

Playing Atari with Deep Reinforcement Learning

Research Review

By Mahad Amer Akhter

Original Paper: <https://arxiv.org/pdf/1312.5602v1.pdf>

Goals and Techniques

The goal of this paper is to design a single neural network which can play as many Atari Games as possible using Reinforcement Learning (RL) algorithms using only the raw video data as input from the games. Unlike supervised and unsupervised learning algorithms which typically require huge datasets to train an efficient model, reinforcement learning algorithms in this case would require the network to learn from rewards which are typically delayed many steps into the future.

However, there are a few challenges that first need to be tackled by RL algorithms that do not typically exist in the case of supervised and unsupervised learning. Most deep learning algorithms assume that the input data is not dependent on one another. However, this is contrary to scenarios encountered in reinforcement learning in which multiple states are usually related to one another. Secondly, deep learning algorithms also usually expect the input to have stationary data distribution. This can create issues because this means that the data distribution is constantly changing as the network learns new strategies

This paper introduces a technique to overcome these challenges. It shows that a Convolutional Neural Network (CNN) can be trained to guide actions using only the raw video data that we get from the game. It is trained using a version of the Q-learning algorithm and stochastic gradient descent is used in order to update the weights. In order to solve the problem of data being dependent and the data distribution constantly changing, a trick is used called "Experience Replay". The agent's experiences are recorded and a data set is made using these experiences. During training, these experiences are drawn at random from the pool of stored experiences in order to update the weights

Experimentation and Results

The experiments were performed on 7 Atari games. The neural network architecture, hyperparameters and the learning algorithm were kept constant among all the experiments.

In three of the seven games, this model achieved performance better than an expert level player. In the fourth game, it achieved a close to human performance. In the last three games, which are more challenging, the performance was not good compared to human performance.