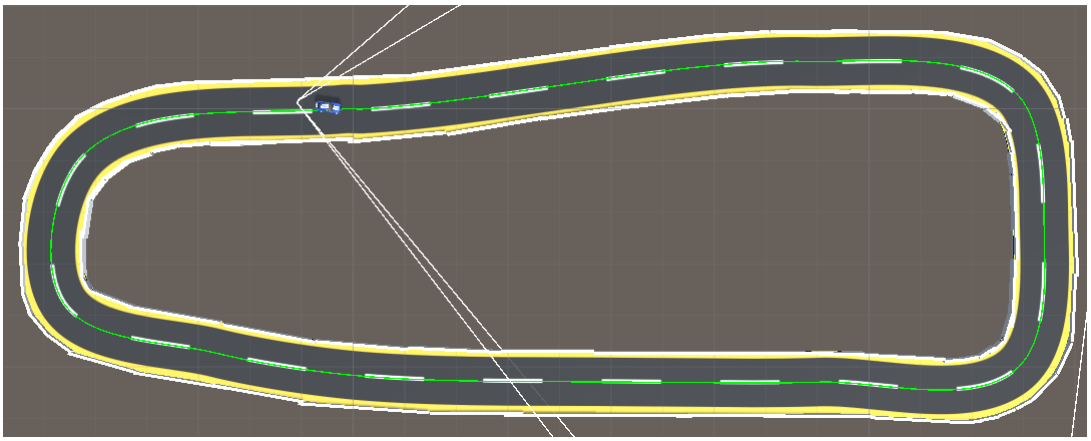
# Instituto Superior de Engenharia de Lisboa

Licenciatura em Engenharia Informática e Multimédia

Interacção em Ambientes Virtuais - 2023/2024 SV

---

# Trabalho nº 2 - Aprendizagem por Reforço com ML-Agents




Docente Arnaldo Abrantes

Realizado por :
Pedro Silva 48965


11 de maio de 2024

# Conteúdo



# Lista de Figuras

# 1 Introdução

Este segundo trabalho pretende consolidar e avaliar os conceitos adquiridos ao longo da disciplina Interação em Ambientes Virtuais sendo o objetivo usar a toolbox ML-Agents para treinar um agente de modo a que este realize uma tarefa, no nosso caso vai aprender a conduzir um carro à volta de uma pista.

Ao longo dos vídeos disponibilizados pelo docente foi demonstrado uma maneira fácil e eficaz de criar pistas, utilizando o *Asset* grátis do Sebastian Lague ”Path Creator”. Este recurso toma partido das funções B-spline e Curvas de Bézier para criar o percurso onde o nosso agente vai atuar.

O agente e o seu controlador também foram disponibilizados pelo docente mas decidimos utilizar o seguinte vídeo. Por fim foi necessário treinar o agente algo que conseguimos através da larga documentação da própria *toolbox* assim como as aulas e vídeos do docente.

# 2 Desenvolvimento

## 2.1 Criação do Ambiente de aprendizagem no Unity

Como já mencionado anteriormente, a criação do ambiente de aprendizagem beneficiou do *Asset* sugerido pelo docente, que nos permitiu criar a nossa pista. Esta é composta por 2 retas e 2 curvas sendo uma destas de grau mais complicado que a outra. Para ajudar o agente foi necessário criar paredes, para que este saiba que é para se manter no meio da pista, e checkpoints, para que ele saiba que está a progredir.

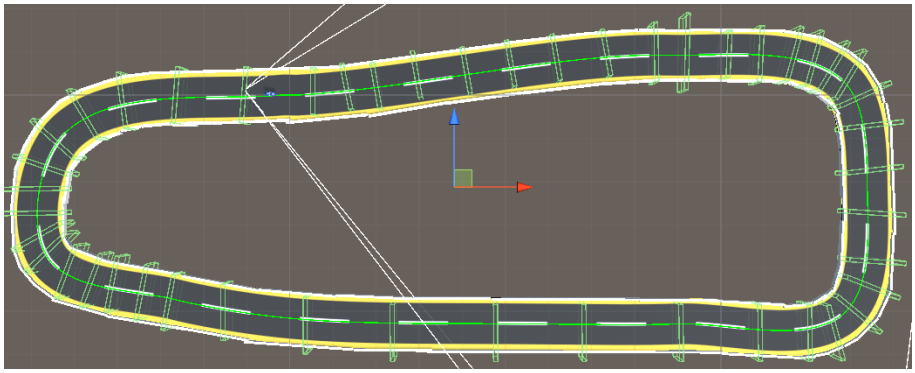


**Figura 1:** Ambiente de Aprendizagem

No total são 46 Checkpoints estando estes distribuídos de forma propositada com um maior número na reta inicial, para que este aprenda mais rapidamente, também um grande número nas curvas pois estas obrigam uma grande destreza do agente e ele precisa de ser recompensado de acordo, na segunda reta já existe um maior espaçamento pois ele já sabe o que tem que fazer.

