Tópicos especiais em Machine Learning

Prof. Henrique Batista da Silva

PUC Minas

Unidade 1 - Introdução ao Aprendizado por Reforço

Introdução

Reinforcement Learning

Reinforcement Learning existe desde a década de 1950, produzindo muitas aplicações interessantes ao longo do anos, particularmente em jogos (por exemplo, TD-Gammon, um programa de jogo de gamão)

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Reinforcement Learning

Uma revolução ocorreu em 2013, quando pesquisadores de da DeepMind demonstraram um sistema que poderia aprender a jogar praticamente qualquer jogo do Atari do zero

Playing Atari with Deep Reinforcement Learning

https://arxiv.org/abs/1312.5602

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Reinforcement Learning

Em alguns casos até superando humanos

Human-level control through deep reinforcement learning

https://storage.googleapis.com/deepmind-media/dqn/DQNNaturePaper.pdf

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Reinforcement Learning

Na maioria dos casos, usando apenas pixels brutos como entradas e sem nenhum conhecimento prévio das regras dos jogos.

DeepMind's system learning to play Space Invaders

https://homl.info/dqn3

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Reinforcement Learning

Nesta evolução, em março de 2016 vitória de seu sistema AlphaGo contra Lee Sedol, um lendário jogador profissional do jogo de Go, e em maio de 2017 contra Ke Jie, o campeão mundial.

A DeepMind foi comprada pelo Google por mais de US$ 500 milhões em 2014.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Aprendendo a otimizar recompensas

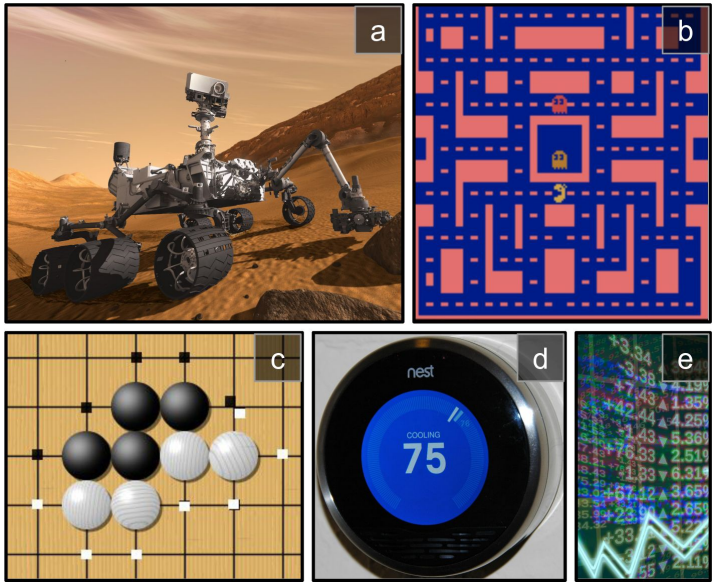
Aprendendo a otimizar recompensas

No Aprendizado por Reforço, um agente faz observações e ações dentro de um ambiente e, em troca, recebe recompensas. Ele aprende a agir de forma a maximizar suas recompensas

O agente atua no ambiente e aprende por tentativa e erro a maximizar seu “prazer” e minimizar sua “dor”.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Aprendendo a otimizar recompensas



Exemplos de Aprendizado por Reforço: (a) robótica, (b) Ms. Pac-Man, (c) Go player, (d) termostato, (e) operador automático

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Robótica

O agente pode ser o programa que controla um robô.

O agente observa o ambiente por meio de um conjunto de sensores e câmeras, e suas ações consistem em enviar sinais para acionar motores. Ele pode ser programado para obter recompensas **positivas** sempre que se aproximar do destino alvo e recompensas **negativas** sempre que perder tempo ou seguir na direção errada.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Jogos

O agente pode ser o programa que controla o Pac-Man. Nesse caso, o ambiente é uma simulação do jogo Atari, as ações são as nove posições possíveis do joystick, as observações são capturas de tela e as recompensas são apenas os pontos do jogo.

Da mesma forma, o agente pode ser o programa jogando um jogo de tabuleiro como Go.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Termostato

O agente pode ser um termostato inteligente, obtendo recompensas positivas sempre que estiver próximo da temperatura alvo e economiza energia, e recompensas negativas quando humanos precisam ajustar a temperatura, então o agente deve aprender a antecipar as necessidades humanas.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Mercado de ações

O agente pode observar os preços do mercado de ações e decidir quanto comprar ou vender a cada segundo. As recompensas são obviamente os ganhos e perdas monetárias.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Mercado de ações

Observe que pode não haver nenhuma recompensa positiva; por exemplo, o agente pode se movimentar em um labirinto, recebendo uma recompensa negativa a cada passo de tempo, então é melhor encontrar a saída o mais rápido possível!

