



APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Jogo da Velha

Agentes podem aprender sem supervisão especializada

Os tipos de problemas mais adequados são aqueles complexos, nos quais parece não haver uma solução óbvia ou facilmente programável.

Jogo da velha: construir um agente que identifica as **imperfeições do oponente** e **aprende a maximizar as chances de sua vitória**.

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Jogo da Velha

Agentes podem aprender sem supervisão especializada

Os tipos de problemas mais adequados são aqueles complexos, nos quais parece não haver uma solução óbvia ou facilmente programável.

Cenários: (1) um jogador experiente jamais perde esse jogo; (2) considere um adversário imperfeito, que às vezes toma decisões incorretas.

X	0	0
0	X	X
		X

Fonte: a autora

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Jogo da Velha

Construir um agente que identifica as **imperfeições do oponente** e **aprende a maximizar as chances de sua vitória?**



APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Jogo da Velha

Um jogo simples, porém, uma solução satisfatória não pode ser obtida através de técnicas clássicas.

A técnica **MiniMax**, por exemplo, baseada na teoria dos jogos, assume um comportamento particular do oponente.

O comportamento do oponente não pode ser conhecido *a priori* na **aprendizagem por reforço**, mas é viável a sua identificação por meio de interações.

A performance do algoritmo passaria a depender da taxa de aprendizagem e da suposta habilidade do adversário.

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Jogo da Velha

Um jogo simples, porém, uma solução satisfatória não pode ser obtida através de técnicas clássicas.

A técnica **MiniMax**, por exemplo, baseada na teoria dos jogos, assume um comportamento particular do oponente.

O comportamento do oponente não pode ser conhecido *a priori* na **aprendizagem por reforço**, mas é viável a sua identificação por meio de interações.

A performance do algoritmo passaria a depender da taxa de aprendizagem e da suposta habilidade do adversário.

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Passos para aplicar aprendizagem por reforço ao jogo da velha

Criamos uma tabela de números, $V(s)$, com uma entrada para cada estado s do jogo

$V(s)$ é a estimativa mais recente de vencermos o jogo a partir do estado s

$V(s)$ é o valor do estado s

A tabela V representa a função valor

Um estado s_1 é considerado melhor do que um estado s_2 , se $V(s_1) > V(s_2)$

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Passos para aplicar aprendizagem por reforço ao jogo da velha

Uma linha com três X's tem probabilidade 1 de vitória— **o agente ganha o jogo**

Uma linha com três O's tem probabilidade 0 de vitória — **o agente perdeu o jogo**

Para os demais estados, chutamos probabilidade $\frac{1}{2}$ de vitória

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Jogamos contra o oponente

Na maioria das vezes o agente escolhe o movimento que o leva ao estado com maior valor, isto é, escolhe os estados com maior probabilidade de vitória (*Exploitation*).

Às vezes o agente seleciona aleatoriamente alguns dos demais movimentos possíveis (*Exploration*).

Atualizamos a tabela V, com isto é possível obter estimativas mais precisas das probabilidades de vencer.

APLICAÇÕES DA APRENDIZAGEM POR REFORÇO

Para a aprendizagem, pode-se utilizar o método de aprendizagem por diferença temporal. Com este método os valores contidos na tabela V são atualizados de acordo com a seguinte regra:

$$V(s) \leftarrow V(s) + \alpha [V(s') - V(s)]$$

Sendo, α é a taxa de aprendizagem, $V(s)$ é o valor da probabilidade atualizada a cada fim de partida $V(s')$ tem o valor inicial de 1 no reforço positivo, ou seja quando há uma vitória e 0 no reforço negativo, ou seja quando há uma derrota do computador.

Ao final de cada partida todos os estados contidos na lista temporária têm suas listas de probabilidades atualizadas proporcionalmente conforme a equação acima.

Ao fim do treinamento o **APRENDIZADO POR REFORÇO** possui uma lista permanente que especifica as melhores jogadas a serem executadas a cada estado aprendido.



Obrigada!

hulianeufrn@gmail.com