Tarea 4 Inteligencia Artificial Preguntas 2.5 y 3

Baltazar Lutjens Oliva 20.161.518-6 N°1862569J

Pregunta 2.5:

a) Su papel es asignar el peso o la importancia a las recompensas futuras en comparación con las recompensas inmediatas. Esto se hace para equilibrar la "miopía" de buscar recompensas instantáneas y la "visión de futuro" de buscar recompensas a largo plazo.
   Si la tasa de descuento es 0, el agente sólo se preocupa por las recompensas inmediatas y no por las recompensas futuras. Dicho de otro modo, busca gratificación instantánea sin considerar posibles recompensas que podría obtener en el futuro tomado diferentes decisiones.
   En contraste, si la tasa de descuento es 1, el agente da igual importancia a las recompensas tanto inmediatas como futuras. Es decir, toma decisiones considerando tanto las recompensas que puede obtener ahora como las recompensas que podría acumular a largo plazo.
b) En mi caso la mejor tasa de descuento fue un valor alto, mas cercano al 1, porque debido a la naturaleza del juego y la ventaja que da poder prevenir la estrategia del contrincante y hacer una jugada acorde, da una ventaja al respecto de usar una tasa mas baja.
c) El learning rate se utiliza para actualizar los valores de Q, que representan la calidad de una acción en un estado dado. Después de que el agente toma una acción y observa el resultado en el nuevo estado, actualiza el valor Q correspondiente al estado y la acción realizada. El learning rate determina la proporción en la que los nuevos conocimientos se incorporan al valor Q existente.
   Si el learning rate es alto, se dará más importancia a la nueva información, lo que significa que los valores de Q se actualizarán más rápidamente. Esto puede llevar a una convergencia más rápida hacia una solución óptima, pero también puede hacer que el modelo sea más sensible a ruido o fluctuaciones en los datos. Por otro lado, si el learning rate es bajo, los valores de Q cambiarán más lentamente, lo que puede ser más estable, pero requerirá más tiempo para converger. En mi caso, utilize un learning rate bajo, por lo que estuve mucho tiempo(alrededor de 4 horas) entrenando el modelo, lo que me dio mejor resultado en comparación a los otros modelos que fueron entrenados.
d) El reward fue básicamente el mismo para los dos agentes pero inverso, se premiaba o se penalizaba por la distancia que tenían con el contrincante, si disminuía premiaba al gato y si aumentaba premiaba al ratón y viceversa, además se implemento un reward con un valor 10 veces mayor, que era positivo si ganaba el gato o negativo si perdía el ratón. Esto lo decidí, porque fue la solución mas rápida y que a mi juicio parecía la mas efectiva. Otro posible reward, podría haber sido no quedar atrapado, junto con el reward de no perder, puesto que viendo el comportamiento, siempre que el ratón perdía, era porque entraba a los pasillos sin salida, esto en comparación a la inicial, quizá podría haber funcionado mejor, pero su dificultad en la implementación es suficiente como para descartala.

Pregunta 3

a) En mi caso el modelo mas eficiente fue la red neuronal, siendo muchísimo mas eficiente con un tiempo estimado de 2 minutos en entrenamiento, en cambio para entrenar el modelo del ratón con Q-learning, me demore aproximadamente 4 horas, Esto se puede deber a una variedad de factores, por ejemplo la velocidad de la tarjeta grafica de mi computador, que pudo haber ayudado a la eficiencia de la red neuronal, pero sobre todo el delta de tiempo se debió al mal rendimiento inicial del modelo de Q-learning, al no tener un Q-map inicial y tener que crear una desde 0, significa que al inicializar todo, el rendimiento sea horrendo, teniendo un rendimiento tal que demoraba 5 minutos en completar 100 juegos, esto fue mejorando con el tiempo, cuando la habilidad del modelo mejoraba y no era necesario hacer cambios en el Q-map tan seguido, por lo que una vez mejoro el modelo siguió mejorando casi exponencialmente, pero en comparación de tiempo rendimiento Q-learning fue incapaz de acercarse a la red neuronal.

b) En mi caso para el ratón, Q-learning tuvo mejor desempeño que la red neuronal contra el ratón de entrenamiento, pero la diferencia no fue significativa. El testeo fue hecho con una ronda de 1000 partidas para cada modelo contra el modelo base y promediando los Mean Steps que tomaba cada uno en realizar cada ciclo de 100 juegos, obteniendo un promedio de 102 para Q-learn y un promedio de 110 aprox para la red neuronal. Para el ratón se ocupó una estrategia similar, solamente que mientras mayor sea la cantidad de pasos se rankea mejor el modelo. En este test, se obtuvieron unos resultados de 90 pasos promedio para la red neuronal y un promedio de 118 para Q-learning, siendo el primero muchísimo mejor que este, esto quizá se deba a la fatla de entrenamiento del modelo de Q-learning, que a medida que iba mejorando el tiempo de cada ciclo de entrenamiento iba aumentando, porque aumentaba la duración del juego.

c) Si el mapa cambiara, habría una gran disparidad entre los modelos, debido a que la red neuronal hizo una gran cantidad de overfiting al tablero, calculando la jugada optima para cada estado de tablero, por lo que ni siquiera tomaría en cuenta nuevos obstáculos, en cambio Q-learn es mas adaptable, debido a que va cambiando retroactivamente su tabla Q-map, lo que haría que se adaptara más fácilmente a nuevos mapas, además que para actualizar su Q-map se necesita una buena comprensión de lo que se quiere lograr a nivel de programador, por lo que es como una ganzúa que quizá se demora más en desbloquear todas las cerraduras, que una llave que abre una cerradura especifico muy rápidamente.

d) En mi opinión, para el gato con Q-learn, se podían observar ocasionalmente pseudo fintas cuando perseguía al ratón, a veces yendo hacia un lado y saliendo para el otro, pero la mayoría de las veces solamente salía y se intentaba acercar lo más posible al ratón, en cambio la red neuronal esperaba más a que el ratón se pusiera en una peor posición por así decirlo y luego “atacaba” mas. Para el ratón la verdad es que los dos modelos funcionaban bastante bien, hasta que decidían esconderse en los pasillos y perdían, por lo que encontré bastante parecido su comportamiento.