DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico Número 2

15 de noviembre de 2016 — Aprendizaje Automático

Integrante	LU	Correo electrónico
Gasco, Emilio	171/12	gascoe@gmail.com
Gatti, Mathias	477/14	mathigatti@gmail.com

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria - (Pabellón I/Planta Baja)
Intendente Güiraldes 2160 - C1428EGA
Ciudad Autónoma de Buenos Aires - Rep. Argentina
Tel/Fax: (54 11) 4576-3359
http://www.fcen.uba.ar

Índice

0. Introducción

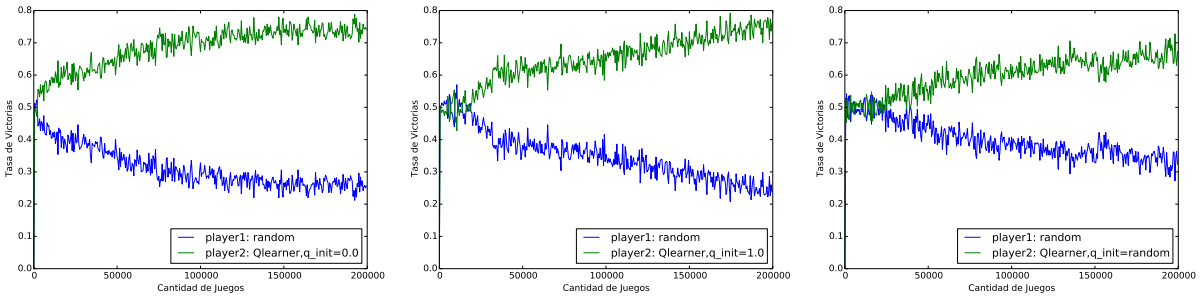
El objetivo de este trabajo fue experimentar con el algoritmo de aprendizaje por refuerzos Q-learning, variando el modelo del estado, refuerzos(recompensas) recibidos por los agentes y hiper-parámetros del algoritmo. . Para ello se construimos diferentes agentes para el juego "4 en linea".

1. Modelo estado agente

2. Resultados

2.1. Hiper-parámetros

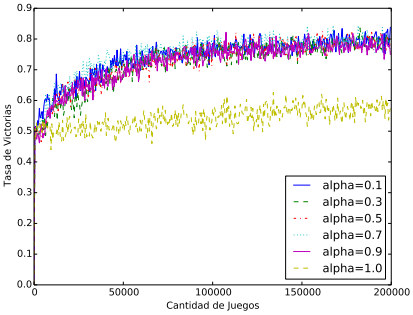
Se experimento con diferentes valores iniciales de Q. Los mejores resultados se obtuvieron utilizando 0 como valor inicial, logrando mayor velocidad en el entrenamiento. En la figura 1 se muestran tasa de victorias de agentes de QLearning contra agentes random para los diferentes valores iniciales de Q. Pasando los 200000 partidos de entrenamiento no se nota mejora los agentes que utilizan 0 como valor inicial. Y para diferentes valores de Q eventualmente convergen a la misma tasa de victorias del agente inicializado a 0.



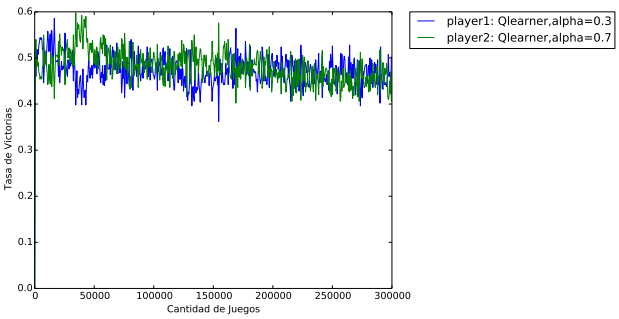
(a) q_init=0.0 (b) q_init=1.0 (c) q_init=random(0,1)

Figura 1: Velocidad de aprendizaje en función de valor inicial de Q

En la figura 2a se compara tasa de victorias para agentes entrenando con diferentes valores para la tasa de aprendizaje. Al utilizar el valor 1.0, que desprecia el conocimiento previo para dicho estado el agente logra una tasa de victorias similar un un agente random. Para todos los valores inferiores a 1.0 se logran una perfomance similar entrenando contra un agente random. La figura 2b muestra resultados de 2 agentes que implementan con Qlearning con diferentes tasa de aprendizaje, y en ningún momento uno de los agentes supera significativamente al otro.



