

Utilising Reinforcement learning Performance in Gaming

Reinforcement Learning (RL) is an interdisciplinary area of machine learning which can learn to make decisions by trial and error, aiming to maximize a cumulative reward signal. RL depends on the choice of parameters which will be the focus of the blog.

Achieving the task at matter requires a combination of strategic decision-making regarding the hyperparameters, agent selection and normalisation of the model. This blog with give an outlook on how Deep Q-Networks (DQN) is deployed and understanding the cartpole environment. The success of the model is determined by the angle that the stick merges in and if the stick falls, the game ends. So, the objective is to keep the stick in an upright position for 200 steps.

* Reinforcement Learning Hyperparameters

Figure 1 depicts the entries plotted to ensure refined iterations. A breakdown of the element has been used to further understand the reasons for the parameter choices.

* + **Iterations:** The number of iterations used was 10000 to allow the agent to spend more time learning from its past experiences and to improve convergence.
  + **Initial collection:** with a small initial collection, the agent can learn from real interactions rather than starting from the beginning.
  + **Steps per iteration:** steps per iteration can depict the number of steps needed to collect the data during the training process. 1 iteration was used per step to improve the models perf romance gradually.
  + **Batch size:** This parameter is the selection of train iteration samples. The appropriate selection of batch size 256 for Cartpole training due to the faster convergence and an efficient memory size.
  + **Learning rate:** the learning rate was set to le-3 which was carefully experimented, however is known as a default input, to increase the stability of learning for the model.
  + **Log and Evaluation Interval:** when determining the frequency of log training and agent performance, a log interval of 500 was used with an evaluation interval of 1000 because there is a need to obtain sufficient information and regular updates from the machine.
* DQN Agent & Normalisation usage
  + The neural network design is already integrated into neural network architectures to preprocess input data effectively. Deep learning and Q learning combined can achieve the collection of experience tuples as well as resetting the training step counter to overall achieve a better outcome.

Training outcome

As shown in Figure 2, the outcome was that the steps consistently reached 200 meaning the game was running at a successful rate where the stick would not delve into an angle where the game would end which highlights the state of the Cartpole environment and underscore the efficiency of reinforcement learning to address decision-making tasks in the best way.

Appendix

A screenshot of a computer

Description automatically generated

Figure 1

A black rectangular object with white text

Description automatically generated

Figure 2