Aprendizaje por Refuerzo en Super Mario

Iván Da Palma Ortega

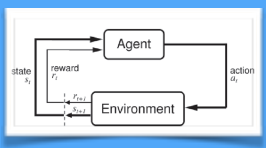
ETSI ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA

Objetivo

Se ha desarrollado el controlador de un agente con el objetivo de superar ciertos niveles en el juego Infinite Super Mario. Para ello el agente analiza los componentes del entorno y toma una decisión basada en ciertos parámetros. No se tendrán en cuenta ni las monedas recogidas ni las flores ni champiñones que convierten a Mario. Se valorará positivamente si Mario golpea desde arriba a un enemigo, y negativamente que el personaje sea alcanzado por algún enemigo o que quede atascado en algún momento.

Procedimiento

El comportamiento del agente ha sido creado a través del algoritmo **Q-Learning**. Éste mide las posibles acciones disponibles en cada estado y premia aquellas que nos acercan al objetivo. Para el entrenamiento se ha considerado una **política epsilon-greedy**, la cual con un cierto porcentaje P se escogía un movimiento al azar y, con 1-P, realizaba la acción que más recompensa disponía en la tabla Q. Al terminar el entrenamiento Mario debe escoger siempre esta última opción.



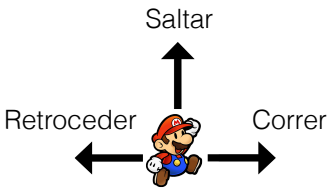
Para la actualización de la Tabla Q hemos tenido en cuenta los siguientes parámetros: **Velocidad de aprendizaje α** (cuánto podemos aprender de cada experiencia) y **Factor de descuento γ** (importancia entre refuerzos inmediatos o a largo plazo).

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$$

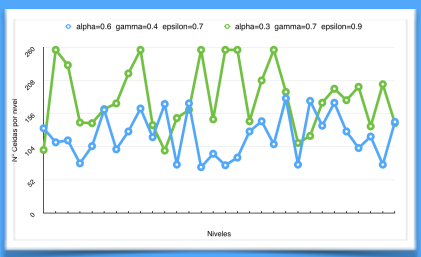
Estados

Corriendo
Saltando
Enemigo
Obstáculo
Atascado

Acciones



Resultado



Tras realizar dos entrenamientos, ejecutamos un total de treinta niveles aleatorios. Vemos que el controlador consigue llegar a una media del 70% del nivel y un 20% de niveles completados en el mejor de los casos. El controlador que explota lo que mejor sabe hacer aprende mejor que explorando otras posibles acciones, ya que al premiar dichas acciones, decidimos premiar más circunstancias negativas que positivas.

Conclusión

Tras entrenar a Mario en distintos niveles, el controlador realiza una media alta del recorrido del nivel, por lo que Mario ha aprendido a evitar enemigos y obstáculos. En determinadas circunstancias Mario queda atascado, por lo que quizá entrenando durante más tiempo el controlador o añadiendo distintos estados más pueda salir de ellos sin problemas.