Simulated Cart-Pole Balancing using Reinforcement Learning

A Project Report

Submitted by:

Sharoon Saxena(0902IT151045)

Yogender Singh(0902IT151060)

In partial fulfilment for the award of degree

of

BACHELOR OF ENGINEERING

IN INFORMATION TECHNOLOGY

Rustamji Institute of Technology

BSF ACADEMY TEKANPUR, GWALIOR

DECLARATION

I hereby declare that the project entitled "**Simulated Cart-Pole Balancing using Reinforcement Learning"** submitted for the B. E. (IT) degree is our original work and the project has not formed the basis for the award of any other degree, diploma, fellowship or any other similar titles.

Place: **Signature of the Student**

Date:

CERTIFICATE

This is to certify that the project titled "**Simulated Cart-Pole Balancing using Reinforcement Learning"** is the bona fide work carried out by Sharoon Saxena and Yogender Singh, students of B.E. (IT) of RUSTAMJI INSTITUTE OF TECHNOLOGY, affiliated to, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (M.P.), during the academic year 2018-19, in partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering (Information Technology) and that the project has not formed the basis for the award previously of any other degree, diploma, fellowship or any other similar title.

Signature of the Guide **Signature of HOD**

Table Of Contents

Team Details

Member 1

Name	Sharoon Saxena
Address	13/2 HIG Geetanjali Complex T.T. Nagar Bhopal
Contact	9425096938
E-mail	saxenasharoon@gmail.com

Member 2

Name	Yogender Singh
Address	House 52/10 Krishna Nagar Jamuhan Dibiyapur Auraiya, U.P.
Contact	9039430719
E-mail	syogender799@gmail.com

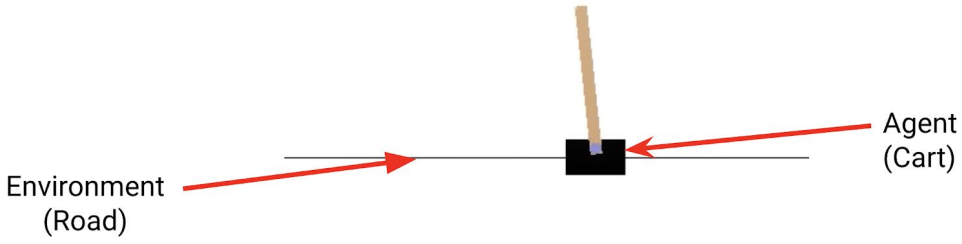
INTRODUCTION

Deep Learning has been flourishing since the the team using the Deep Learning approach won the international Image-Net competition with a significant score Advantages. It has been a significant decade in the domain of Deep Learning, where machines have starting to achieve feats which were earlier considered impossible.

Among the most Prominent requirement of the deep learning is the need of labelled data. A good neural network requires a very large amount of the Labelled data in order to learn from it, typically ranging from 100,000 examples to several million examples.

This has always proven itself as a caveat, since gathering data of this size is no mean task and requires extensive manual labelling. In some scenarios, the data may never present to begin with.

In order to fight this drawback, We are proposing the use of Reinforced Learning. The special thing about the Reinforcement Learning is that it requires no training data and the system can learn the on its own with experience.
In Reinforcement Learning, there are fundamentally 3 components.



Environment

It is finite and well defined, It is the region or the boundary within which the agent can interact and move around.

In our Example of Cart-Pole, the road is the environment, upon which the cart can move.

Agent

Agent is the entity which interacts with the environment. In our example the agent is the Cart, which moves along the road (environment) to produce some results.

Policy

Policy defines the criteria, or a set of rules; which tells agent to choose the right actions depending upon the rewards obtained by the agent. In our example, the policy is quite simple. Maximise the score by surviving as long as the the pole is balanced, falling of pole ends the game.

OBJECTIVE

Objective of this project is to implement an algorithm of Reinforcement Learning, which would allow us to work around the caveat of the deep learning; problem of obtaining a dataset.

In order to scale our success, we will be comparing the model trained with reinforcement learning against the Neural Network Approach and other approaches such as the Brute Force and weighed vectors.

