Instituto Tecnológico de Aeronáutica - ITA
Inteligência Artificial para Robótica Móvel - CT-213
Aluno:

Relatório do Laboratório 12 - Deep Q-Learning

1. Breve Explicação em Alto Nível da Implementação

Para o ajuste da recompensa, seguiu-se as modificações propostas:

```
reward = reward + (state[0]-START_POSITION_CAR)**2 + state[1]**2
#
eig = 0
#
if next_state[0] >= 0.5:
#
eig = 1
#
reward = reward + 50*eig
#
```

A implementação da rede neural por sua vez:

```
model = models.Sequential()
#
# Layer 1
#
model.add(layers.Dense(24, activation=activations.relu, input_dim=self.state_size))
#
# Layer 2
#
model.add(layers.Dense(24, activation=activations.relu))#
# Layer 3
#
model.add(layers.Dense(self.action_size, activation=activations.linear))
#
```

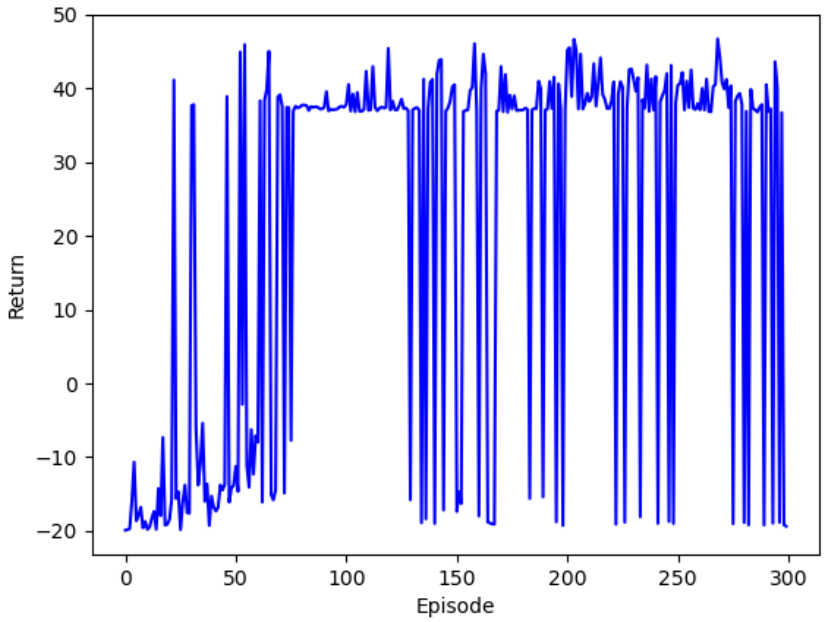
A implementação do Epsillon-greedy foi feita de forma análoga aos laboratórios anteriores.

2. Figuras Comprovando Funcionamento do Código

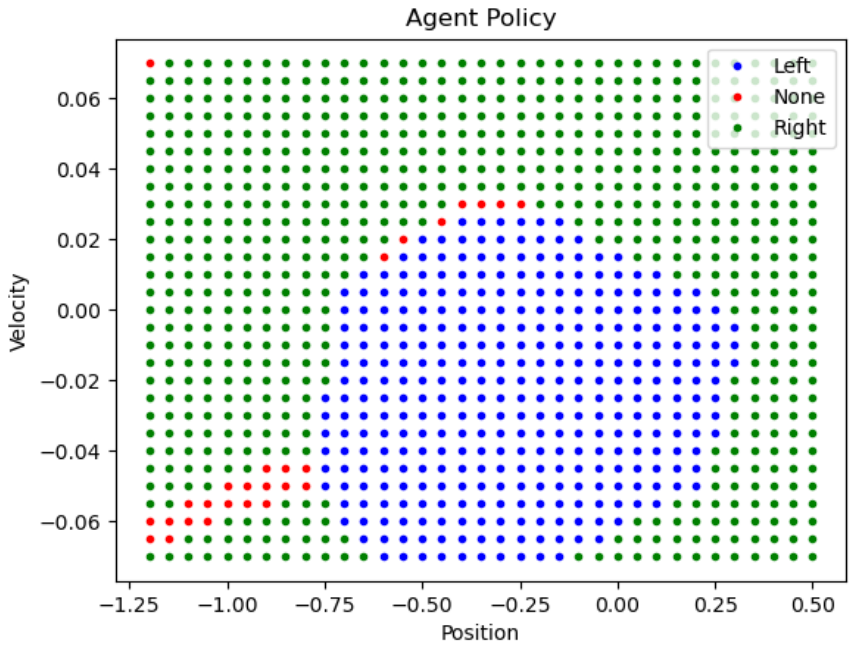
2.1. Sumário do Modelo

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
=
 dense (Dense)               (None, 24)                72

 dense_1 (Dense)             (None, 24)                600

 dense_2 (Dense)             (None, 3)                 75

=================================================================
=
Total params: 747
Trainable params: 747
Non-trainable params: 0
_________________________________________________________________
Loading weights from previous learning session.
```

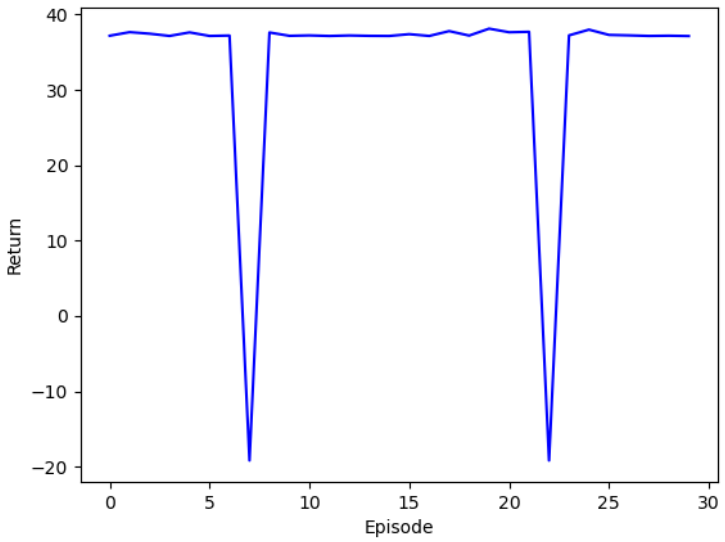
2.2. Retorno ao Longo dos Episódios de Treinamento



2.3. Política Aprendida pelo DQN



2.4. Retorno de 30 Episódios Usando a Rede Neural Treinada



3. Discussão dos Resultados

Observa-se a possível randomicidade do algoritmo pela seleção do batch de experiências pelos ruídos no treinamento. É especialmente instável no início do treino, começando a estabilizar ao serem acertadas mais vezes a tarefa proposta.

Para a política montada, observa-se o desenho de conjuntos viáveis de ação que resolvem o problema, sendo interessante por descrever limitações físicas. Observa-se por fim que no fim a política treinada e presente no arquivo <mountain_car.h5> funcionou bem no espaço de teste.