Relatório Robótica Computacional Insper-Pong

Bruno Costa

Engenharia da Computação 4º Semestre

Funcionamento:

O código é baseado num arquivo da Medium: *Write an AI to win at Pong from scratch with Reinforcement Learning* e um de Andrej Karpathy: *Deep Reinforcement Learning: Pong from Pixels*, foi utilizado o OpenAi Gym e sua biblioteca de Atari. Foi utilizado os algorítimos de *Policy Gradients* com *backpropagation*, sendo que o input da rede é a imagem do jogo.

Primeiramente, foram definidos os parâmetros de aprendizagem: o tamanho do batch, o *gamma*, a taxa de decadência, taxa de aprendizado e o número de neurônios da rede, além disso foram gerados os primeiros pesos aleatoriamente (depois foi usado um sistema de *save* e *load* através do *pickle*). Começando o aprendizado, gera-se o ambiente, que pode ser renderizado, e a imagem passa por um processamento, no qual ela é cortada, sofre um *downsample*, convertida para preto e branco, o fundo é removido e a matriz resultante dos pixels é achatada (de 80 x 80 para 6400 x 1). Em seguida é calculada a probabilidade do bloco ir para a cima, através da aplicação da rede neural na matriz da imagem processada, ou *forward pass*, que é feita por uma série de multiplicação das matrizes, resultante em uma probabilidade. Após calculada, a probabilidade é rodada e o bloco realiza a ação a cada *step*. Para obter o gradiente por ação é aplicado uma *Loss Function* que assume que a ação tomada é a correta.

Após o episódio terminar, alguém atingir 21 pontos, são compiladas todas as *observations*, gradientes, recompensas e valores da *hidden layer*. Em seguida é aplicada a *Discount Function*, que reduz o efeito das ações tomadas no começo do episódio sobre os gradientes, além do cálculo dos próprios gradientes (feito a partir das quatro equações de *backpropagation)*, ou seja, a direção que os pesos devem ir para que a recompensa seja mais alta. Finalmente, após o número de episódios do *batch* os pesos são movidos na direção dos

gradientes através do *RMSProp*, método de aplicação destes gradientes. Além disso, a cada episódio o ambiente e as listas de valores sofrem um *reset*.

Para rodar o treinamento basta rodar o arquivo Pong.py.