

Aplicação de redes neurais artificiais em aprendizado por reforço

Módulo 10: Ciência da Computação

- RESUMO EXECUTIVO

Neste módulo, equipes de alunos desenvolvem agentes inteligentes que aprendem a tomar decisões sequenciais em ambientes incertos usando aprendizado por reforço (Reinforcement Learning). A solução é construída em ambiente de alto desempenho com computação paralela e aborda problemas reais de otimização de ações ao longo do tempo.

Durante 10 semanas, uma turma dedicada de alunos e seus professores irá entender o seu desafio e propor soluções utilizando metodologias ágeis. Ao final do cronograma, sua empresa receberá até 5 protótipos funcionais, permitindo que você acelere e valide ideias.

- O DESAFIO

A empresa traz um problema onde um sistema precisa aprender a tomar decisões repetidas em um ambiente dinâmico, com o objetivo de maximizar um resultado ao longo do tempo. O desafio pode envolver controle de processos, otimização de estratégias ou simulação de comportamentos adaptativos.

- OBJETIVO DO MÓDULO

1. Modelar problemas de tomada de decisão como Processos de Decisão de Markov (MDPs).
2. Implementar algoritmos de aprendizado por reforço como Q-Learning, SARSA e métodos de Monte Carlo.
3. Aplicar computação paralela para acelerar o treinamento de agentes em ambientes complexos.
4. Analisar estratégias de negócio com IA, fusões e aquisições e estruturas corporativas em contexto de transformação digital.
5. Desenvolver técnicas de influência ética e argumentação para líderes em contextos de mudança tecnológica.

- PROBLEMAS QUE PODEMOS RESOLVER

- Otimização de estratégias de precificação dinâmica em mercados com demanda variável e concorrência.
- Controle autônomo de processos industriais com múltiplas variáveis interdependentes.
- Agendamento e alocação de recursos em sistemas com incerteza e variações em tempo real.
- Personalização de recomendações ou ações sequenciais baseadas no comportamento acumulado do usuário.
- Simulação de estratégias de negociação ou leilão em mercados financeiros ou de commodities.

- POSSÍVEIS CASOS DE APLICAÇÃO

1. Casos de Sucesso Reais (O que já entregamos)

- **Serviços Financeiros e Bancos:** Construímos com o **Bank of America (BOFA)** um simulador de estratégias de execução de ordens financeiras utilizando agentes de Aprendizado por Reforço (RL), visando a minimização do impacto de mercado e a otimização de negociações.
- **Agronegócio e Sustentabilidade:** Desenvolvemos com a **A de Agro** um agente de RL focado na otimização de estratégias de aplicação de insumos agrícolas ao longo de toda a safra, tomando decisões adaptativas com base em variáveis climáticas e nas condições do solo.
- **Indústria e Calibração:** Estruturamos para a **Minipa** um sistema de controle adaptativo focado na otimização de processos de calibração e testes de instrumentos, utilizando aprendizado contínuo para ajustar as calibrações com precisão.

2. Outras Aplicações Possíveis (Onde mais podemos gerar valor)

Treinar agentes de IA para tomar decisões sequenciais em ambientes incertos permite automatizar estratégias complexas de negócios que antes dependiam exclusivamente da tentativa e erro humana. Nossos alunos podem construir modelos de Aprendizado por Reforço para atuar de forma preditiva e adaptativa em cenários como:

- **Varejo e E-commerce:** Sistemas de otimização de precificação dinâmica, ajustando preços em tempo real frente a flutuações de demanda, níveis de estoque e ações da concorrência para maximizar o lucro e o escoamento.
- **Logística e Cadeia de Suprimentos (Supply Chain):** Agentes autônomos para controle inteligente de estoque e agendamento de frotas, adaptando-se a imprevistos, gargalos de distribuição e incertezas de demanda de produtos.
- **Mobilidade e Veículos Autônomos:** Sistemas adaptativos de planejamento inteligente de rotas para navegação autônoma em ambientes logísticos ou urbanos, lidando de forma inteligente com obstáculos em movimento.

- POSSÍVEIS ENTREGÁVEIS

- Agente treinado com documentação do ambiente, função de recompensa e política aprendida
- Relatório de avaliação com curvas de aprendizado e análise de convergência
- Pipeline de treinamento reproduzível com configurações documentadas
- Apresentação executiva com análise de viabilidade de aplicação no negócio e métricas de impacto

- **PREMISSAS E REQUISITOS**

1. Da Empresa Parceira:

- Definir claramente o ambiente de decisão: estados, ações possíveis e critério de recompensa.
- Fornecer dados históricos ou um simulador do ambiente de negócio para treinamento do agente.
- Disponibilizar especialista de domínio para validar a função de recompensa e os resultados ao longo do módulo.

2. Do Projeto:

- O ambiente de treinamento deve ser simulável computacionalmente dentro das 10 semanas do módulo.
- A solução será um agente funcional em simulação, não necessariamente integrado a sistemas de produção em tempo real.
- Problemas com ambientes de altíssima complexidade devem ser delimitados em escopo antes do início do módulo.

- **ESCOPO E RESTRIÇÕES**

1. Escopo Incluído:

- Modelagem do problema como MDP e definição do ambiente de simulação.
- Implementação e treinamento de algoritmos de RL selecionados.
- Avaliação de desempenho com análise de trade-offs de exploração e exploração.
- Documentação técnica e relatório de resultados para o parceiro.

2. Fora do Escopo:

- Implantação do agente em sistemas de produção ou controle físico real.
- Desenvolvimento de hardware de controle ou integração com sistemas SCADA.
- Garantia de desempenho ótimo global em ambientes não-estacionários de alta dimensão.

- **POSSÍVEIS TECNOLOGIAS / ARQUITETURA / FERRAMENTAS DE NEGÓCIOS**

- Python
- PyTorch / TensorFlow
- Stable-Baselines3 / RLlib (frameworks de RL)
- OpenAI Gym / Gymnasium (ambientes de simulação)
- NumPy / SciPy (estatística e otimização)
- Ray (computação paralela e distribuída)
- Docker / Jupyter Notebooks