

2º Projeto Inteligência Artificial – 2018/2019

Grupo 105

André Guerra nº86382

Tomás Zaki nº79690

P1 – Inferência exata em Redes Bayesianas

Em relação aos resultados obtidos, correspondem aos resultados esperados nos testes fornecidos pelo corpo docente.

A complexidade deste algoritmo para a inferência exata de uma variável dada uma evidência com valores desconhecidos, corresponde a $O(2^n)$ em que n corresponde ao número de variáveis presentes na rede.

Um dos métodos alternativos seria a implementação do algoritmo de eliminação de variáveis, que consiste em fazer as somas, na árvore de avaliação, da direita para a esquerda guardando os resultados intermédios (fatores) para evitar ter de os voltar a calcular.

No algoritmo por nós implementados, estes fatores são sempre recalculados sempre, que queremos calcular uma probabilidade de uma variável à 'posteriori' (dada uma evidência com valores desconhecidos).

P2 – Aprendizagem por Reforço

EXER 1

Neste tipo de aprendizagem, os agentes podem aprender o que fazer sem conhecerem o ambiente em que se inserem. Na ausência total de conhecimento sobre o ambiente, um agente pode aprender um modelo de transição para os seus próprios movimentos, no entanto, o agente precisa de saber se algo de bom aconteceu mediante uma ação sua, ou pelo contrário, se algo de mau aconteceu.

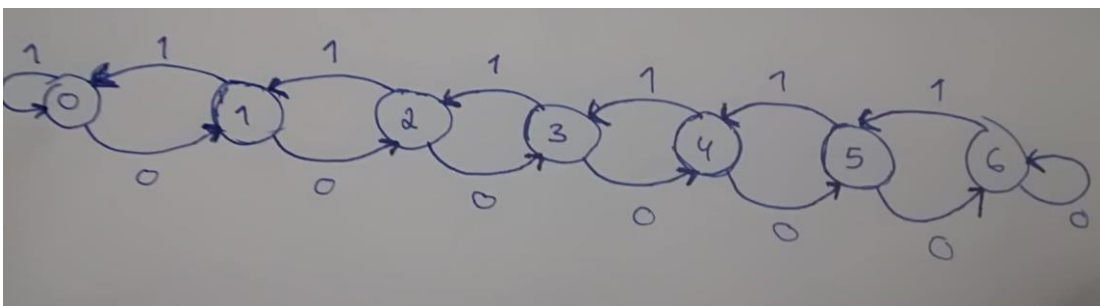
A este tipo de informação chamamos recompensa. Uma recompensa serve assim para definir políticas ótimas nos processos de decisão (será ótima se maximiza a recompensa total esperada). Assim, esta abordagem consiste em usar recompensas observadas para aprender uma política ótima para o ambiente.

Ao implementarmos o Q-learning, definimos o valor de alpha, ou learning rate, como 0.2. Learning rate corresponde à quantidade de informação nova que vamos usar para calcular o novo valor de Q. Se o valor for 0 o agente não aprende nada de novo e se for 1 o agente considerará sempre apenas a nova informação, ignorando completamente a antiga.

Através de uma análise do traj gerado do exerc 1, chegamos a conclusão que o mapa total do ambiente e suas possíveis ações seria o seguinte:

Estados: {0,1,2,3,4,5,6}

Ações: {0,1}



Função de recompensa:

$R(\text{Estados } 0 \text{ e } 6) = 1$

$R(\text{Estados } 1 \text{ a } 5) = 0$

Política Ótima:

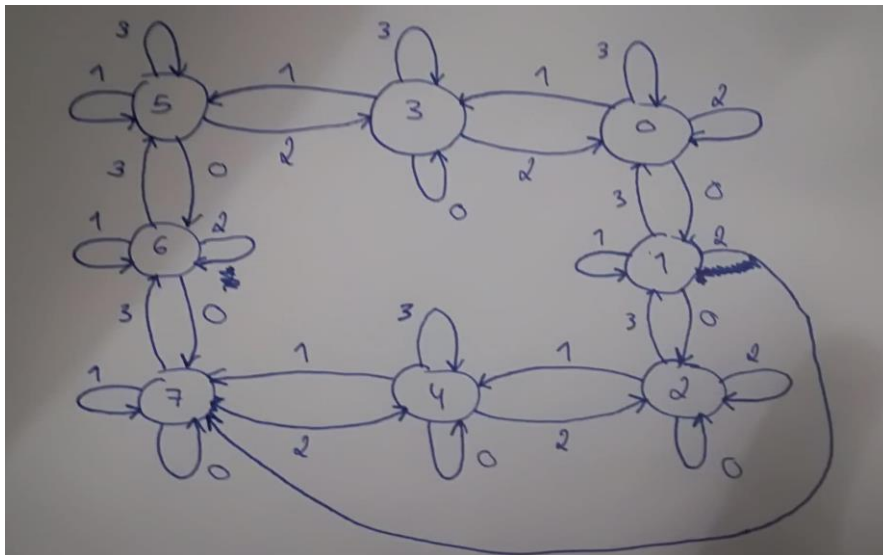
Depois de executado Q-Learning, a política ótima é:

0 -> 1	1 -> 1	(Estado -> Ação Ótima)
2 -> 1	3 -> 1	
4 -> 0	5 -> 0	
6 -> 0		

EXER 2

Estados: {0,1,2,3,4,5,6,7}

Ações: {0,1,2,3}



Função de recompensa:

$R(\text{Estados } 0 \text{ a } 6) = -1$

$R(\text{Estado } 7) = 0$

Política Ótima:

Depois de executado Q-Learning, a política ótima é:

0 -> 0	1 -> 2	(Estado -> Ação Ótima)
2 -> 1	3 -> 2	
4 -> 1	5 -> 0	
6 -> 0	7 -> 1	