PUCRS - Escola Politécnica
Inteligência Artificial

T2 - Labirinto das Moedas com Aprendizagem por Reforço

1. **Definição**

   O trabalho 2 da disciplina de IA visa fixar e exercitar conceitos relativos aprendizagem de máquina. O trabalho consiste na simulação de um jogo, no qual um agente deve aprender a percorrer um labirinto recolhendo sacos de moedas e encontrar a saída. O objetivo do jogo recolher o máximo de moedas e encontrar a porta de saída. Ao longo do percurso até a porta, o agente deve contornar paredes e recolher sacos de moedas.

2. **Ambiente: Labirinto**

   Será uma matriz 10x10. Da mesma forma que no T1, as paredes serão codificadas com o caracter 1, a entrada com E e a saída com S. Os sacos de moeda serão identificados por M.

```
10
E 0 0 1 0 M 0 0 M 1
0 1 0 M 0 1 0 1 0 M
0 0 M 0 1 1 M 0 1 1
M 1 1 0 M 1 1 M 0 1
M 0 0 0 0 1 1 0 1 1
1 1 1 1 0 1 1 0 1 1
1 0 1 1 0 1 1 M 0 M
M 0 M M 0 1 1 1 1 1
1 M 1 M 0 0 M M 1 1
1 M 1 M 1 M 0 0 0 S
```

3. **Movimentação do Agente**

   Somente o agente pode se mover no ambiente. Ele pode se mover nas seguintes direções: ←→↑↓, uma célula de cada vez. Agentes não caminham sobre paredes e nem as transpassam. Sua percepção é de duas casas em qualquer direção. Ele não sabe o que há atrás das paredes. Para recolher um saco de moedas, ele deve estar na mesma célula do saco. Cabe mencionar que no inicio da simulação, o agente não sabe a localização de nenhum elemento do ambiente, a não ser as coordenadas (linha e coluna) da célula que corresponde à saída. O agente sempre inicia na célula (0,0). Se o agente bater em uma parede, ele perder o jogo (game over). O agente anda uma célula por vez e tem como objetivo achar a saída e recolher as moedas do ambiente.

4. **Pontuação do agente no jogo**

   Ao longo da execução, o agente deve manter os sacos de moedas que recolheu. Na tela devem ser exibida a quantidade atual de moedas coletadas pelo agente. Considere que em cada saco há 50 moedas. A pontuação geral do jogo considera:

   (a) quantidade de moedas coletadas (incrementa seu score)

(b) a distância que o agente está da saída (incrementa seu score)

(c) colisão com as paredes (decrementa seu score)

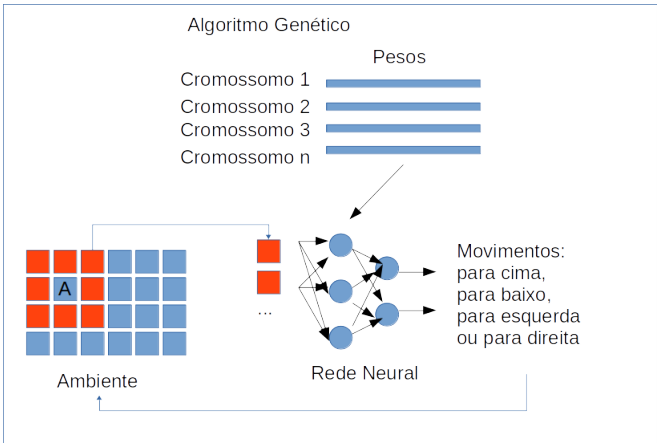
Você pode acrescentar outros elementos como o tempo que ele leva para chegar à saida.

5. **Processo de Aprendizagem**

   O agente deve aprender a se mover corretamente no ambiente usando aprendizagem por reforço. A escolha dos movimentos deve ser feita por uma rede neural cujos pesos devem ser definidos por um algoritmo genético. Cada cromossomo do algoritmo representa uma instância dos pesos da rede. A função de aptidão a ser definida deve favorecer o comportamento correto do agente. O objetivo é evoluir os pesos até que a rede consiga aprender a função que permite que o agente encontre a saída com a maior quantidade de moedas.

6. **Simulação**

   A simulação deve exibir informações que permitam acompanhar as decisões e ações realizadas pelo agente. Ao final da simulação, deve ser exibido o sucesso ou o insucesso do agente, bem como a quantidade de moedas recolhidas e a sua pontuação.



7. **Forma de Avaliação**

   (a) O trabalho pode ser realizado pelo mesmo grupo do T1.

   (b) A **apresentação do trabalho** será em aula e terá dois momentos. No primeiro momento, o grupo exibe **um vídeo curto de até 3 minutos**, mostrando o trabalho (evolução do agente). O vídeo deve mostrar todas as funcionalidades implementadas. No segundo, momento, serão realizadas perguntas sobre a implementação.

(c) A entrega dos fontes, do executável e do vídeo no moodle será dia: 02/12/2021. **A apresentação também será nesse dia.**Todos os integrantes do grupo devem estar presentes na apresentação do trabalho.

(d) A nota será distribuída da seguinte forma,

i. Rede Neural: responsável pelas decisões tomadas pelo agente. A entrada será baseada na percepção do agente (conteúdo das células ao seu redor) e outras informações que julgar relevante. A saída da rede será o movimento $\leftarrow\rightarrow\uparrow\downarrow$. Os pesos da rede serão definidos pelos cromossomos de um algoritmo genético. Modelagem e propagação da rede: 2,0 **pontos**

ii. Algoritmo Genético: a codificação dos cromossomos usará números reais (pesos da rede neural). Para implementar os operadores genéticos consulte o pdf sobre algoritmos genéticos que está disponível no moodle. Construa uma função de aptidão que valorize o bom desempenho do agente (quantidade de moedas recolhidas, distância percorrida sem bater nas paredes e distância que está da saída): 4,0 **pontos**

iii. Simulação: Cada cromossomo definirá um conjunto de pesos para a rede. A rede definirá passo a passo o comportamento do agente. O score do agente no jogo definirá o aptidão do cromossomo. O objetivo é encontrar o conjunto de pesos que melhor atende ao problema (mais moedas e busca pela saída). Sua simulação deve permitir o acompanhamento da evolução do Algoritmo Genético, as decisões da rede e as ações executadas pelo agente: 3,0 **ponto**s.

iv. Video do trabalho: **1,0 ponto**.