

UNIVERSIDAD CARLOS III DE MADRID

APRENDIZAJE AUTOMÁTICO

GRADO EN INGENIERÍA INFORMÁTICA

GRUPO 83

Competición

Autores:

Daniel MEDINA GARCÍA
Alejandro RODRÍGUEZ SALAMANCA

10 de mayo de 2016

Índice

1. Agente de la competición	2
-----------------------------	---

1. Agente de la competición

El agente entregado en la competición se trata de una versión del agente desarrollado para la práctica 3.

El conjunto de estados elegido fue el formado por la dirección del fantasma más cercano y la dirección del último movimiento realizado por PacMan. Esto nos da un total de 64 posibles estados. Estos atributos fueron elegidos por dos razones:

- Simplifica mucho el conjunto de estados — Al tener en cuenta únicamente tres atributos, el conjunto de estados queda muy reducido, lo que era uno de los objetivos que tratábamos de cumplir, puesto que un conjunto de estados muy grande puede contener estados que no se lleguen a alcanzar nunca.
- Aporta información necesaria y genérica — Otro de los requisitos que hemos intentado cumplir ha sido obtener un agente genérico que no esté *atado* a ningún mapa, y que sea capaz de desenvolverse con soltura en diversos escenarios. Por ello, los atributos escogidos son independientes del mapa, pero aportan la información que necesita PacMan para determinar hacia qué dirección debe moverse.

En cuanto a los refuerzos, hemos decidido dividirlos en tres tipos:

- **PacMan se come un fantasma.** Cuando PacMan se come un fantasma, recibe un refuerzo positivo con un valor de 199, coincidiendo con la diferencia de puntuación entre el estado actual y el próximo estado al que se transiciona.
- **PacMan se acerca al fantasma más cercano.** En este caso, entendemos que PacMan está realizando un movimiento beneficioso, por el cual recibe un refuerzo positivo con valor 10.
- **PacMan se aleja del fantasma más cercano.** Si PacMan está *persiguiendo* a un fantasma y su distancia a éste aumenta, se penaliza el movimiento con un refuerzo negativo con valor -1.

Además, se ha usado la técnica ϵ -greedy, por la cual nuestro agente toma la acción devuelta por la política obtenida hasta el momento un ϵ % de las veces, y $(1-\epsilon)$ % una acción aleatoria, para que el mapa sea explorado y se prueben diversas acciones para un mismo estado, con el objetivo de quedarnos con la que se considere mejor.