Existem muitos outros exemplos de tarefas para as quais o Aprendizado por Reforço é adequado, como carros autônomos, sistemas de recomendação, etc.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações

Política de ações (policy)

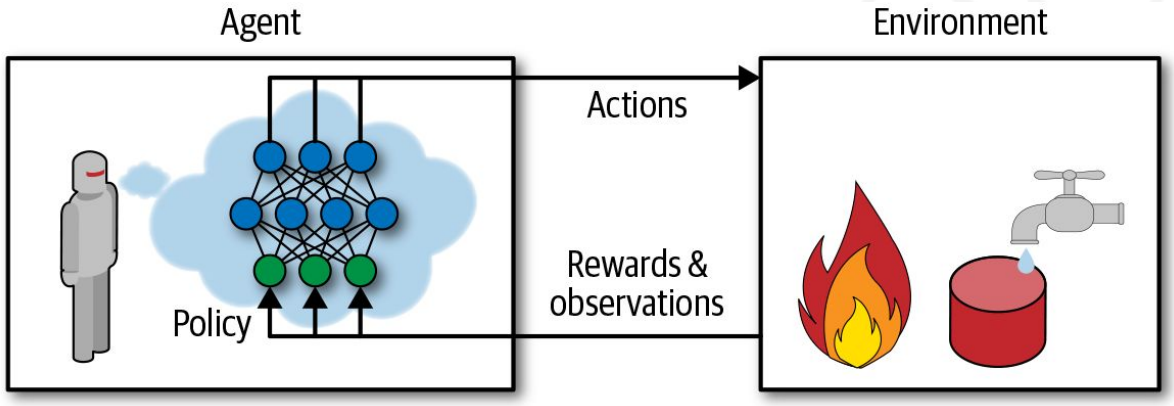
O algoritmo que um agente de software usa para determinar suas ações.

Pode ser uma rede neural que recebe observações como entradas e emite a ação a ser tomada

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)

Aprendizado por reforço usando uma política de rede neural



Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)

A política pode ser qualquer algoritmo e não precisa ser determinista.

Por exemplo, considere um aspirador de pó robótico cuja recompensa é a quantidade de poeira que ele pega em 30 minutos. Sua política poderia ser avançar com alguma probabilidade “p” a cada segundo, ou girar aleatoriamente para a esquerda ou para a direita com probabilidade “1 – p” e ângulo de rotação “-r” e “+r”.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)

Como essa política envolve alguma aleatoriedade, ela é chamada de política estocástica. O robô terá uma trajetória errática, o que garante que eventualmente chegará a qualquer lugar que possa alcançar a poeira. A questão é, quanta poeira ele vai pegar em 30 minutos?

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)

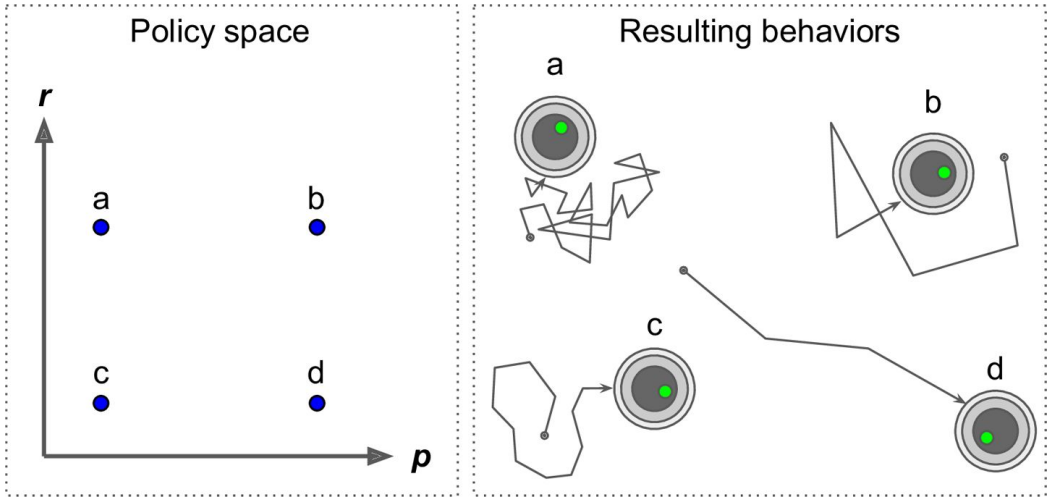
Como você treinaria um robô desses?

Existem apenas dois parâmetros de política que você pode ajustar: a probabilidade “p” e a faixa de ângulo “r”.

Um algoritmo de aprendizado possível poderia ser experimentar muitos valores diferentes para esses parâmetros e escolher a combinação que apresenta o melhor desempenho

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)



Este é um exemplo de busca de políticas (por força bruta). Espaço de busca muito grande (muito difícil encontrar o melhor conjunto de parâmetros)

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)

Outra forma é usando algoritmos genéticos.

