It will do this by selecting what it believes is the most rewarding action. When the outcome of the action is less than expected, that action will lose value, lowering itself below other possible actions. As a result, the agent will choose a different action from that state next time. If all actions lead to less than expected outcomes, the agent will systematically explore all actions from that state.

3 Hypothesis and Model

As discussed earlier, if hyperactivity may be characterized by frequently switching between behaviours or, in RL terms, changes in the policy, hyperactivity may simply be the result of having an above average initial value for potential actions from a given state. Having higher initial state values (say, twice as much for hyperactive agents than healthy agents), will lead to steeper decreases in action-value when the action taken is unrewarding. The hyperactive agent would thus switch more frequently. Note that it was this prospective mechanism rather than a clinical finding which motivated this work.

In order to simulate this situation, we built a simple MDP with only three states: the initial state and two terminal states, which is commonly referred to as the “two-armed bandit problem” [1]. From the initial state, the agent must pick one of two actions, which move it to one of the two terminal states, which we have chosen to give zero (neutral) reward value. Exploring the environment is thus reduced to taking both actions and discovering the most valuable state amongst these two terminal states. In our setting, the action-values are initialized to some positive value. Then, for each trial, the agent will pick the most valuable action and, because it receives zero reward for its efforts, decreases its value for that action. It will pick the same action until its decreasing action-value dips below the other action’s value, where it will switch to the other action.

The update of action-value is based on an algorithm called Q-Learning [1]. When using this approach for the two-armed bandit case, the action-value update rule can be simplified to:

$$Q(s, a) \leftarrow Q(s, a) + \alpha[R(s') - Q(s, a)], \quad (1)$$

where s is the initial state, a is the action chosen, $Q(s, a)$ is

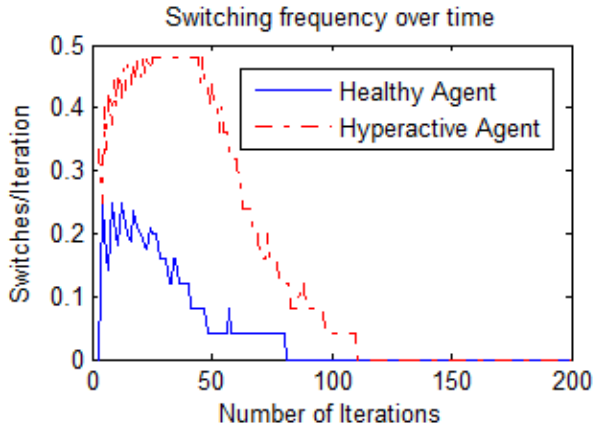


Figure 1: Switching frequency over time with a threshold. Under this amended hypothesis, the higher the initial value, the more frequent the switching.

the action-value, α is the learning rate (we used $\alpha = 0.1$), and $R(s')$ is the reward received at the next state, s' .

The policy (i.e. the action to choose) changes when the value of taking the most recent action sinks below the other action's value. Note that we focus on the use of action-values, but a form using state-values would work similarly.

4 Results

Initial results (not shown) were unexpected, although reasonable. The rate of switching was the same for both the agent with high initial action-values ($Q(s, a) = 10$) and the agent with low initial action-values ($Q(s, a) = 5$). Indeed, both agents switched at exactly the same time.

Although unexpected, these results make sense. No matter the initial value used, one action loses value as fast as the other and eventually switching occurs with almost every trial. One possibility, however, is that switching between actions incurs a cost. Although two actions may have similar action-values, the action which had been taken last time may be selected because it is “easier” than switching to the other. This can be expressed by introducing a threshold to the switching operation. The action-value of the recently selected action must then drop by a certain amount (the cost) below the next highest action-value before a switch will be made. That is, a switch will be made when taking the alternate action is more valuable than the latest action, even after accounting for the effort required. Consider the terminal states A and B, and their action-values $Q(s, a_A) < Q(s, a_B)$, for example. With the initial hypothesis, the switch occurred as soon as $Q(s, a_A) > Q(s, a_B)$. Now, with the threshold, $\theta = 0.5$, it occurs when

$$Q(s, a_A) > Q(s, a_B) + \theta \quad (2)$$

Since the initially higher action-values of the hyperactive agent decrease faster according to Q-Learning, the threshold is reached more readily and thus switching occurs more frequently, as shown in Figure 1.

5 Discussion, Conclusions, and Future Work

We have suggested a mechanism to explain hyperactivity. Neural correlates may involve the basal ganglia (BG),

where reinforcement learning is believed to take place [2]. One prediction from this model is that hyperactivity should decrease if the value-extinguishing learning rate (α) is made to decrease. This could be done by increasing extracellular dopamine. In the striatum, a region of the BG, direct pathway neurons promote actions and have been implicated in representing action-values [2]. From Shen et. al. [3], it appears that these neurons lose synaptic strength when active with low dopamine, but gain strength with high dopamine. Thus, perhaps a higher dopamine level hinders these neurons from extinguishing their activity and thus slows action-value reduction and switching. One study [4] found that extracellular dopamine is increased with therapeutic dosages of oral methylphenidate, a key drug used to reduce the symptoms of ADHD.

Another interesting possibility along this theme is that our simulation (with cost thresholds incorporated) will give similar results with equally optimistic agents when the hyperactive agent has a larger α . Below normal dopamine levels in hyperactive people should facilitate this through greater reduction of synaptic strengths of direct pathway neurons, as interpreted from Shen et. al [3].

Another prediction of this mechanism is that hyperactivity will decline with the number of trials or time spent in a certain state. This is counterintuitive since, for example, it seems that students naturally get more hyperactive with time and not less. It may be that the initial action of “listening to the teacher” is rewarding enough to sustain that activity and only after a period of time does its value decrease to the level of other actions. These other actions then begin to take part and hyperactivity increases. However, if given enough time, perhaps all actions will get boring, and switching will slow.

References

- [1] Sutton, R., Barto, A. (1998). Reinforcement Learning: An Introduction.
- [2] Samejima, K. (2005). Representation of Action-Specific Reward Values in the Striatum *Science*, 310, 1337-1340.
- [3] Shen, W., Flajolet, M., Greengard, P., Surmier, D. J. (2008). Dichotomous Dopaminergic Control of Striatal Synaptic Plasticity, *Science*, 321, 848-851.
- [4] Volkow, N., Wang, G., Fowler, J., Logan, J., Gerasimov, M., Maynard, L., Ding, Y., Gatley, S., Gifford, A., Franceschi, D. (2001). Therapeutic Doses of Oral Methylphenidate Significantly Increase Extracellular Dopamine in the Human Brain *J. Neurosci.*, 21, RC121.