Universidad del Valle de Guatemala
Inteligencia artificial, sección 10
Integrantes:
- Gabriel García - 21352
- Luis Pedro Montenegro - 21699
- Javier Prado - 21486
- Emilio Solano - 21212

Laboratorio 7

Task 1 - Teoría

Responda las siguientes preguntas de forma clara y concisa, pueden subir un PDF o bien dentro del mismo Jupyter Notebook.

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**

   TD representa una estrategia no supervisada diseñada para anticipar el valor esperado de una variable en una secuencia de estados. TD emplea un ingenioso truco matemático que simplifica los complejos razonamientos sobre el futuro, sustituyéndolos por un proceso de aprendizaje directo y efectivo que arroja resultados similares. En lugar de hacer cálculos complicados para averiguar todas las recompensas futuras, el aprendizaje TD hace algo más sencillo: trata de adivinar cuánta recompensa obtendremos ahora mismo y qué recompensa esperamos recibir después. Esto lo hace muy útil para predecir lo que sucederá a continuación en una serie de eventos. El aprendizaje TD se diferencia de los métodos tradicionales de aprendizaje supervisado en que no requiere tener ejemplos de entrada y salida para aprender. En cambio, se centra en predecir valores futuros basados en valores actuales, sin necesidad de un conjunto de datos de entrenamiento completo con respuestas correctas. El **Error de TD** es la diferencia entre la recompensa final correcta (V*_t) y nuestra predicción actual (V_t). El papel del error de diferencia temporal es fundamental en los algoritmos de aprendizaje por refuerzo porque guía el proceso de actualización de las estimaciones del agente sobre cuánto valor esperado existe en ciertas acciones o estados.
   (Doung, 2021)

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

   Existe una variedad de juegos simultáneos, por ejemplo el **juego de Morra,** donde ambos jugadores realizan una acción en simultáneo para ver el número de dedos que el oponente dio. En el contexto de los juegos simultáneos las decisiones se hacen, como su nombre lo dice, al mismo tiempo. En este caso las estrategias a tomar.

3. **¿Qué distingue los juegos de suma cero de los juegos de 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 juegos de no suma cero y discuta las consideraciones estratégicas únicas involucradas**

   Los juegos de suma cero son aquellos en los que la ganancia total de un jugador es exactamente igual a la pérdida total de otro jugador, de manera que la suma de estas ganancias y pérdidas siempre es cero. Por otro lado, en los juegos de suma no cero, los intereses de los jugadores no son necesariamente opuestos, lo que hace que estos juegos sean más complejos de resolver. En estos juegos, los jugadores pueden adoptar una posición cooperativa, lo que significa que colaboran entre sí en lugar de competir directamente. En este contexto, las decisiones pueden afectar a todos los participantes de manera conjunta, lo que puede resultar en que todos ganen o todos pierden, dependiendo de la estrategia adoptada y de la cooperación entre los jugadores.
   (Mattos, 2023)

   Un ejemplo de juego de no suma cero es el "Dilema del Prisionero Iterado", donde dos personas son arrestadas por cometer un delito y se les da la oportunidad de confesar o permanecer en silencio. En esta versión iterada del juego, que consiste en varias rondas, las estrategias como "Ojo por ojo", donde un jugador comienza cooperando y luego refleja el movimiento anterior del oponente en rondas posteriores, y "Cooperar siempre", donde se elige consistentemente la cooperación independientemente de las

acciones del oponente, son comunes. En este juego, las consideraciones estratégicas incluyen la reciprocidad, la toma de decisiones a largo plazo y las estrategias condicionales, lo que lo hace más complejo que un juego de suma cero.
(*Cooperación la Clave del Éxito En el Dilema del Prisionero Repetido - FasterCapital*, s. f.)

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 representa una solución estable en la que ningún jugador tiene un incentivo para desviarse unilateralmente de la estrategia elegida. Esto significa que, una vez alcanzado el equilibrio, ningún jugador puede mejorar su situación cambiando su estrategia sin que los otros jugadores también cambien la suya y, en consecuencia, sin que se produzca un resultado peor para él mismo.

   Un ejemplo claro de esto se puede observar en el juego del dilema del prisionero. En este juego, dos prisioneros enfrentan la decisión de cooperar o traicionar a su compañero de crimen. Si ambos cooperan, reciben una sentencia moderada. Si ambos traicionan, reciben una sentencia más larga. Sin embargo, si uno coopera y el otro traiciona, el traidor recibe una sentencia muy corta mientras que el cooperador recibe una sentencia muy larga.

   En este escenario, el equilibrio de Nash se alcanza cuando ambos prisioneros deciden traicionar, ya que si uno de ellos decide cooperar mientras el otro traiciona, el que coopera recibe una sentencia muy larga mientras que el traidor recibe una sentencia corta. Por lo tanto, ningún prisionero tiene un incentivo para cambiar unilateralmente su decisión, ya que cualquier cambio resultaría en un resultado peor para él mismo si el otro prisionero sigue traicionando.
   (Martinez, 2018)

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?**

   El TD se utiliza para modelar y optimizar procesos de toma de decisiones en entornos dinámicos debido a su capacidad para aprender de la experiencia en tiempo real y ajustar las decisiones en consecuencia. El TD maneja el equilibrio entre exploración y explotación mediante la adopción de estrategias que permiten a los agentes explorar nuevas acciones o estados mientras explotan el conocimiento existente para maximizar las recompensas a largo plazo. Al utilizar esta estrategia, el TD garantiza que los agentes exploran continuamente el espacio de acciones y estados, lo que les permite descubrir nuevas estrategias óptimas mientras evitan quedarse atrapados en lugares óptimos. Uno de los desafíos principales es la selección adecuada de los parámetros del algoritmo, como la tasa de aprendizaje y el factor de descuento. Una elección incorrecta de estos parámetros puede afectar significativamente el rendimiento del TD y su capacidad para converger a soluciones óptimas.
   (*Temporal Difference Learning (TD Learning) | Engati*, s. f.)

Referencias:

- ***Temporal difference learning (TD Learning) | Engati.* (s. f.). Engati. https://www.engati.com/glossary/temporal-difference-learning** Martinez, A. (2018, 6 marzo). Teoría de juegos II: Equilibrio de Nash - Policonomics. *Policonomics - Economics made simple*. https://policonomics.com/es/lp-teoria-juegos2-equilibrio-nash/
- *Cooperacion la clave del exito en el dilema del prisionero repetido - FasterCapital*. (s. f.). FasterCapital. https://fastercapital.com/es/contenido/Cooperacion--la-clave-del-exito-en-el-dilema-del-prisionero-repetido.html#Introducci-n-al-dilema-de-los-prisioneros-iterados
- Mattos, A. A. (2023, 30 noviembre). ¿Qué es un juego de suma cero y no cero? *Rankia*. https://www.rankia.co/blog/analisis-colcap/4595268-que-juego-suma-cero-no
- Duong, V. H. T. (2021, 31 diciembre). Intro to reinforcement learning: temporal difference learning, SARSA vs. Q-learning. *Medium*. https://towardsdatascience.com/intro-to-reinforcement-learning-temporal-difference-learning-sarsa-vs-q-learning-8b4184bb4978