Departamento de Informática
Universidad Carlos III de Madrid

Grado en Ingeniería Informática

Aprendizaje Automático

Curso 2020-2021

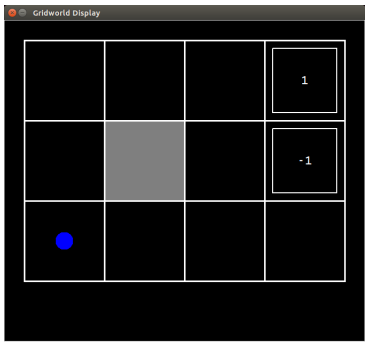
Tutorial 4: Introducción al Aprendizaje por Refuerzo

15 de abril de 2021

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

1. Ejercicio 1

1. Descargar el código fuente para este tutorial de Aula Global (fichero **tutorialRefuerzo.tar.gz**).
2. Descomprimir el fichero anterior.
3. Abrir un terminal o consola de comandos y entrar dentro de la carpeta *tutorialRefuerzo*.
4. Para empezar vamos a ejecutar GridWorld en el modo de control manual, que utiliza las teclas de flecha.
   python3 gridworld.py -m -n 0
5. El objetivo es lograr llegar lo antes posible a la celda etiquetada con un 1, evitando caer en la celda con un -1.



Preguntas

1. ¿Cuántas celdas/estados aparecen en el tablero? ¿Cuántas acciones puede ejecutar el agente? Si quisieras resolver el juego mediante aprendizaje por refuerzo, ¿cómo lo harías?

2. Abrir el fichero *qlearningAgents.py* y buscar la clase *QLearningAgent*. Describir los métodos que aparecen en ella.

3. Ejecuta ahora el agente anterior con:

   python3 gridworld.py -a q -k 100 -n 0

4. ¿Qué información se muestra en el laberinto? ¿Qué aparece por terminal cuando se realizan los movimientos en el laberinto?

5. ¿Qué clase de movimiento realiza el agente anterior?

6. Dibujar el MDP considerando un entorno determinista.

7. ¿Hay varias políticas óptimas? Describe todas las políticas óptimas para este problema.

8. Escribir el método *update* de la clase *QLearningAgent* utilizando las funciones de actualización del algoritmo *Q-Learning*.

9. Establece en el constructor de la clase *QLearningAgent* el valor de la variable *epsilon* a 0,05. Ejecuta nuevamente con:

   python3 gridworld.py -a q -k 100 -n 0

   ¿Qué sucede?

10. Después de la ejecución anterior, abrir el fichero *qtable.txt*. ¿Qué contiene?

2. Ejercicio 2

Ahora vamos a crear un MDP estocástico:

1. Ejecuta y juega un par de partidas con el agente manual:

   python3 gridworld.py -m -n 0.3

   ¿Qué sucede? ¿Crees que el agente *QLearningAgent* será capaz de aprender en este nuevo escenario?

2. Reiniciar los valores de la tabla Q del fichero *qtable.txt*. Para ello ejecutar desde el terminal:

   cp qtable.ini.txt qtable.txt

3. Ejecutar el agente *QLearningAgent*:

   python3 gridworld.py -a q -k 100 -n 0.3

4. Tras unos cuantos episodios, ¿se genera la política óptima? Y si se genera, ¿se tarda más o menos que en el caso determinista?

3. 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 21 de abril de 2021**.

El nombre del archivo comprimido debe contener los últimos 6 dígitos del NIA de los dos alumnos, `ej. tutorial4-123456-234567.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 función *update* implementada.
   - Los ficheros generados con las distintas tablas Q.

2. El archivo de código fuente modificado por los alumnos **qlearningAgents.py**.

ANEXO: Parámetros útiles del código

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

```
python3 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.
- **-k episodios** Número de episodios de aprendizaje. Por defecto es 1.
- **-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.