UNIFECAF

AI & CHATBOT

**APRENDIZADO POR REFORÇO**

Edson Natan da Silva Celço RA 32418

Arthur

Lucas

Octávio

Taboão da Serra, São Paulo

2023

Edson Natan da Silva Celço

Arthur

Lucas

Octávio

**APRENDIZADO POR REFORÇO**

Trabalho apresentado no curso de Análise e Desenvolvimento de sistemas da Centro universitário faculdade capital federal (UNIFECAF).

Orientador Rafael Venâncio de Oliveira.

Taboão da Serra, São Paulo

2023

# Resumo

O Aprendizado por reforço é um tipo de programação de software que permite que um algoritmo aprenda por conta própria através da experiência. O algoritmo é colocado em um ambiente e é incentivado a tomar ações que levem a uma recompensa, enquanto evita ações que levem a uma punição. Essa abordagem é semelhante a como os seres humanos aprendem com a experiência e com as consequências de suas ações. O objetivo é que o algoritmo aprenda a tomar as melhores decisões para alcançar um objetivo específico, sem a necessidade de intervenção humana. Em outras palavras, o algoritmo aprende por conta própria o que funciona e o que não funciona, aprimorando suas habilidades ao longo do tempo. O aprendizado por reforço tem uma ampla gama de aplicações em áreas como robótica, jogos, publicidade online e muitas outras.

O termo Aprendizado por reforço (Reinforcement Learning) surgiu no ano de 1951, com Marvin Minsky, estudante de matemática pela Harvard. O mesmo criou um sistema que imitava um rato em um labirinto.

Mas, só em março de 2016, que a maioria dos pesquisadores de Inteligência Artificial entenderam como utilizar o conceito do Reinforcement Learning e como ele pode refinar a performance do algoritmo referente a tomada de decisões.

Palavra-chave: Aprendizado por reforço, programação de software, algoritmo, ambiente, decisões, aprimoramento de habilidades, Marvin Minsky, performance, tomada de decisões.

# Summary

Reinforcement Learning is a type of software programming that allows an algorithm to learn on its own through experience. The algorithm is placed in an environment and is encouraged to take actions that lead to a reward, while avoiding actions that lead to a punishment. This approach is similar to how human beings learn from experience and from the consequences of their actions. The goal is for the algorithm to learn to make the best decisions to achieve a specific goal, without the need for human intervention. In other words, the algorithm learns on its own what works and what doesn't, improving its abilities over time. Reinforcement learning has a wide range of applications in areas such as robotics, gaming, online advertising, and many others.

The term Reinforcement Learning was coined in 1951 by Marvin Minsky, a math student at Harvard. He created a system that mimicked a mouse in a maze.

But it was only in March 2016 that most Artificial Intelligence researchers understood how to use the concept of Reinforcement Learning and how it can refine the performance of the algorithm regarding decision making.

Keywords: Reinforcement learning, software programming, algorithm, environment, decisions, skill enhancement, Marvin Minsky, performance, decision making.

Sumário

[Resumo 3](#_Toc134522134)

[Summary 4](#_Toc134522135)

[Introdução 6](#_Toc134522136)

[Desenvolvimento 7](#_Toc134522137)

[Conclusão 9](#_Toc134522138)

[Referências 10](#_Toc134522139)

# Introdução

Reinforcement Learning, ou Aprendizado por Reforço, é uma abordagem de aprendizado de máquina que permite que um algoritmo aprenda por conta própria a tomar decisões a fim de alcançar um objetivo específico. É uma forma de programação de software em que o algoritmo é colocado em um ambiente e incentivado a tomar ações que levem a uma recompensa, enquanto evita ações que levem a uma punição.

A ideia por trás do Reinforcement Learning é simular a forma como os seres humanos aprendem através de experiências e consequências. O objetivo é que o algoritmo aprenda a tomar as melhores decisões para alcançar o objetivo sem a necessidade de intervenção humana. Ao longo do tempo, o algoritmo aprende por conta própria o que funciona e o que não funciona, aprimorando suas habilidades.

O Aprendizado por Reforço tem aplicações em uma ampla variedade de áreas, incluindo robótica, jogos, publicidade online e muitas outras. Na robótica, por exemplo, um robô pode aprender a navegar em um ambiente desconhecido através do Aprendizado por Reforço. Nos jogos, um algoritmo pode aprender a jogar jogos complexos como o xadrez ou o Go.

Embora o conceito de Aprendizado por Reforço tenha sido introduzido em 1951 por Marvin Minsky, estudante de matemática pela Harvard, a maioria dos pesquisadores de Inteligência Artificial começou a entender como utilizar essa abordagem apenas em março de 2016, e desde então tem sido objeto de muitas pesquisas e avanços.

# Desenvolvimento

O processo de aprendizado por reforço envolve um agente (o algoritmo) tomando ações em um ambiente para maximizar uma recompensa. O agente aprende com a experiência, ajustando suas ações para maximizar a recompensa ao longo do tempo.

Para uma melhor explicação, “o algoritmo, experimenta um processo de aquisição de conhecimento, tomando suas próprias decisões em um ambiente. Assim, algo naquele ambiente o informa que a ação realizada foi correta, compensando o algoritmo, caso a ação seja errada, o punindo” (Araújo, 2021).

Existem vários algoritmos diferentes de aprendizado por reforço, cada um com seus próprios pontos fortes e fracos. Alguns dos mais comuns incluem Q-learning, SARSA e Policy Gradient methods.

