Reinforcement Learning: Artificial Feeling

Phani Ram Sayapaneni
Department of Computer Science
University at Buffalo
Buffalo, NY 14226
phaniram@buffalo.edu

Abstract

In this paper, we demonstrate that effects of emotions on learning could be in fact implemented during the reinforcement learning process. This enables any AI agents to not just learn the actions from rewards, but also feel the experience of winning vs losing during the training process.

1 Introduction

Recent innovations in deep reinforcement learning has enabled us to train AI agents to learn specific tasks [2]. These agents, indiscriminately learn the task, with out any feeling/emotion. They learn the task as if their performance, experience during the learning process doesn't matter to their state at all, this is highly in contrast to how humans learn. Human beings feel the learning process, they physically feel the winning rush or losing atrophy, which is missing in robots/AI agents. The goal of this paper is to develop such experience for AI agents.

2 Artificial Feeling and Implementation

Dopaminergic Neurotransmission plays a key role while learning any experience. Also, it has been mentioned [3] that dopamine actually modulates the learning and working memory in the prefrontal cortex of mammals, birds. This physical distinction is important, since our ability to learn new skills faster and our working memory capacity both are depreciated when we experience failure.

Here, we recreated this effect by considering Replay Buffer as a representation for the Working Memory.

Usually the replay buffer is fixed during the entire life time of an AI agent. But here, we express the size of replay buffer as the function of recent experience. If the agent has been performing well in the recent n experiences, the size of replay buffer is increased. Similarly, if the agent has been performing poor in the recent n experiences, the size is decreased.

$$S_{\mathrm{rb}}\,(t) = F_{\mathrm{rb}}\,(r_{\mathrm{t-n}}\,, r_{\mathrm{t-n+1}}, \dots \dots r_{\mathrm{t}}) \qquad \forall\, t \in [0,\, \infty]$$

Here, S is the size of replay buffer, F is the function that encodes this functionality and t is the time step during which the agent is interacting with environment and n is a hyper parameter. The function F could be any function, we have chosen a linear function which is a fractional change in sum of rewards between recent n/2 experiences and it's next n/2 experiences.

Similarly, same functionality has been applied to the learning rate parameter in the

Optimizer.

$$S_{\mathrm{lr}}\ (t) = F_{\mathrm{lr}}\ (r_{\mathrm{t-n}}\ , r_{\mathrm{t-n+1}}, \ldots \ldots r_{\mathrm{t}}\ ) \qquad \forall\ t\ \in [0,\ \infty]$$

The Replay Buffer, Optimizer are the considered as the important physical resources for the AI agent during it's learning process.

3 Results

The below figure shows the learning curve obtained from training the agent in an Atari Break Out environment. The green curve represents the learning curve obtained from a stable model, where as the red curve represents the learning curve obtained from the new model(feeling).

We can observe that these curves diverge after few experiences i.e., when the mean reward per episode is substantial (greater than 4 in our case). We expect to uncover instances of learning experiences which improve the learning outcomes for only positive ones as opposed to all experiences.

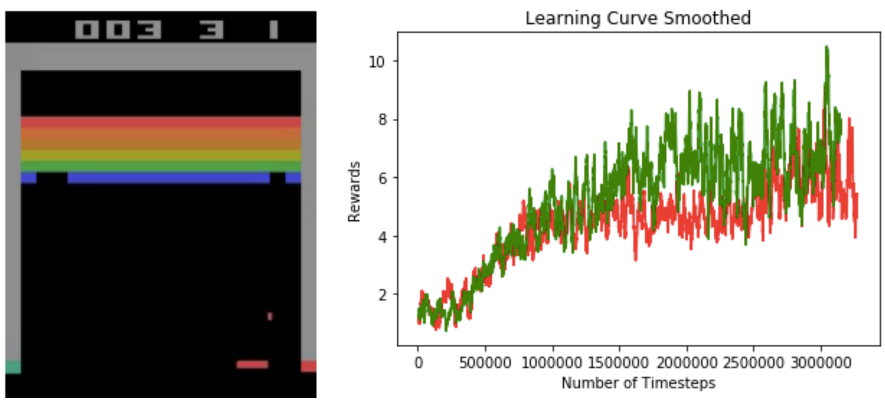

Fig. 1 Shows the environment used and the learning curve obtained after 3M steps.

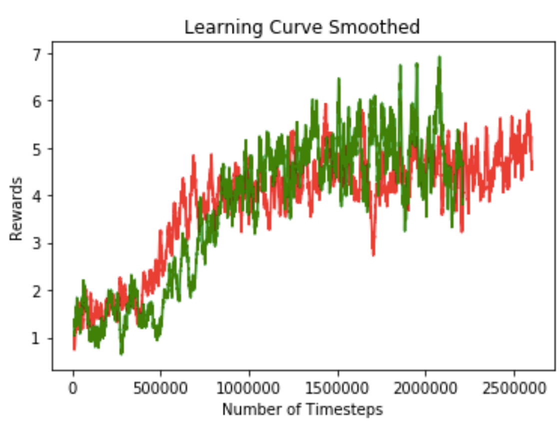

Fig. 2 Shows the learning curve obtained while encoding the experience through learning rate vs replay buffer.

In the figure 2, the green curve represents the learning curve modulated by changes in size of replay buffer and the red curve represents the learning curve modulated by changes in learning rate, both of them influenced by experiences. However, we can see that the encoding of replay buffer size substantially effects the learning in the initial phase, where as the learning rate effects it more during it's later phase.

3 Implications to AI safety

By abstracting the access to resources such as replay buffer, optimizer with function F, we can hardcode this feature and make sure that any AI agent physically feels the difference of learning some thing good vs bad, provided the reward system is biased towards the safety of humans.

4 Conclusion

Essentially the function F, is abstracting all the processes within Neuroendocrinology, which is actually responsible for the real feel in mammals [4]. This function could further be broken down into multiple functions as the AI system gets more complex.

5 References

[1] Sutton, Richard S. Reinforcement Learning: An Introduction

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller, DeepMind Technologies, Playing Atari with Deep Reinforcement Learning, 2013

[3] Victoria Puig, Jonas Rose, Robert Schmidt, Nadja Freund, Dopamine modulation of learning and memory in the prefrontal cortex: insights from studies in primates, rodents, and birds, 2014

[4] JB Becker, SM Breedlove, D Crews, MM McCarthy, Behavioral endocrinology, 2002

[5] Elizabeth Gould, Liisa Galea, Stress and the Brain, Frontiers in Neuroendocrinology, 2018