Relatório do Projeto 2 de Inteligência Artificial

Mihail Brinza 83533 Ricardo Brancas 83557

7 de Dezembro de 2017

1 Métodos de Classificação

Para classificar as plavras, considerámos as seguintes *features*:

1. Tamanho da palavra;
2. Número de vogais;
3. Paridade do número de vogais;
4. Paridade do número de consoantes;
5. Número de caracteres "a";

2 Métodos de Regressão

3 Aprendizagem por Reforço

3.1 Trajetórias Aprendidas

3.1.1 Exemplo 1

$5_0 \xrightarrow{0} 6_1 \xrightarrow{0} 6_1 \xrightarrow{0} 6_1 \xrightarrow{0} 6_{(1)}$

3.1.2 Exemplo 2

$5_0 \xrightarrow{0} 6_1 \xrightarrow{0} 1_0 \xrightarrow{1} 0_1 \xrightarrow{1} 0_{(1)}$

3.2 Modelo do Mundo

3.2.1 Exemplo 1

Neste primeiro exemplo o ambiente consiste numa série de quadrículas sequênciais, tal como demonstrado na figura 1, em que a ação 1 corresponde a dar um passo para a esquerda e a ação 0 corresponde a dar um passo para a direita. Tentar andar para fora dos limites não tem qualquer efeito. Para além disto, no estado 5 a ação 0 não é determinística; ao realizar esta ação, o agente pode ir ter ao estado 6 ou permanecer no estado 5.

Por fim determinámos que o agente é recompensado ($r = 1$) sempre que o estado inicial (antes da ação) é um estado limite. (i.e. 0 ou 6), não sendo recompensado nas outras situações ($r = 0$).

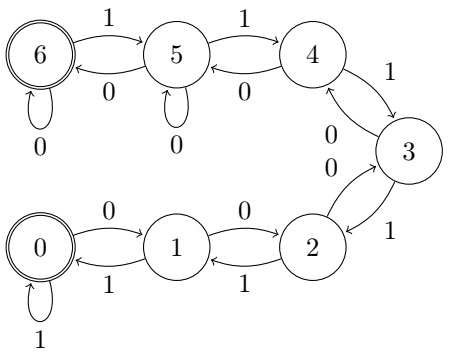

Figura 1: Ambiente 1. Os nós com duplo contorno são os nós de recompensa.

Como tal, quando começamos no estado 5 o melhor curso de ação é andar rapidamente para o estado limite mais próximo, o 6, e depois mantermo-nos lá.

3.2.2 Exemplo 2

No segundo exemplo o ambiente é muito semelhante ao primeiro com exceção do estado 6. Agora quando tentamos andar para a direita (ação 0) nesse estado voltamos para o estado 1, tal como representado na figura 2. A função de recompensa mantem-se também inalterada, sendo 1 quando o estado inicial é o zero ou o seis, e 0 caso contrário.

Como tal [quando começamos no estado 5] já não é possível usar a estratégia anterior de ficar parado no estado 6. Assim, o melhor a fazer nesta situação é tentermo-nos dirigir para o estado 0, de modo a maximizar a recompensa a longo prazo, que é exatamente o que obtemos com o algoritmo Q-learning.

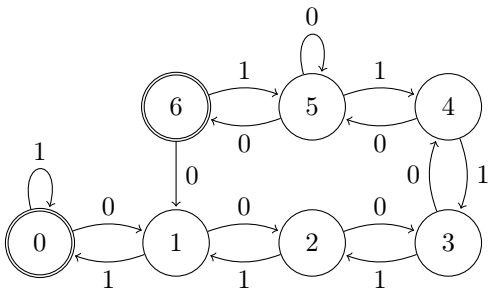

Figura 2: Ambiente 2. Os nós com duplo contorno são os nós de recompensa.