
ALUNO(S): Luiz Guilherme Monteiro Padilha

1. INTRODUÇÃO

É possível imaginar incontáveis cenários em que algoritmos de aprendizagem de máquina possam ser implementados e experimentados, a fim de restringir tal universo, foi selecionado uma aplicação em jogos, mais especificamente ao famoso jogo “Space Invaders” de 1978, para a plataforma de vídeo games Atari, aplicado ao jogo foram utilizados dois algoritmos de aprendizagem por reforço, Deep Q-Networks e Cross-Entropy Method, a fim de otimizar a quantidade de pontos obtidos em uma partida com 3 tentativas.

2. ALGORITMOS DE APRENDIZAGEM DE MÁQUINA

Os algoritmos escolhidos utilizam o processo de decisão de Markov, onde o agente utiliza um conjunto de ações, possui um conjunto de estados, recompensas, políticas de tomadas de decisão e valor.

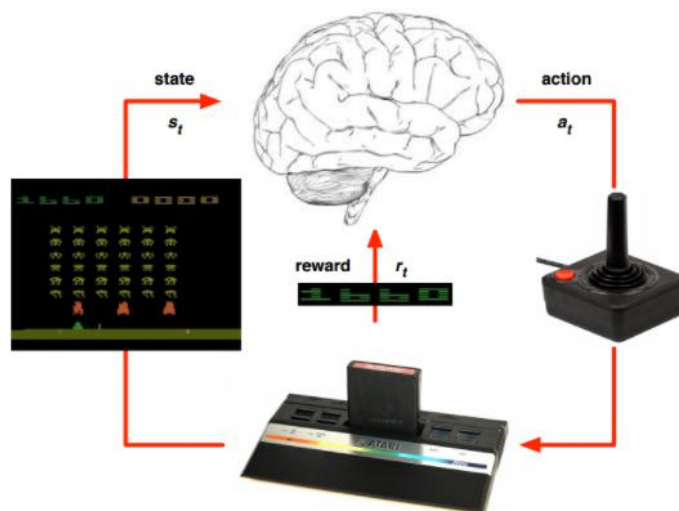


Imagem 1: Processo de decisão de Markov no jogo.

2.1 Deep Q-Networks

Utilizando a ideia de Deep Q-Learning que combina aprendizagem por reforço com deep learning, empregando uma rede neural onde o agente aprende a tomar decisões por tentativa e erro, ponderando as ações do mesmo com recompensas, no caso do jogo “Space Invaders” as recompensas estimulam as ações a fim de maximizar a quantidade de pontos acumulados em um episódio(3 vidas do jogador).

O algoritmo recebe como entrada os pixels renderizados pela tela do simulador Stella, processa essa entrada em sua rede neural a fim de obter o valor Q do peso da ação tomada, então para saber qual ação deve ser tomada seguindo alguma política dado o estado atual, a fim de maximizar sua pontuação atual.

2.2 Cross-Entropy Method

É um método de Monte Carlo para amostragem aleatória e otimização de importância, sendo um algoritmo evolutivo, onde algumas amostras são executadas e apenas as que possuam as melhores características serão carregadas para as gerações futuras, logo as sequências de ações que empregaram melhores pontuações para dado estado, deve ser a

3. METODOLOGIA

Os algoritmos foram executados primeiramente como treino, onde passaram um total de dez mil passos(ações) de jogo, em uma instância determinada do jogo. Após isso foi executado um teste de cinquenta episódios de jogo para se extrair uma média de pontos de ambos os algoritmos.

4. RESULTADOS OBTIDOS

Como resultado se obteve a seguintes tabelas:

Tabela 1: Resultados obtidos

| Episódio | Deep Q-Network | Cross-Entropy Method |
|----------|----------------|----------------------|
| | Score | |
| 1 | 460 | 80 |
| 2 | 290 | 20 |
| 3 | 210 | 185 |
| 4 | 555 | 165 |
| 5 | 210 | 45 |
| 6 | 410 | 40 |
| 7 | 460 | 190 |
| 8 | 210 | 30 |
| 9 | 410 | 180 |
| 10 | 180 | 90 |
| 11 | 180 | 35 |
| 12 | 210 | 410 |
| 13 | 210 | 65 |
| 14 | 440 | 220 |
| 15 | 180 | 45 |
| 16 | 410 | 15 |
| 17 | 410 | 110 |
| 18 | 260 | 35 |
| 19 | 180 | 115 |
| 20 | 210 | 155 |
| 21 | 410 | 120 |

| | | |
|-------|-------|-------|
| 22 | 180 | 90 |
| 23 | 210 | 80 |
| 24 | 460 | 120 |
| 25 | 180 | 45 |
| 26 | 210 | 155 |
| 27 | 410 | 115 |
| 28 | 180 | 90 |
| 29 | 185 | 70 |
| 30 | 310 | 70 |
| 31 | 180 | 80 |
| 32 | 180 | 260 |
| 33 | 310 | 205 |
| 34 | 210 | 75 |
| 35 | 180 | 110 |
| 36 | 580 | 35 |
| 37 | 540 | 120 |
| 38 | 210 | 50 |
| 39 | 410 | 205 |
| 40 | 410 | 255 |
| 41 | 340 | 35 |
| 42 | 410 | 160 |
| 43 | 155 | 35 |
| 44 | 210 | 170 |
| 45 | 410 | 85 |
| 46 | 210 | 60 |
| 47 | 180 | 95 |
| 48 | 260 | 70 |
| 49 | 310 | 115 |
| 50 | 180 | 65 |
| Media | 293,9 | 109,4 |

5. CONCLUSÃO

Pode-se observar que o desempenho em maximizar a pontuação durante cada partida foi mais eficiente no emprego do algoritmo Deep Q-Network, dado o treinamento curto empregado aos dois algoritmos, nenhum obteve sucesso em terminar a primeira fazer do jogo, porem com um treinamento mais longo seria possível que ambos os algoritmos aprendessem a lidar com os desafios do game e superar as fazes seguintes maximizando ainda mais a pontuação.