Departamento de Informática
Universidad Carlos III de Madrid

Grado en Ingeniería Informática

Aprendizaje Automático

Curso 2015-2016

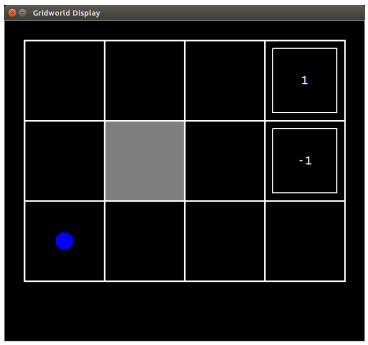
Tutorial 4: Introducción al Aprendizaje por Refuerzo

5 de abril de 2016

- El objetivo de este tutorial es familiarizarse con el software donde se tiene que implementar *Q-learning* en el dominio del Pac-Man.
- Se puede realizar en Linux, Windows o Mac.
- Es importante ir realizando los ejercicios en orden.

1. Ejercicio 1

1. Descargar el código fuente del Pac-Man de Aula Global (fichero **refuerzo.zip**).
2. Descomprimir la carpeta *refuerzo*.
3. Abrir un terminal o consola de comandos y entrar dentro de la carpeta *refuerzo*.
4. Para empezar vamos a ejecutar GridWorld en el modo de control manual, que utiliza las teclas de flecha.
   python gridworld.py -m
5. Ejecútalo ahora con el agente por defecto
   python gridworld.py



Preguntas

1. ¿Qué aparece por terminal cuando se realizan los movimientos en el laberinto?

2. ¿Qué clase de movimiento realiza el agente por defecto?
3. Dibujar el MDP.
4. Cambia el laberinto que viene por defecto por AAGrid. Dibujar el MDP correspondiente a este nuevo laberinto. *python gridworld.py -g AAGrid -n 0*
5. Abre el fichero *gridworld.py* y analiza cómo se almacenan los laberintos en el código.
6. Crea un laberinto nuevo.
7. ¿Se pueden sacar varias políticas óptimas? Describe todas las políticas óptimas para este problema.

2. Ejercicio 2

- Se tienen que implementar los métodos para el agente *QLearningAgent* en fichero *ValueIterationAgent*:
  - Constructor de la clase.
  - *readQtable* leer de fichero los valores de la tabla Q.
  - *writeQtable* escribir en un fichero los valores de la tabla Q.
  - *computeActionFromValues(state)* extrae la mejor acción teniendo en cuenta el valor de la tabla Q. Cuando se trate de un estado meta no tiene que devolver ninguna acción.
  - *computeQValueFromValues(state, action)* devuelve el Q-valor del par estado-acción.

Consideraciones:

- Escribir en un fichero la tabla $Q^{\star}(s)$ para $\gamma = 0{,}9$.
- El agente tiene que inicializarse con la tabla leída por fichero.
- El agente tiene que devolver la acción correspondiente con la política más avara, es decir, la que maximiza Q.

3. Ejercicio 3

Ahora vamos a crear un MDP estocástico, para ello cambia el parámetro de -n 0.3 y comprueba las tuplas de aprendizaje que se crean en el programa. Genera una nueva tabla Q y responde a las siguientes preguntas:

1. Dibujar el MDP correspondiente a este nuevo problema.
2. Crear una tabla Q para este nuevo problema ¿Es Q óptima?
3. ¿Se genera la política óptima?

4. Documentación a entregar

El tutorial se debe realizar **obligatoriamente** en grupos de 2 personas y se entregará a través del entregador que se publicará en Aula Global **hasta las 23:55 horas del miércoles 13 de abril de 2016**.

El nombre del archivo comprimido debe contener los últimos 6 dígitos del NIA de los dos alumnos, `ej.tutorial3-123456-234567.zip` El archivo comprimido debe incluir lo siguiente:

1. Un documento en formato **PDF** que debe contener:
   - Las respuestas a todas las preguntas planteadas en los ejercicios.
   - Descripción de las funciones implementadas para cumplir los requisitos del enunciado.
   - Laberinto creado en el ejercicio 1.
   - Los ficheros generados con las distintas tablas Q.
2. El archivo de código fuente modificado por los alumnos **valueIterationAgents.py**.

El peso de este tutorial sobre la nota final de la asignatura es de 0.3 puntos.

ANEXO: Parámetros útiles del código

Para poder ver todas las opciones disponibles hay que introducir el siguiente comando:

```
python gridworld.py --help
```

Los principales argumentos que se pueden cambiar son:

- **-d descuento** Parámetro de descuento. Por defecto 0,9
- **-n ruido** Hace que las acciones sean no deterministas. Por defecto es 0,2.
- **-i episodios** Número de rondas por cada iteración. Por defecto es 10.
- **-g laberinto** Laberinto usado. Por defecto es *BookGrid*. Se puede elegir entre *BookGrid, BridgeGrid, CliffGrid, MazeGrid* y *getAAGrid*.
- **-a agente** Tipos de agente. Por defecto es *BookGrid*. Se puede elegir entre *random*, *value* y *q*.
- **-m** Modo manual.
- **-r recompensa** por vida para un intervalo de tiempo.