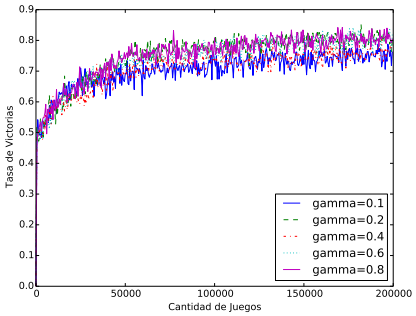
(a) Bots entrenando con diferentes α(tasa de aprendizaje) contra Bot random



(b) Bot α=0.3 vs Bot α=0.7

Figura 2

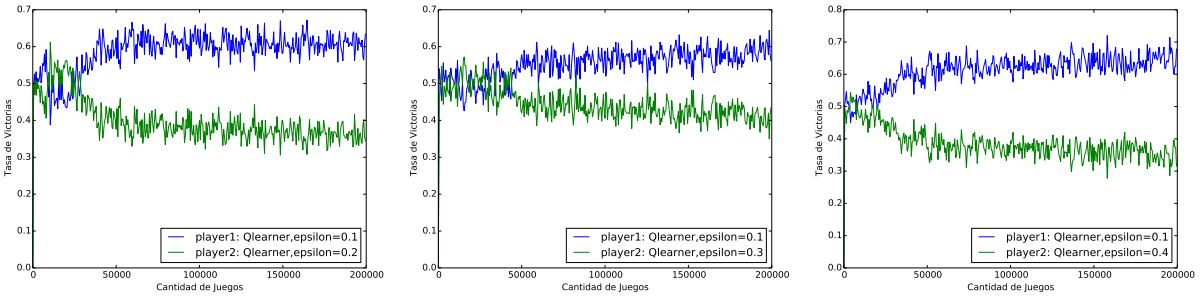
Por ultimo experimentamos variando el factor de descuento(γ). Este parámetro controla el peso que tienen los recompensas de estados posteriores. Los resultados obtenidos no fueron concluyentes. Antes de experimentar esperamos mejores resultados para valores de γ cercanos a 1. Pero como se puede observar en la figura 3, para valores de $\gamma = 0{,}2$ se obtienen resultados parecidos a valores cercanos a 1.



(a)

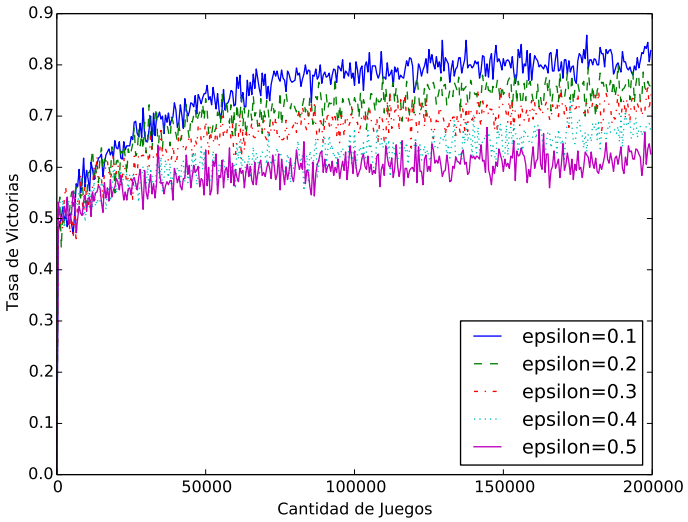
Figura 3: Bots entrenando con diferentes γ contra agente random

2.2. Exploración vs Explotación



(a) (b) (c)

Figura 4: Bots con diferentes valores de epsilon



(a)

Figura 5: Bots entrenando con diferentes epsilon contra Bot random

2.3. Estrategias de entrenamiento

También experimentamos con diferentes combinaciones de entrenamiento para intentar mejorar la performance del agente los agentes. Enfrentamos agentes entrenados contra un agente random contra agentes entrenados entre Agentes Qlearning y por ultimo se probo entrenar primero contra random y después contra Agente de Qlearning. No se observaron mejoras modificando el entrenamiento y en promedio los agentes sin importar como se entrasen ganaban mitad de partidas cada uno.

3. Discusión

A partir de los experimentos pudimos confirmar que la variación hiper-parámetros tiene un impacto en la velocidad de aprendizaje de los agentes. A excepción de algunos valores patológicos como puede ser una laza de aprendizaje igual a 1.0 la variación de los mismo tiene un impacto en la velocidad de aprendizaje pero no en la calidad del agente resultante. La calidad de juego del agente esta fuertemente relacionada con como se modela los estados Q. Realizamos varios intentos de mejorar la calidad del agente sin éxito, por ejemplo guardar un estado que tuviese en cuenta el color con el que jugaba el el jugador, pero no se logran mejoras apreciables.