We are aiming to achieve the objective using the Crt-Pole environment of the OpenAI gym, which allows us to use their pre-rendered environments, so as to avoid the overwork of managing the graphical visualisation.

Research Methodology

In order to undertake this research, we will be implementing the 4 approaches in order to benchmark the performance of the Reinforcement learning model.

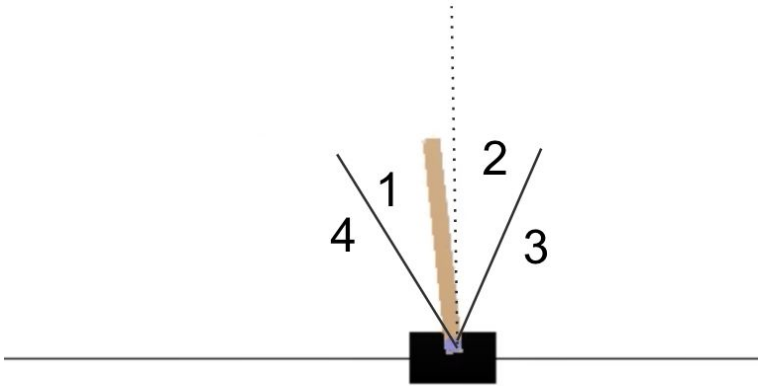
Brute Force Approach

This follows the most straightforward approach, depending upon the current angle of the pole with respect to the cart, we accelerate the card based on the hardcoded acceleration.

Weighted Vectors Approach

In case of the weighted Vectors, we divide the position of the poles in either direction in to 2 regions, 4 in total when both sides are combined. Then we choose a randomly generated vector of size 4 between 1 and -1 and the values are chosen using the uniform distribution.

This implementation of weighted vectors allows the pole to survive for longer because the pole is accelerated with respect to the region in which the pole is in in the current time frame.



In this figure the numbers 1,2,3 and 4 represents the 4 regions and the acceleration is stronger in case of region of 3 and 4.

Neural Network Approach

First step would be to obtain a dataset to train the neural network upon. We ran 50,000 episodes of brute force approach. Minimum threshold of selection = 100 time steps

- If an episode lasts longer than that, add that to the training set
- Else discard

  (we need quality data to train on)

We train a Sequential Neural Network on the the dataset we obtained

It learns the pattern of successful cart pole balancing.

Reinforced Approach

Uses a technique in which the model is rewarded if it makes correct action, penalty otherwise

Actions are made, given the observations of a state.

initially the model will not be very good at guessing the output.

slowly it will become good at predicting the output.

Exploration and exploitation is carried simultaneously to find new improved solutions and to find the good solution in explored search space.

Expected Outcomes

Brute force

This approach is expected to perform very poorly on this task, as there are many variables involved, such as angle of the pole and different degrees of acceleration for the cart, the brute force approach does not consider them and therefore the results are expected to be sub par.

Weighted Vectors

This method is expected to outperforms the brute force Approach.We are still not using any machine learning algorithm therefore the results will often saturate out after certain limit.

Neural Networks

No wonder in what a Neural network is capable of doing. It could easily learn the cart pole balancing from the dataset we created.

Using a suitable dataset, which we will be creating ourself, this methods will easily outperform the two previous methods of Brute Force and Weighted Vectors.

"Many a times it is not possible to obtain a large dataset in order to train a Neural Network, It has always been a bottle neck" Because of this reason, we will need to invest a great amount of computational resources in creating and training the dataset.

Reinforced Learning

Exploration and exploitation is carried simultaneously to find new improved solutions and to find the good solution in explored search space, therefore we can see that initially the model was not able to perform very good. Eventually it will learn from its mistakes and performs very good.

Conclusion

Brute Force Approach

Approach performs very poorly because it does not take into consideration the degree of present state, i.e. how much the pole is tilted. On 10 trail of runs the max time of survival is 118 timesteps and average survival time of about 21 time steps which is sub par as expected.