## 2.2 Criação do Agente

Ambos o modelo e o controlador do carro foram retirados do vídeo que está nesta hiperligação.

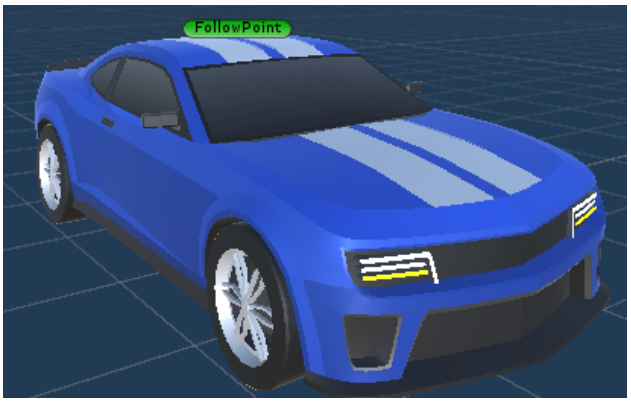



**Figura 2:** Agente

### 2.2.1 Observações

Como observações vamos ter a direção onde está o próximo Checkpoint, assim como o agente vai ter um *Ray Perception Sensor 3d* que o vai permitir identificar o que é uma parede e um Checkpoint.

### 2.2.2 Ações

Existem 2 ações continuas, o agente pode andar para a frente ou para trás com uma força de 1000 e pode rodar o volante entre 30º. Também foi testado adicionar uma ação discreta que permitia o agente travar mas esta implicava uma aprendizagem mais longa e como não tínhamos muito tempo optámos apenas pelas duas contínuas.

### 2.2.3 Recompensa extrínsecas

Para as recompensas o agente recebe +1 cada vez que passa por um Checkpoint correto e perde -1 cada fez que passa por um errado. Como castigo tem também na colisão com uma parede (-5) e por cada frame que fica a tocar numa parede (-0.5). Foi tomada a decisão de não acabar logo o episódio quando este tocava numa parede porque isso tornava os episódios demasiado pequenos o que dificultava a aprendizagem.

À medida que o agente foi aprendendo foi necessário acrescentar novas recompensas. Como o agente estava mais confiante este começou a acelerar cada vez mais o que fazia com que ele virasse o carro cada vez que chegasse a uma curva. Isto foi corrigido ao castigar por -10 e acabar o episódio.

Outra consequência desta aprendizagem foi na mesma situação mas desta vez não chegava a virar o carro apenas batia na parede. Muitas das vezes ele ficava virado para o lado errado mas continuava a ir, então decidimos que assim que ele passasse 7 Checkpoints errados castigávamos com -10 e acabávamos o episódio. Se o agente se enganar no Checkpoint mas perceber o seu erro é compensado com +3.

## 2.3 Validação do setup de treino usando o modo Heurístico

O modo Heurístico permitiu-nos afinar quer os detalhes do agente quer os detalhes da pista. Por exemplo tivemos que reduzir para metade o tamanho do agente e aumentar o tamanho da pista para que este tenha uma maior chance de sucesso. Depois destas afinações o agente estava pronto para ser treinado.

## 2.4 Treino do Agente

### 2.4.1 Algoritmo

O algoritmo e os seus parâmetros foram os seguintes:

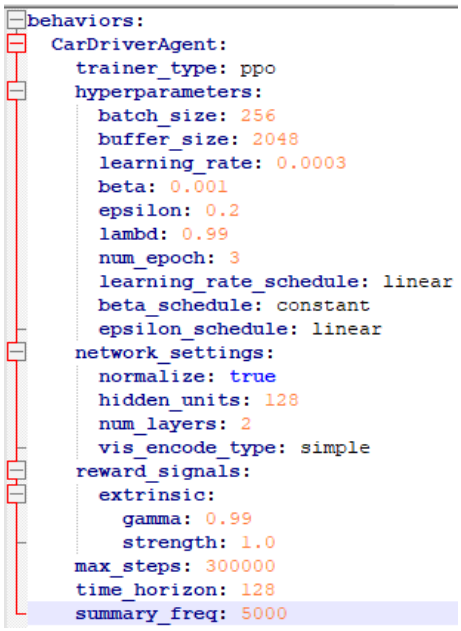



**Figura 3:** Algoritmo e Parâmetros

### 2.4.2 Parâmetros de Treino

Os parâmetros de treino foram os seguintes:

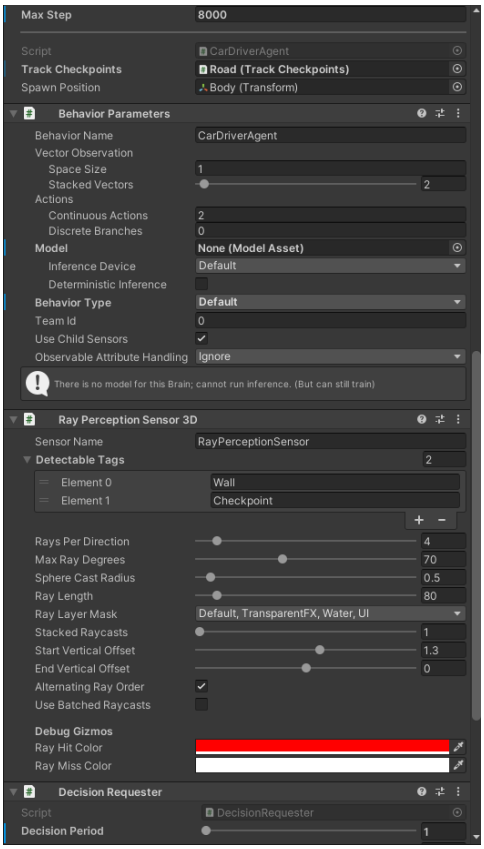



**Figura 4:** Algoritmo e Parâmetros

Todos os parâmetros mantiveram-se constantes ao longo do treino tirando o *Max Step*. Este começou com um número menor mas foi aumentando à medida que o agente aprendia, com o objetivo de o permitir dar cada vez mais voltas seguidas, sendo estes 8000 suficientes para o agente dar 3 voltas tranquilamente.

### 2.4.3 Evolução do Agente

O agente evoluiu da forma esperada, começando por aprender a andar em frente e a manter-se afastado das paredes, ou seja, no meio da pista. O próximo problema foram as curvas onde este teve que aprender a desacelerar antes de chegar a uma curva e a contorná-la. Por fim, o grande problema deste agente foi a segunda curva pois como esta é mais apertada que a primeira criou uma nova dificuldade ao condutor porque não bastava desacelerar como na última tinha que ser muito antes ou de forma a maximizar o valor das recompensas continuar a acelerar, bater na parede mas virar para o caminho correto e continuar a sua corrida. Esta última opção poderia ser evitada com a introdução do travão porque o agente só percebe que está numa curva mesmo em cima dela e com este utensílio deixa de precisar duma reação tão prévia e pode reagir mais no momento. Infelizmente não houve tempo de introduzir esta mecânica.

Os resultados deste treino foram que o agente consegue completar 1 volta 75% das tentativas, 2 voltas 50% e 3 voltas 25%. Estes resultados devem-se ao extremo grau de dificuldade da segunda curva algo que pode ser remediado com a alteração ou se tivéssemos mais tempo mais treino do agente.

## 2.5 Avaliação do Agente

O agente comporta-se de acordo com o seu objetivo tendo também algo inesperado que foi depois de bater voltar para o caminho correto. Ao fazermos inferência, usando a rede neuronal que resultou do treino, conseguimos observar o agente a realizar todos os passos descritos anteriormente com sucesso. Gostava de ter tido mais tempo para criar um ambiente só de teste para observar o desempenho do agente mas isto não foi possível.

# 3 Conclusões

O objetivo deste segundo trabalho era o de consolidar e avaliar os conceitos adquiridos na disciplina Interação em Ambientes Virtuais usando-os para criar e treinar o nosso agente. Conseguimos utilizar todo o conhecimento obtido ao longo desta segunda metade do semestre como B-spline e Curvas de Bézier, para a construção do ambiente de aprendizagem, e todas as mecânicas da *toolbox ML-Agents*, para a criação do agente.

Como foi dito anteriormente, gostava de ter tido mais tempo para introduzir a mecânica do travão assim como um ambiente apenas de teste pois acho que ambas funcionalidades iriam enriquecer este trabalho. Apesar disto, creio que o projeto desenvolvido está de acordo com o pretendido pelo docente, demonstrando o domínio que tenho sobre toda a matéria lecionada ao longo do semestre.