Ao longo das décadas seguintes, vários pesquisadores contribuíram para o desenvolvimento do Reinforcement Learning. Em 1988, Andrew Barto e Richard Sutton publicaram um livro chamado "Reinforcement Learning: An Introduction", que se tornou uma referência na área.

A partir da década de 1990, o Reinforcement Learning começou a ser aplicado em problemas mais complexos, como jogos de tabuleiro e robótica. Em 1997, a IBM criou o programa Deep Blue, que derrotou o campeão mundial de xadrez, Garry Kasparov, em uma partida.

“Os três tipos de aprendizado de máquina”, “O futuro é uma variável aleatória: como não se sabe a priori o que irá acontecer, é desejável uma abordagem que leve em consideração essa incerteza, e consiga incorporar as eventuais mudanças no ambiente do processo de tomada da melhor decisão. Essa ideia de fato deriva do conceito de “aprendizagem por reforço” da psicologia, no qual uma recompensa ou punição é dada a um agente, dependendo da decisão tomada; com o tempo e a repetição dos experimentos, espera-se que o agente consiga associar as ações que geram maior recompensa para cada situação que o ambiente apresenta, e passe a evitar as ações que geram punição ou recompensa menor.”, (Honda, 2017)

Nos anos seguintes, a aplicação do Reinforcement Learning em jogos de vídeo, como o jogo de corrida TORCS, e em jogos de tabuleiro, como o jogo de Go, se tornou cada vez mais comum.

O Reinforcement Learning tem sido visto como uma das áreas mais promissoras da Inteligência Artificial, com o potencial de revolucionar a forma como os sistemas inteligentes são desenvolvidos. À medida que a tecnologia continua a evoluir, é provável que o Reinforcement Learning continue a ser uma área importante de pesquisa e desenvolvimento.

De acordo com Cardeal, Enzo do Turing Talks, em sua postagem “Aprendizado por Reforço #1- Introdução”, o mesmo diz que no aprendizado por reforço existem alguns conceitos importantes como o Agente, Ambiente, Estado, Política e Recompensa.

Agente: O sistema em que o aprendizado irá rodar, podendo ser um software ou mesmo o hardware, é ele que irá tomar as decisões no ambiente, assim o mesmo terá que tomar decisões para poder receber as punições/recompensas.

Ambiente: É a simulação do problema, ele pode ser um software ou um ambiente real, neste local onde o agente irá atuar.

Estado: É como o ambiente, o sistema e o agente se encontram naquele instante. Sempre que o agente realiza alguma ação o ambiente entrega sua recompensa/punição e muda de estado.

Política: Esta é a estratégia que o agente aplica para poder se mover no ambiente com base no seu atual estado, sempre que o agente atinge um determinado comportamento ele se atualiza e para poder chegar ao novo comportamento ideal.

Recompensa: É o que o agente recebe quando chega a um determinado local do ambiente, podendo ser tanto positiva (recompensa) ou negativa (punição)

# Conclusão

No aprendizado por reforço, não há avaliação direta se as ações executadas pelo agente foram boas ou ruins. Após realizar uma ação, o agente recebe a observação do ambiente e o sinal de recompensa. Esta recompensa é uma indicação do quão útil é o agente ao executar uma ação escolhida no estado em que se encontra no momento e serve como avaliação de eficiência do agente em alcançar seus objetivos. O agente visa realizar ações que potencializem as recompensas recebidas. Essa recompensa recebe o nome de recompensa instantânea, pois é recebida do estado atual onde o agente se encontra. Por outro lado, existe a recompensa cumulativa, que leva em consideração todo o histórico de recompensas do início ao fim. O objetivo deste agente é aumentar o prêmio cumulativo recebido. Ou seja, a cada passo o trabalho do agente é realizar uma ação que maximize a recompensa que você espera receber. Um agente pode encontrar situações em que suas ações tenham diferentes benefícios. Por exemplo: um helicóptero autônomo pode se encontrar em uma situação em que tem que decidir entre chegar mais rápido ao destino ou com mais combustível. Mesmo que ambos os resultados não sejam alterados, o agente é capaz de selecionar apenas uma das opções, e para isso, uma recompensa pode ser um valor escalado, positivo ou negativo, de modo que o retorno esperado para as duas ações possa ser comparável e, portanto, essas próprias ações também possam ser comparadas.

O Aprendizado por Reforço tem aplicações em uma ampla variedade de áreas e tem o potencial de revolucionar muitas indústrias. Embora o conceito de Aprendizado por Reforço tenha sido introduzido em 1951, foi apenas em 2016 que a maioria dos pesquisadores de Inteligência Artificial começou a entender como utilizá-lo adequadamente. Desde então, o Reinforcement Learning tem sido objeto de muitas pesquisas e avanços, e está se tornando cada vez mais comum em muitas aplicações. À medida que a tecnologia continua a evoluir, é provável que o Aprendizado por Reforço se torne uma ferramenta ainda mais poderosa para resolver problemas complexos em muitas áreas diferentes.

# Referências

<https://repositorio.unesp.br/handle/11449/239535>

<https://people.com.ai/blog/reinforcement-learning/#:~:text=De%20onde%20surgiu%20o%20Reinforcement,a%20navegar%20em%20um%20labirinto.>

<https://www.scielo.br/j/ca/a/3YPYK8P6ff97tgJwJn8xhXs/?lang=pt>

<https://lamfo-unb.https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/.io/2017/07/27/tres-tipos-am/>

<https://medium.com/turing-talks/aprendizado-por-refor%C3%A7o-1-introdu%C3%A7%C3%A3o-7382ebb641ab>