Universidade Federal de Santa Catarina
Inteligência Artificial
Professor: Jomi F. Hübner

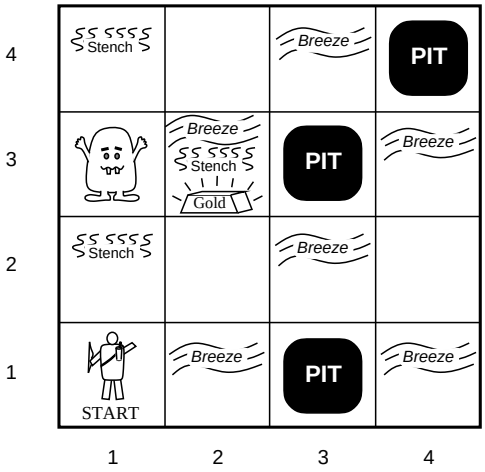
DAS / PPGEAS

Exercícios de Aprendizado por Reforço

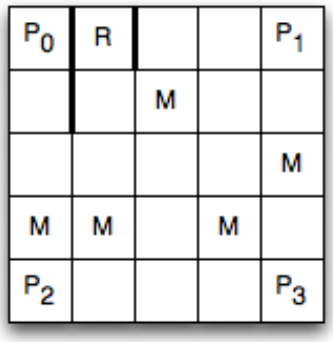
Para cada um dos problemas enunciados abaixo, avalie a aplicação da técnica de aprendizado por reforço (AR) procurando definir:

- O MDP (quais os estados, ações, recompensas e transições)
- Uma possível política que será aprendida
- Parâmetros como γ
- A complexidade do processo de aprendizado

1. O problema clássico do Wumpus World:



2. Considerando o cenário da figura abaixo, onde há 25 lugares para um agente estar:



   Um prêmio *pode* aparecer em algum dos cantos. Quando o agente está no mesmo lugar do prêmio, ele recebe uma recompensa de 10 e o prêmio desaparece. Nos casos onde não há prêmio no cenário, há uma probabilidade do prêmio aparecer em algum dos cantos. Monstros podem aparecer e desaparecer em qualquer momento nos lugares marcados com M. O agente sofre dano se estiver no mesmo lugar que um monstro. Sempre que estiver com dano, o agente recebe uma punição de 10. Um agente com dano é consertado se visitar a estação de reparação marcada com R. O agente deve aprender a coletar prêmios percebendo o ambiente e se movimentando em quatro direções.

3. Um robô que tem a tarefa de percorrer o ambiente procurando por latas de refrigerante. Quanto encontrar, deve coleta-las e levar para um depósito. Considere que o robô tem sensores para detectar as latas e braços para pegá-las. O robô funciona com uma bateria recarregável.

   As decisões que deve tomar é uma entre as seguintes:

   - pegar lata
   - soltar lata
   - ficar parado
   - procurar latas
   - ir para o depósito
   - ir para a estação de recarga da bateria

4. Um controlador de elevadores. São basicamente duas decisões:

   - Se tem chamados, qual atender
   - Se não há chamada, onde esperar

   Considere um prédio de 10 andares e 4 elevadores.

5. O problema enunciado abaixo, retirado de "*Treinando seu Cérebro.* Reader's Digest, 2002. pg. 117.", era para ser resolvido pelo leitor.

   > O romancista inglês Charles Dickens nunca foi, até onde se sabe, um fã de passatempos, mas ele bem poderia ter sido um. Sua mente era capaz de produzir frases como "..e o infortunado Millier passou a sentir-se tão fora do seu elemento como um golfinho na guarda de uma sentinela", mostrando criatividade para idéias contraditórias, que é uma das chaves para se decifrar um enigma. E o último livro de Dickens, *O mistério de Edwin Drood*, era uma espécie de quebra-cabeça literário, um romance de mistério tornado ainda mais enigmático pelo fato de o grande escritor ter morrido antes de concluí-lo. Mas neste passatempo, não existe mistério nenhum; basta um pouco de pensamento lógico ao examinar os quadrados abaixo. Em cada

fila horizontal, coluna vertical e linha diagonal, todas formadas por sete quadradinhos, aparecem as sete letras D, I, C, K, E, N e S, em uma certa ordem. Preenchemos alguns dos quadrados para servirem de ponto de partida, e as pistas abaixo irão ajudar você a preencher os restantes.

(a) Os quadrados 3 e 42 contêm a mesma letra;

(b) Os quadrados 4 e 26 contêm a mesma letra, que é diferente daquela do quadrado 43;

(c) Os quadrados 5 e 49 contêm a mesma letra;

(d) Os quadrados 9 e 36 contêm a mesma letra; e

(e) Os quadrados 22 e 48 contêm a mesma letra.

1 **D**	2	3	4	5	6 **I**	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23 **C**	24	25	26	27	28 **K**
29	30	31	32	33	34	35
36	37	38	39 **E**	40	41 **N**	42
43	44	45	46	47 **S**	48	49

6. O Q-Learning generaliza? O que acontece se o agente precisa decidir uma ação para um estado onde nunca esteve? Seria possível utilizar técnicas de aprendizado indutivo para auxiliar o agente nestes casos?