

Ciência da Computação / Sistemas de Informação

## **Utilização de Técnicas de Aprendizado por Reforço em Cadeias de Suprimentos**

Marcos Carvalho Ferreira - 8º módulo de Engenharia de Controle e Automação, UFLA, bolsista PIBIC/CNPq

Julio César Alves - Professor do Departamento de Computação Aplicada - ICET/UFLA - juliocesar.alves@ufla.br - Orientador - Orientador(a)

### **Resumo**

Este Plano de Trabalho faz parte do início de um projeto que visa aplicar técnicas de Aprendizado por Reforço (AR) e Otimização Estocástica em problemas de decisão sequencial com incertezas de Cadeias de Suprimentos. Como conceitos de AR não fazem parte das ementas das disciplinas do curso da UFLA, e este plano foi o primeiro plano do projeto, foi necessário, primeiramente, estudar toda a base teórica da área de Aprendizado por Reforço para aplicação de algumas das técnicas clássicas em algum problema de Cadeias de Suprimentos. O problema escolhido foi o Beer Game, que é muito utilizado como ferramenta didática para entendimento de conceitos de cadeias de suprimentos, como o efeito chicote. E o objetivo era utilizar um algoritmo clássico de AR, no caso o Q-Learning, para resolver o problema. Mas, como o Beer Game original tem espaços de estados e ações muito grandes para a aplicação do algoritmo citado, foi necessário simplificá-lo (o algoritmo decide a mesma quantidade para todos os nós da cadeia, e ele enxerga, como estado, apenas as informações do primeiro elo da cadeia). O problema e o algoritmo foram implementados em Python e foram feitos testes de validação da implementação. Em seguida, foram realizados diversos experimentos variando-se os parâmetros do algoritmo e o tamanho do treinamento. Os resultados dos experimentos, com 10 milhões de episódios, mostraram que o Q-Learning conseguiu um resultado melhor que um agente aleatório, indicando que houve aprendizado. Pretende-se, como continuidade deste trabalho, avaliar possíveis melhorias na forma de aplicação do algoritmo, comparar os resultados com baselines melhores, e estudar e aplicar técnicas de Aprendizado por Reforço Profundo que são mais apropriadas para a resolução de problemas com grandes espaços de estados e ações. Os principais resultados alcançados pelo trabalho foram: o estudo, implementação e documentação dos conceitos teóricos e técnicas clássicas de AR, a formação do aluno envolvido no presente trabalho e a efetivação do grupo de pesquisa em Otimização sob Incerteza. A documentação dos conceitos estudados, os códigos implementados e a análise de resultados está disponível na forma um notebook Python, na plataforma Google Colab, com o objetivo de facilitar a continuidade dos trabalhos e divulgar os resultados alcançados. Agradecimentos: UFLA, CNPq.

Palavras-Chave: aprendizado por reforço, cadeias de suprimentos, Q-Learning.

Instituição de Fomento: Conselho Nacional de Desenvolvimento Científico e Tecnológico

Link do pitch: <https://youtu.be/MtYTwe4jcVI?si=XjRYVZc8m6naK6P9>