Título do relatório projeto final da disciplina de *Reinforcement Learning*

Giancalo Vanoni Ruggiero[a], **Luciano Felix Dias**[a], **Tales Ivalque**[b]

[a]*Engenharia da Computação, INSPER*
[b]*Engenharia Mecatrônica, INSPER*

Prof. Fabrício Barth

Abstract—Neste projeto foi desenvolvido e validado um agente capaz de atuar no ambient XXX. O método de treinamento empregado foi ... Os resultados obtidos foram ...

keywords—*Reinforcement Learning, DQN, Taxi Driver, ...*

1. Introdução

Na introdução é importante apresentar o contexto e objetivo do trabalho. **Por que usar aprendizagem por reforço para resolver este problema?** Qual é otimização desejada?

2. Ambiente

Nesta seção é importante responder as seguintes perguntas:

- **Como os estados são representados?** Não basta dizer que é um Box ou uma matriz RGB. É necessário explicar o motivo desta representação. Considerando o problema que pretende-se resolver, porque os autores do ambiente decidiram utilizar esta representação?
- **Qual é o espaço de ações do agente?** As ações são discretas ou contínuas? Se discretas, quantas são? O ambiente é determinístico ou não?
- **Como é definida a função de reward?**

Para responder algumas das perguntas acima, talvez seja necessário utilizar equações. Desta forma, segue alguns exemplo de como definirir equações em LaTeX. Neste caso, são apresentados dois exemplos, a equação 1 e a equação 2.

$$\frac{\hbar^2}{2m}\nabla^2\Psi + V(\mathbf{r})\Psi = -i\hbar\frac{\partial\Psi}{\partial t} \tag{1}$$

$$f(x) = x^2 + 2x + 10 \tag{2}$$

3. Método

Nesta seção é importante descrever os **algoritmos testados**. Uma breve descrição, não é necessário descrever detalhes do funcionamento do algoritmo. A equipe pode assumir que o leitor deste texto é alguém familiarizado com os algoritmos de reinforcement learning. Não esqueçam de colocar as **devidas citações**. Por exemplo:

> **Exemplo**
>
> Este artigo irá fazer uso dos algoritmos DQN [1] e PPO [2] para treinar um agente para o ambiente XXX...

Também é importante destacar a implementação utilizada. Se foi uma implementação feita do zero pela equipe ou se foi utilizada uma biblioteca. Se a equipe optou por utilizar uma biblioteca é importante citar a biblioteca. Em ambos os casos, se a equipe fez a sua própria implementação ou se utilizou uma biblioteca, é importante **citar o repositório** onde os scripts se encontram.

Nesta seção também é importante descrever quais são os principais **indicadores** que a equipe está avaliando. Isto está diretamente relacionado com a função que pretende-se otimizar - que foi descrita na introdução deste relatório.

Eventualmente, a equipe deseja adicionar algum trecho de código nesta parte do relatório. Isto pode ser feito da seguinte maneira:

```python
env = gym.make(
        'MountainCarContinuous-v0',
        render_mode='human')

(obs,_) = env.reset()

for i in range(1000):
    a, _s = model.predict(obs, deterministic=True)
    obs, reward, done, truncated, info = env.step(a)
    env.render()
    if done:
      obs = env.reset()
```

Code 1. Exemplo de código em Python

4. Resultados

O objetido desta pararte do relatório é descrever os resultados obtidos. Durante a disciplina nós vimos as principais técnicas para mostrar o aprendizado do agente e o desempenho do mesmo para realizar a tarefa. Esta seção precisa mostrar dados quantitativos que descrevem isto: a curva de aprendizagem e as validações feitas depois do treinamento.

Para apresentar os resultados obtidos com certeza a equipe terá que fazer uso de figuras, como a apresenta em 1.

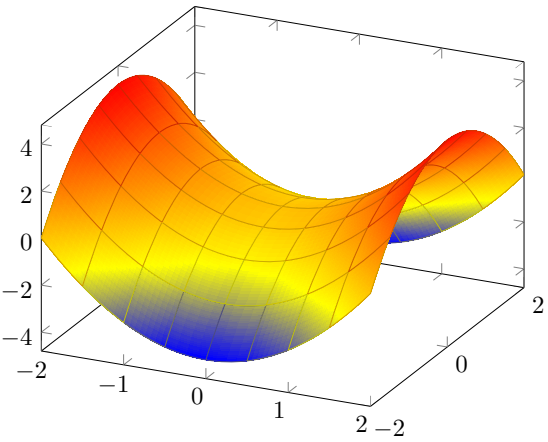

Figure 1. Exemplo de figura obtido a partir de *PGFPlots - A LaTeX package to create plots.* [Online]. Available: https://pgfplots.sourceforge.net/).

Além de tabelas, como a apresentada em 1.

Table 1. Exemplo de tabela

Column 1	Column 2
Data 1	Data 2
Data 3	Data 4

5. Considerações finais

O objetivo deste trabalho foi ...

References

[1] V. Mnih, K. Kavukcuoglu, D. Silver, *et al.*, "Playing atari with deep reinforcement learning", *CoRR*, vol. abs/1312.5602, 2013. arXiv: 1312.5602. [Online]. Available: http://arxiv.org/abs/1312.5602.

[2] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms", *CoRR*, vol. abs/1707.06347, 2017. arXiv: 1707.06347. [Online]. Available: http://arxiv.org/abs/1707.06347.