Weighed Vectors

This method outperforms the previous method. With proper number of games played this approach can last for more than 1000 time steps. On 10 episodes of this algorithm max score achieved was 762 and avg score of about 315

Can obtain better results with more calculated weight vectors

Neural Networks

No wonder in what a Neural network is capable of doing. It could easily learn the cart pole balancing from the dataset we created

Max Score was 1701 time frames and averaging 1200 time frames

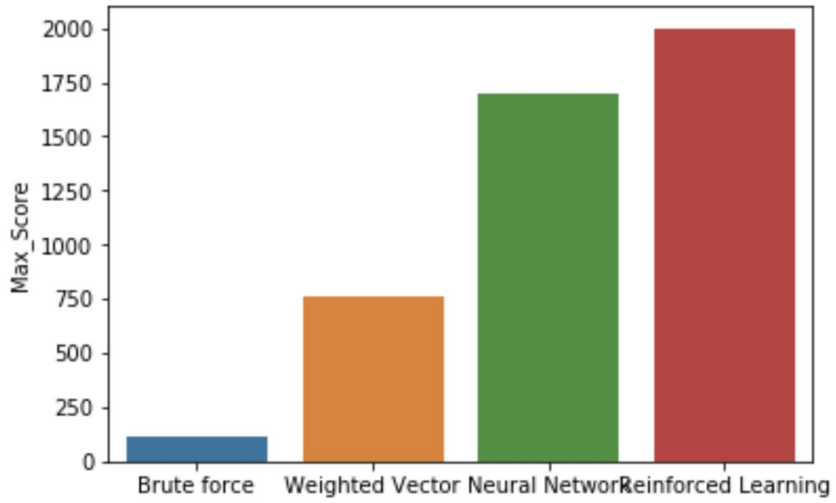
“Many a times it is not possible to obtain a large dataset in order to train a Neural Network, It has always been a bottle neck”

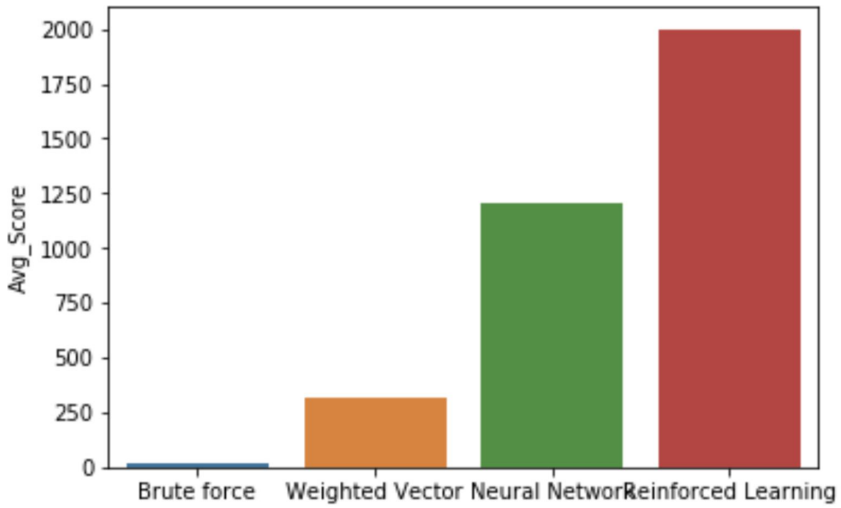
Reinforced Learning

We can see that initially the model was not able to perform very good. Eventually it learns from its mistakes and performs very good.

1199 is the upper time limit ...after this game is forcefully closed

Even higher avg score can be achieved by training longer and increasing the time limit





Looking at the graphs above, we can clearly see that the Reinforced approach clearly out performed the Neural Networks in both the Maximum Score and the Average Score.

Note that the Max Score was capped at 1999, if the threshold could be removed, Reinforced Learning approach can reach the score of infinity.

References

Barto Andrew and Sutton Richard; 2017, Reinforcement Learning: An Introduction

Aurélien Géron; 2017 , Hands on Machine Learning and Deep Learning using Scikit-Learn and Tensorflow

OpenAI; 2016, OpenAI Gym