Por exemplo, você pode criar aleatoriamente uma primeira geração de 100 políticas e testá-las, depois “matar” as 80 piores e fazer com que os 20 sobreviventes produzam 4 descendentes cada.

Um filho é uma cópia de seu pai mais alguma variação aleatória.

As políticas sobreviventes e seus descendentes juntos constituem a segunda geração.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Política de ações (policy)

O processo continua de geração em geração até encontrar um conjunto de parâmetros satisfatórios.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

OpenAI Gym

OpenAI Gym

No Aprendizado por Reforço,para treinar um agente, é necessário um ambiente de trabalho.

Para programar um agente que aprenda a jogar um jogo Atari, você precisará de um simulador de jogo Atari.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

OpenAI Gym

O OpenAI Gym é um kit de ferramentas que fornece uma ampla variedade de ambientes simulados (jogos Atari, jogos de tabuleiro, simulações físicas 2D e 3D e assim por diante), para treinar agentes, compará-los ou desenvolver novos algoritmos de RL.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

OpenAI Gym

Prática: Abra o Google Colab e utilize o material de aula

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Um exemplo de uma policy

OpenAI Gym

Prática: Abra o Google Colab e utilize o material de aula

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Neural Network Policies

Neural Network Policies

Vamos criar uma rede neural que receberá observações como entradas e emitirá as probabilidades de ações a serem tomadas para cada observação.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

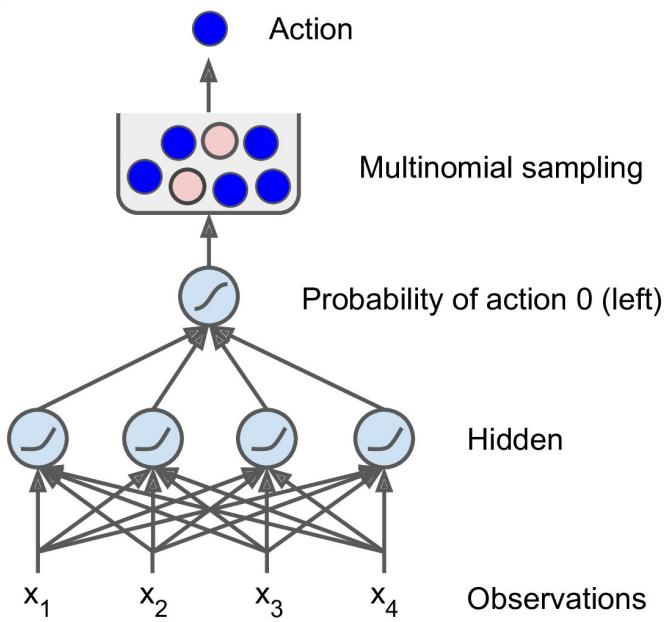
Neural Network Policies

Para escolher uma ação, a rede estimará uma probabilidade para cada ação. Selecionaremos uma ação aleatoriamente de acordo com as probabilidades estimadas.

No caso do ambiente Cart-Pole, existem apenas duas ações possíveis (esquerda ou direita), então precisamos apenas de um neurônio de saída: ele produzirá a probabilidade “p” da ação 0 (esquerda) e a probabilidade de ação 1 (direita) será 1 - p.

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Neural Network Policies



Por exemplo, se ele produzir 0,7, escolhemos a ação 0 com 70% de probabilidade ou a ação 1 com 30% de probabilidade.

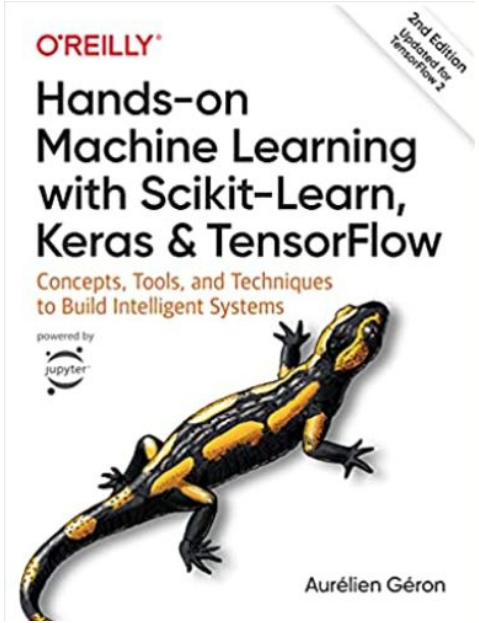
Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

OpenAI Gym

Prática: Abra o Google Colab e utilize o material de aula

Fonte: Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow. O'Reilly Media; 2nd ed. 2019

Principais Referências



Aurélien Géron. **Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow:** Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media; 2nd ed**.** 2019