Summary of Human-level control through deep reinforcement learning
By Morgan Moskalyk

Goals

The paper *Human-level control through deep reinforcement learning*, delves into the concept of reinforcement learning and a new approach of evaluating the best score of a given state, called Q-learning and how it can be applied to game agents to conquer Atari 2600 games.

With typical deep learning problems, the solutions are biologically inspired by being modelled after the way that humans learn through dopamine triggers and repetitive practice of a given task. Similarly, in a deep learning implementation you need millions of data points to train a model to perform the most correct actions. However, by using reinforcement learning, one is able to get continuous feedback on the agents' actions, in order achieve a better result; similar to how dopamine rewards the brain on successful actions.

Therefore, with the use of a combined deep learning architecture of a convolutional network for image recognition and a new Q-network architecture which was then combined with reinforcement learning, the goal of the paper hoped to achieve a general level of intelligence with a single algorithm in the realm of game playing agents.

Techniques

Deep Q-Network

The paper used an architecture called Deep Q-network with reinforcement learning by receiving only the pixels and the game score as inputs. The "Q-learning" part of the network helps establish what the 'quality' of a given position is, by not only taking into consideration of the current state, but all the next possible states and those discounted reward values. When evaluating and updating the value of a given states, this is done only periodically to reduce correlations in observations.

Reinforcement Learning

The goal of the agent is to beat games and win with a higher score by making decisions on an accumulation of the agents rewards. The paper was able to achieve higher rewards, by making an observation, performing an action, and feeding that into the sum of the possible rewards for every action an agent might take on an environment. This iterative feedback loop of state, action, and reward is the fundamental basis of reinforcement learning.

Training

The team used stochastic gradient descent to train the model and find a minimum in the loss function.

Experience Replay

This technique hopes to achieve a reduction in the correlation of various observations, by randomly pulling mini-batches from the memory, instead of the most recent state transition. The result of this will make sure that the Q-values calculated will converge when using a convolutional network (which is typically an unstable calculation due to the possibility of reaching a local minimum).

Visualization of High Dimensional Data

In order to validate the generality of the approach, the team utilized a technique called t-SNE. This approach allows for highly dimensional data to reduce its dimensionality to a space of two or three dimensions.

Results

The DQN agent, achieved more than 75% of the score, on more than half of the games of professional human game players across a set of 49 games. It achieved this with minimal training data while deriving efficient representations of an environment from complex inputs, like the pixels of a game.

Additionally, the result of the t-SNE visualization displayed that there were groupings of similar reward state values, even though they were perceptually dissimilar. This showcased the generality of the algorithm across different games, which they were able to achieve with the same architecture and hyper parameters.