INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

ADEETC

Inteligência Artificial e Sistema Autonomos

2º Semestre 2016/2017

3º Trabalho prático

Rui Santos nº 39286

2 de Maio 2017

Conteúdo

1 Problema

No 3º trabalho prático da cadeira de Inteligência Artificial para Sistemas Autonomos, é-nos pedido com base nos trabalhos anteriores e na nova materia lecionada, que simulemos 4 comportamentos diferentes sobre um Agente Propector.

- Concepção e implementação de um agente reactivo para recolha de alvos com base em esquemas comportamentais reactivos;

- Concepção e implementação de um agente para recolha de alvos, tendo por base raciocinio automatico com base em procura em espaços de estados, capaz de minimizar a distância percorrida na recolha de alvos;

- Concepção e implementação de um agente para rercolha de alvos, tendo por base processos de decisão de Markov, capaz de minimizar a distância percorrida na recolha de alvos;

- Concepção e implementação de um agente para recolha de alvos, capaz de aprender a partir da experiência, tendo por base mecanismos de aprendizagem por reforço.

2 Modelo

2.1 Comportamento Reactivo

O agente que implementa um comportamento reactivo é exatamente o que o nome em si indica, através de percepções do ambiente onde se encontra, internamente forma uma reacção composta por todas as acções que dão finalidade ao agente, e determina qual a acção a escolher. Acções estas que têm hierarquias e prioridades de escolha, de forma que a precisão de escolha de acção é mais elevada.

2.2 Comportamento Deliberativo

O agente que implementa um comportamento deliberativo, é aquele que tem em conta o Passado-Presente-Futuro e que antecipa soluções perante o objectivo que quer atingir, avaliando as mesma com base no custo e utilidade. Este comportamento pode ser dividido em duas partes:

-Raciocinio sobre fins (Finalidade): Em que o agente decide o que fazer de forma a obter objectivos.

-Raciocinio sobre meios (Planeamento): Em que o agente decide como fazer, isto é, quais as acções a executar face aos objectivos, resultando num plano de acções.

Concluindo a tomada de decisões e acções é descrita por:

1. **Observar** o mundo;
2. **Atualizar** crenças;
3. **Reconsiderar** - caso seja necessário:
    3.1 **Deliberar**;
    3.2 **Planear**;
4. **Executar** plano de acção.

2.3 Comportamento Deliberativo - Decisão Markov

Processo de Decisão de Markov, é um processo que implementa uma decisão sequencial, isto é, de acordo com o estado atual, é feita uma escolha para o estado seguinte, baseado em algumas variáveis sendo que os valores dessas mesmas variam consoante as escolhas feitas nessa sequência de estados.

O agente que implementar o Processo de Decisão de Markov enfrenta alguns problemas:

- Utilidade de uma acção depende de uma sequência de decisões;
- Possibilidade de ganhos e perdas;
- Incerteza na decisão;
- Efeito Acumulativo;

Representação do Mundo sob a forma de PDM:

S - Conjunto de estados do mundo

A(s) - Conjunto de acções possíveis no estado $s \epsilon S$

T(s,a,s') - probabilidade de transição de s para s' através de a

R(s,a,s') - retorno esperado na transição de s para s' através de a

γ - taxa de desconto para recompensas diferidas no tempo

t - 0,1,2,... - tempo discreto

De forma a escolher o estado seguinte, é calulada a utilidade do estado com base nas diferentes partes de representação do mundo em forma de PDM:

$$U(s) = max\Sigma T(s,a,s')[R(s,a,s') + \gamma U(s')]$$

2.4 Comportamento Aprendizagem

Neste comportamento implementamos uma novo conceito: Aprendizagem.

Aprendizagem singifica a melhoria de desempenho para um dada tarefa com a experiência, não deve ser confundida por memória.
A aprendizagem por reforço é obtida a partir da intração com o ambiente, estados, acções, reforço - Ganho ou Perda.

Um estado pode evoluir no tempo de acordo com o estados observados, acções realizadas, reforços obtidos e o valor de num estado realizar uma acção. Com isto a nossa aprendizagem irá ser baseada no algoritmo Q-Learning.

Q-Learning inicia num estado, escolhe uma acção possivel nesse estado com derivação de um politica, politica esse por exemplo neste trabalho ε-greedy, executar essa acção e observar o seu reforço, atualizar a sua aprendizagem nesse estado, e repetir o mesmo processo até que o estado atual seja um estado final.

3 Arquitetura

3.1 1ª Fase - Reactivo

Para a simulação desta primeira fase, foram criadas as devidas classes que implementam os comportamentos que dão finalidade ao agente:

- AproximarAlvoDIR - acção que permite ao agente alcançar o alvo;

- EvitarObstaculo - acção que o agente escolhe sempre que a sua percepção do mundo lhe retorna um obstaculo, em que efetivamente roda para a esquerda ou para a direita.

- Contornar - acção que permite ao agente contornar obstáculos.

- Explorar - acção em que aleatoriamente escolhe um movimento para o agente explorar o ambiente.

Recolher - agrupa todas aas acções possiveis do agente numa so classe, de forma a depois poder ser ativada no controlo do mesmo por outras classes.

3.2 2ª Fase - Deliberativo

Para a simulação desta segunda fase que implementa o conceito de procura em espaço de estados do trabalho anterior, foram necessárias a criação das classes:

- PlanPEE - efetivamente cria todas as acondições para que a deliberação do agente seja possivel, recebendo um tipo de procura em espaço de estados (A*, Sôfrega) permitindo a heuristica. Aqui é simulado o plano de acções para o objectivo alvo presente no ambiente.

- OperdadorMover - Visto que agora não existe nenhuma acção predefinida

como o comportamento reactivo, é necessário definir todos os tipos de movimentos possiveis num so estado do ambiente, estes operadores irão ser instânciados na classe seguinte;

- ModeloMundo - classe que trata de ser mediador do mesmo, tendo todas as caracteristicas do mesmo e em que o agente se baseias, desde os Operadores aos estados possiveis e os seus conteudos.

3.3 3ª Fase - PDM

Para a simulação desta fase, foi necessário a criação das classes de implementação dos conceitos anteriormente vistos do Processo de Decisão Markov.

Em diferenciação da fase anterior, o modelo do mundo é visto de outras maneiras pelo PDM, consequentemente foi necessário a seguinte classe:
- ModeloPDMPlan - classe que visa implementar os conceitos de estados S do mundo, Acções A, Transições T e Recompensas R do PDM, e que permite à seguinte classe, dar corpo a um Planeamento PDM.

- PlanPDM - função semelhante à descrita na fase anterior do PlanPEE mas que permite ao agente trabalhar sobre esse plano.

- PDM - classe que calcula a utilidade e politicas de um estado em função do processo de decisão sequencial de markov.

3.4 4ª Fase - Aprendizagem Reforço

Para a implementação da aprendizagem por reforço é necessário a criação de classes que tenham como função os conceitos anteriormentes falados.

- AprendQ - implementa o conceito de Q-Learning com ajuda da classe SelAccaoEGreedy.

- SelAccaoEGreedy - Através do algoritmo de escolha seleciona uma acção com base num valor escolhido pelo utilizador ε.

- MemoriaEsparsa - com base nas acções efetuadas, e escolhidas pelo Q-Learning, são guardadas nesta classe em memoria, para assim serem comparadas mais tarde e atualizadas.

- MecAprend - classe que agrupa todos os conceitos anteriormente referidos neste tópico e que permite dar vida a um novo controlo no agente.