Universidad Del Valle de Guatemala
Departamento de Computación
Inteligencia Artificial
Grupo 6

Diana Lucía Fernández Villatoro - 21747
Jennifer Michelle Toxcón Ordoñez - 21276
Daniel Esteban Morales Urizar - 21785
Brandon Rolando Sicay Cumes - 21757

Laboratorio 7

Task 1 - Teoría

1. **¿Qué es el temporal difference learning y en qué se diferencia de los métodos tradicionales de aprendizaje supervisado? Explique el concepto de "error de diferencia temporal" y su papel en los algoritmos de aprendizaje por refuerzo**

   El temporal difference learning es un enfoque de aprendizaje automático que se basa en aprender a partir de la experiencia que resulta de la secuencia de interacciones que un agente tiene con su entorno a lo largo del tiempo. El TD learning se diferencia de métodos tradicionales de aprendizaje supervisado pues este no requiere un conjunto de datos etiquetados para entrenar el modelo, sino que aprende directamente de la experiencia y la retroalimentación que encuentra a base de recompensas y penalizaciones.

   Por su parte, el error de diferencia temporal es una medida que se utiliza para cuantificar la diferencia entre las predicciones actuales de un agente sobre una recompensa futura y la recompensa real que se obtuvo después de tomar la acción específica. El cálculo de este error depende del método específico de aprendizaje que se esté usando, sin embargo la forma general de calcularlo se expresa como:

   $$TD\ error = R_t + \gamma V(S_{t+1}) - V(S_t)$$

   Donde:

   - R_t es la recompensa real observada después de tomar la acción
   - $V(S_t)$ es la estimación actual del estado S_t
   - $V(S_{t+1})$ es la estimación del valor del siguiente estado
   - γ es el factor de descuento

   Este error juega un rol importante en algoritmos de aprendizaje por refuerzo pues proporciona la señal de retroalimentación que se necesita para ajustar y mejorar las estimaciones del valor de los estados o acciones, lo cual ayuda a que el agente vaya aprendiendo y mejorando en su proceso de aprendizaje.

2. **En el contexto de los juegos simultáneos, ¿cómo toman decisiones los jugadores sin conocer las acciones de sus oponentes? Dé un ejemplo de un escenario del mundo real que puede modelarse como un juego simultáneo y discuta las estrategias que los jugadores podrían emplear en tal situación**

   En este tipo de juegos existen diferentes estrategias para emplear decisiones, una de ellas es evaluar y considerar las posibles acciones a realizar, tanto para uno como jugador como para el oponente, teniendo en cuenta las posibles consecuencias de cada una de ellas. A su vez se pueden hacer uso de probabilidades, donde se le asigna una probabilidad distinta a las posibles acciones que podría tomar el oponente, con las que luego se calcula cual es la mejor respuesta considerándolas.

   El uso de estrategias dominantes es un enfoque en que los jugadores identifican estrategias donde se maximizan las ganancias de su propio juego sin tomar en cuenta las acciones del oponente. Del mismo modo, se puede tratar de predecir las acciones próximas del oponente, a pesar de no conocerlas con exactitud, esta puede ser una opción para la toma de decisiones si se tiene un conocimiento medio o avanzado en el juego, en donde se de analizar las acciones pasadas y tratar de adivinar la próxima acción apoyándose del uso de señales y pistas.

   El juego de azar póker es un ejemplo de la vida real que se modela como un juego simultáneo. Las posibles estratégias a usar en un juego como el póker pueden llegar a ser muchas, sin embargo, entre ellas se encuentra el tratar de predecir las acciones de los oponentes a base de las acciones pasadas que se han tomado, también es posible combinar esta estratégia con la asignación de probabilidades a las posibles acciones que el oponente va a realizar.

   Sin embargo, una de las mejores estrategias que se puede emplear para un juego como este es la adaptación y el ajuste continuo a medida que se desarrolla el juego y se van revelando acciones tomadas por los jugadores, adaptando las estratégias que se tomaron a base del conocimiento del juego en función de la nueva información que se va desbloqueando.

3. **¿Qué distingue los juegos de suma cero de los juegos de no suma cero y cómo afecta esta diferencia al proceso de toma de decisiones de los jugadores? Proporcione al menos un ejemplo de juegos que entren en la categoría de suma cero y discuta las consideraciones estratégicas únicas involucradas**

   Los juegos de suma cero se identifican por ser juegos de ganancia o pérdida total entre jugadores, es decir, que un jugador pierde exactamente la misma cantidad de lo que el otro jugador gana. Por otra parte, en el juego de no suma cero un jugador puede ganar sin que el otro jugador pierda la misma cantidad necesariamente, en ellos se pueden tener estrategias en donde todos los jugadores pueden verse beneficiados o perjudicados a la vez. Esta diferencia entre el tipo de juegos afecta al proceso de toma de decisiones pues en los juegos de suma cero cada jugador trata de maximizar su propia ganancia, por lo que su estrategia para la toma de decisiones suele ser agresiva y orientada a explotar las debilidades de los oponentes.

   Por otra parte, en los juegos de no suma cero las estrategias de toma de decisiones se basan en la negociación, colaboración y la búsqueda de soluciones que

beneficien a todas las partes involucradas, por lo que los jugadores pueden tener incentivos para cooperar y trabajar en conjunto hacia un objetivo común.

