Unyleya
EDUCACIONAL

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

Além de aprender uma função de recompensa de um especialista, podemos também aprender diretamente uma política para ter um desempenho comparável ao especialista.

Útil se tivermos uma política especializada que seja apenas aproximadamente ideal.

Abbeedl e Andrew (2004) propuseram um algoritmo que usa um MDP (*Markov Decision Process*) e uma política de um "especialista" aproximadamente ideal, e em seguida, aprende uma política com desempenho comparável ou melhor do que a política do especialista, usando exploração mínima.

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

Essa propriedade mínima de exploração acaba sendo muito útil em tarefas frágeis como um voo autônomo de helicóptero.

Um algoritmo tradicional de aprendizagem por reforço poderia começar a explorar aleatoriamente, o que quase certamente levaria a um acidente de helicóptero no primeiro teste.

Idealmente, poderíamos usar dados de especialistas para começar uma política de linha de base que pode ser melhorada com segurança ao longo do tempo.

Essa política de linha de base deve ser significativamente melhor do que uma política inicializada aleatoriamente, o que acelera a convergência.

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

Algoritmo de aprendizagem

Ideia principal: usar ensaios da política de especialistas para obter informações sobre o MDP subjacente e, em seguida, executar iterativamente uma melhor estimativa da política ótima para o MDP real.

A execução de uma política também nos fornece dados sobre as transições do ambiente, que pode ser usada para melhorar a precisão do MDP estimado.

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

Funcionamento do algoritmo em um ambiente discreto

Primeiro, usamos uma política de especialistas para aprender sobre o MDP:

1. execute uma quantidade fixa de testes usando a política de especialistas, registrando cada trajetória de ação do estado;

2. estime as probabilidades de transição para cada par de ação do estado usando os dados registrados por meio da estimativa de máxima verossimilhança;

3. estime o valor da política especializada, calculando a média da recompensa total em cada tentativa.

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

Aprende-se uma nova política:

1. aprenda uma política ótima para o MDP estimado usando qualquer algoritmo de aprendizagem por reforço padrão;

2. teste a política de aprendizado no ambiente real;

3. se o desempenho não for suficientemente próximo do valor da política especializada, adicione as trajetórias de ação do estado dessa avaliação ao conjunto de treinamento e repita o procedimento para aprender uma nova política.

Vantagem: em cada estágio, a política que está sendo testada é a melhor estimativa para a política ótima do sistema. Há uma diminuição na exploração, mas a ideia central do aprendizado é que podemos supor que a política especializada já está próxima do ideal.

APRENDIZAGEM DA APRENDIZAGEM: APRENDER COM UM ESPECIALISTA

Referências

RUSSELL, S.; NORVIG, P. **Artificial Intelligence:** A Modern Approach. 3.ed. New Jersey: Pearson Education, 2010.

ABBEEL, P.; ANDREW, Y. Apprenticeship learning via inverse reinforcement learning. **Proceedings of the twenty-first international conference on Machine learning**. ACM, 2004.

Obrigada!

hulianeufrn@gmail.com