

## Relatório Ficha 1 – Reinforcement Learning

### Carlos Guerra – nº78299 METI

- 1) Esta função é a descrita no código como a `master_movement(state, action)` que por sua vez vai chamar funções do ficheiro `aux_file` dependendo do estado em que o robot se encontra, por exemplo, se o robot se encontrar no estado 10 apenas se pode mover para baixo ou para a sua direita
- 2) Esta função está no código como `reward(state)` que ao receber um estado se este for o estado 100 devolve como reward um 100 e 0 para qualquer outro estado
- 3) Para correr esta função é necessário correr a função `run_xk_random(valor)` onde o valor é o número de passos que pretendemos que o robot dê. Tendo devolvido estes valores:

```

RUN: 1 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.01039700000000878
RUN: 2 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.00872700000000151
RUN: 3 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.008097999999996608
RUN: 4 VALOR MÉDIO REWARD: 0.3 TEMPO: 0.008990999999997307
RUN: 5 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.008261999999996217
RUN: 6 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.008577000000002055
RUN: 7 VALOR MÉDIO REWARD: 0.2 TEMPO: 0.008187999999997642
RUN: 8 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.009183999999999637
RUN: 9 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.013653999999999389
RUN: 10 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.012244000000005473
RUN: 11 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.009355999999999659
RUN: 12 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.012055999999999829
RUN: 13 VALOR MÉDIO REWARD: 0.2 TEMPO: 0.008852000000000748
RUN: 14 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.008764000000002437
RUN: 15 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.00865199999999966
RUN: 16 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.009217999999997062
RUN: 17 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.00842599999999849
RUN: 18 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.009149999999999991
RUN: 19 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.00805499999999948
RUN: 20 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.008683999999998804
RUN: 21 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.00995499999999714
RUN: 22 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.00942199999999871
RUN: 23 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.009009999999999296
RUN: 24 VALOR MÉDIO REWARD: 0.2 TEMPO: 0.009498999999998148
RUN: 25 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.00967199999999857
RUN: 26 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.010079999999998979
RUN: 27 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.00971900000000181
RUN: 28 VALOR MÉDIO REWARD: 0.0 TEMPO: 0.009297999999997808
RUN: 29 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.0087810000000015
RUN: 30 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.00952200000000697
DESVIO DE TEMPO: 0.4956364999999999
MEDIA DE TEMPO: 0.0002824929999999808

```

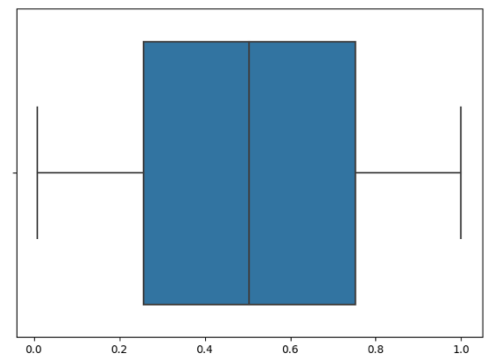


Figura 1- Gráfico dos valores de tempos de execução (1000 passos)

Figura 2- Valores da recompensa média, tempo de execução, desvio de tempo e média do tempo de execução (1000 passos)

```

RUN: 1 VALOR MÉDIO REWARD: 0.115 TEMPO: 0.20546999999999627
RUN: 2 VALOR MÉDIO REWARD: 0.095 TEMPO: 0.31632600000000032
RUN: 3 VALOR MÉDIO REWARD: 0.06 TEMPO: 0.17166500000000004
RUN: 4 VALOR MÉDIO REWARD: 0.065 TEMPO: 0.212951000000001543
RUN: 5 VALOR MÉDIO REWARD: 0.095 TEMPO: 0.165868999999999183
RUN: 6 VALOR MÉDIO REWARD: 0.09 TEMPO: 0.170523000000001663
RUN: 7 VALOR MÉDIO REWARD: 0.085 TEMPO: 0.16924999999996704
RUN: 8 VALOR MÉDIO REWARD: 0.1 TEMPO: 0.179152000000004372
RUN: 9 VALOR MÉDIO REWARD: 0.075 TEMPO: 0.260288000000001695
RUN: 10 VALOR MÉDIO REWARD: 0.085 TEMPO: 0.17380299999999593
RUN: 11 VALOR MÉDIO REWARD: 0.085 TEMPO: 0.16760600000000014
RUN: 12 VALOR MÉDIO REWARD: 0.075 TEMPO: 0.1626599999999918
RUN: 13 VALOR MÉDIO REWARD: 0.095 TEMPO: 0.17389100000000038
RUN: 14 VALOR MÉDIO REWARD: 0.09 TEMPO: 0.162451999999998317
RUN: 15 VALOR MÉDIO REWARD: 0.08 TEMPO: 0.173641000000001859
RUN: 16 VALOR MÉDIO REWARD: 0.085 TEMPO: 0.162578999999997516
RUN: 17 VALOR MÉDIO REWARD: 0.09 TEMPO: 0.172687999999991558
RUN: 18 VALOR MÉDIO REWARD: 0.075 TEMPO: 0.16985000000000028
RUN: 19 VALOR MÉDIO REWARD: 0.075 TEMPO: 0.17814299999999533
RUN: 20 VALOR MÉDIO REWARD: 0.105 TEMPO: 0.163197000000001092
RUN: 21 VALOR MÉDIO REWARD: 0.105 TEMPO: 0.194134999999993888
RUN: 22 VALOR MÉDIO REWARD: 0.065 TEMPO: 0.170390000000002513
RUN: 23 VALOR MÉDIO REWARD: 0.12 TEMPO: 0.19501600000000336
RUN: 24 VALOR MÉDIO REWARD: 0.08 TEMPO: 0.190070999999995258
RUN: 25 VALOR MÉDIO REWARD: 0.08 TEMPO: 0.17270200000000632
RUN: 26 VALOR MÉDIO REWARD: 0.07 TEMPO: 0.195951000000004528
RUN: 27 VALOR MÉDIO REWARD: 0.085 TEMPO: 0.171067000000002295
RUN: 28 VALOR MÉDIO REWARD: 0.055 TEMPO: 0.23375000000000361
RUN: 29 VALOR MÉDIO REWARD: 0.105 TEMPO: 0.21090500000000834
RUN: 30 VALOR MÉDIO REWARD: 0.06 TEMPO: 0.195901999999995228
DESVIO DE TEMPO: 0.3418369999999984
MEDIA DE TEMPO: 0.0002820946499999989
TEMPO: 5.641892999999978

```

Figura 4-Valores da recompensa média, tempo de execução, desvio de tempo e média e valor do tempo de execução (20000 passos)

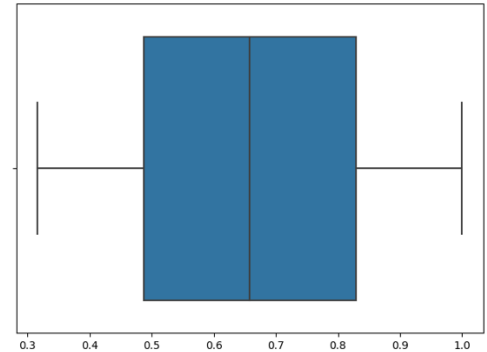


Figura 3- - Gráfico dos valores de tempos de execução (20000 passos)