Un ejemplo del juego de suma cero es el de Damas, donde dos jugadores compiten entre sí mientras capturan las fichas del oponente mientras evita que capturen sus propias fichas. Entre las estrategias está el control del centro del tablero, donde las fichas ubicadas aquí tienen una mayor posibilidad de movimientos. A su vez, se tiene el intercalado entre ataques y defensas activas, donde se alterna entre mover fichas para amenazar fichas enemigas y proteger las fichas propias. El anticiparse a los movimientos contrarios es una estrategia bastante buena, donde se trata de prepararse para posibles ataques enemigos y poder responder de una manera efectiva, esto mediante la evaluación de las posibles jugadas del oponente y anticipar las consecuencias de dichos movimientos.

4. **¿Cómo se aplica el concepto de equilibrio de Nash a los juegos simultáneos? Explicar cómo el equilibrio de Nash representa una solución estable en la que ningún jugador tiene un incentivo para desviarse unilateralmente de la estrategia elegida**

El equilibrio de Nash es una herramienta que se usa para analizar el comportamiento estratégico de los jugadores en juegos simultáneos en donde ningún jugador puede mejorar su situación cambiando su estrategia, mientras se mantienen constantes las estrategias de los demás jugadores. Este equilibrio representa una solución estable pues ningún jugador tiene el incentivo de desviarse, pues si se tiene el caso en que un jugador decide cambiar la estrategia y adoptar una nueva, esta no será beneficiosa mientras los demás jugadores mantengan su estrategia actual y estén en el equilibrio de Nash. Lo que hace que el equilibrio de Nash sea la opción preferida para un jugador en la ausencia de cooperación o coordinación con los otros jugadores es que el desviarse no generará ninguna ganancia adicional a la conseguida con dicha estrategia.

5. **Discuta la aplicación del temporal difference learning en el modelado y optimización de procesos de toma de decisiones en entornos dinámicos. ¿Cómo maneja el temporal difference learning el equilibrio entre exploración y explotación y cuáles son algunos de los desafíos asociados con su implementación en la práctica?**

En el modelado y optimización de procesos de toma de decisiones en entornos dinámicos, el temporal difference learning es una técnica muy efectiva que se usa dentro de entornos donde la retroalimentación que se puede obtener es escasa, no está disponible de inmediato o es un conjunto estocástico. El equilibrio entre explotación y exploración es una característica bastante sensible que se debe tener en cuenta dentro de un modelado, y dentro del TD learning esto se logra mediante métodos como lo es el algoritmo ε-greedy, donde el agente procede a seleccionar la mejor opción conocida con probabilidad $1-\varepsilon$ y elige una acción aleatoria con probabilidad ε, lo cual permite que el agente explorer el espacio de acciones mientras se sigue explotando las acciones que se consideran como efectivas.

Sin embargo, en la práctica algunos de los desafíos están en la selección de hiperparámetros críticos, como lo es la tasa de aprendizaje y el parámetro de explotación, dentro del algoritmo ε-greedy, lo cual puede resultar como un desafío y puede requerir ajustes iterativos. A su vez, la convergencia y estabilidad del algoritmo puede ser difícil de alcanzar, especialmente en entornos no lineales o con una complejidad alta, para los cuales se pueden necesitar de técnicas auxiliares adicionales.

Dentro de dichos entornos con complejidad y dimensionalidad alta se puede tener el desafío de lograr un aprendizaje oportuno para el modelo o la dificultad de generalizar de una buena manera en estados similares, en este caso se necesita realizar la reducción de la dimensionalidad o el uso de arquitecturas más robustas y complejas para lograr abordar de buena manera el desafío. Del mismo modo, en entornos de gran escala, una exploración exitosa puede llegar a ser muy costosa temporal y computacionalmente.

Task 2 - Connect Four

Imagen 1 - Conteo de partidas ganadas por algoritmo Q-learning Vs. Minimax

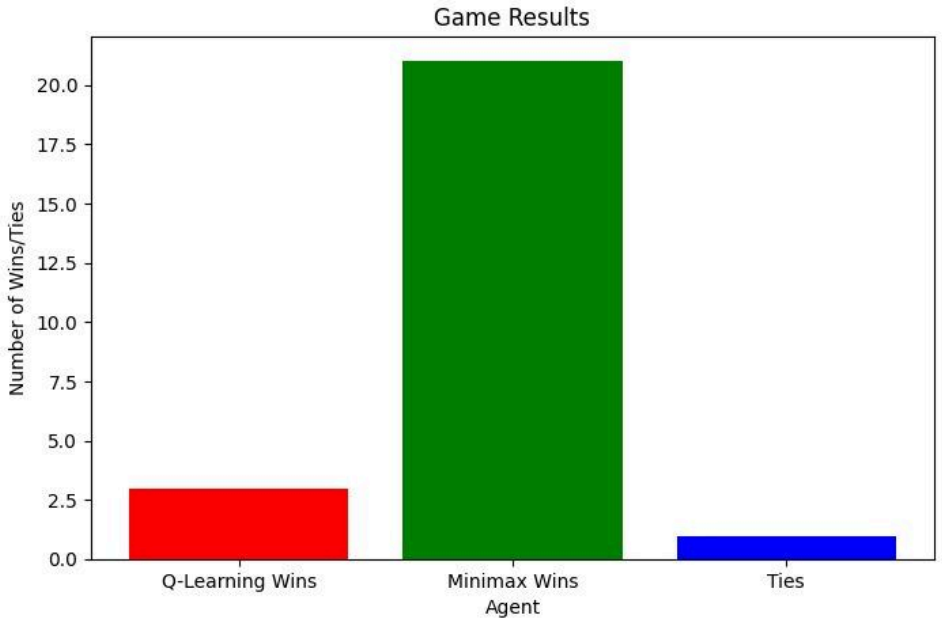

Imagen 2 - Conteo de partidas ganadas por algoritmo Q-learning Vs. Minimax con poda Alpha-Beta