

Inteligencia Artificial - Hoja de Trabajo 2 -

Semestre I - 2024 **Instrucciones:**

- Esta es una actividad en grupos de no más de 3 integrantes.
 - Este grupo aún no existe en Canvas por lo que deberán unirse a uno con el nombre de **Grupo C [Hasta 4 Integrantes]** -
 - Si alguien necesita cambiar de grupo por favor notificar al profesor antes del viernes de laboratorio
- No se permitirá ni se aceptará cualquier indicio de copia. De presentarse, se procederá según el reglamento correspondiente.
- Tendrán hasta el día indicado en Canvas.
 - No se confíen, aprovechen el tiempo en clase para entender todos los ejercicios y avanzar lo más posible.

Task 1 - Preguntas Teóricas

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible.

1. Defina el proceso de decisión de Markov (MDP) y explique sus componentes.

El proceso de decisión de Markov es un framework empleado en problemas de toma de decisiones donde los resultados son parcialmente aleatorios y parcialmente bajo el régimen de un agente de toma de decisiones. (Lark Editorial Team, 2023).

Sus componentes son los siguientes:

- **States:** Estados distintos que representan diferentes situaciones o condiciones en las que el agente se puede hallar envuelto.
- **Actions:** Denota las opciones disponibles que el agente puede tomar durante las transiciones.
- **Transaction Probability Function:** Otorga la probabilidad de transicionar de un estado a otro al tomar una opción.
- **Reward Function:** Evalúa los beneficios inmediatos o costos asociados con tomar una acción en un estado específico,

(Lark Editorial Team, 2023)

2. Describa cual es la diferencia entre política, evaluación de políticas, mejora de políticas e iteración de políticas en el contexto de los PDM.

Una **política** es aquella que determina la acción óptima a tomar dado el estado actual para que tome la mejor recompensa(reward). La **evaluación de política** consiste en verificar qué tan buena es la recompensa que va obteniendo en cada estado. Ahora bien, la **mejora de políticas** es el proceso en el cual la política es optimizada.

(Kanade, 2022)

3. Explique el concepto de factor de descuento (gamma) en los MDP. ¿Cómo influye en la toma de decisiones?

El **factor de descuento** es una variable para definir las recompensas a largo o corto plazo. Si gamma va a 0 las recompensas inmediatas son priorizadas; si tiende a 1 se enfoca en recompensas a largo plazo. Esta se emplea en la función de valor, la cual retorna la recompensa por cada estado. Gamma es la variable que pone énfasis si es recompensa inmediata o a largo plazo dependiendo de la política.

(Kanade, 2022)

4. Analice la diferencia entre los algoritmos de iteración de valores y de iteración de políticas para resolver MDP.

Iteración de valores: Algoritmo que va mejorando el estimado de manera iterada el valor de cada estado. Lo hace tomando en consideración las recompensas inmediatas y a futuro al tomar distintas acciones. Eventualmente la secuencia converge dando como tal un estado óptimo de política de estado.

Iteración de políticas: Algoritmo que toma una política aleatoria. Luego similar al algoritmo de arriba se evalúa iteradamente hasta que esta converja en la más óptima.

Viendo esto, podemos ver la diferencia entre ambos donde uno elige una política a optimizar y la otra va por la optimización de los estados para obtener convergencia.

(Tokuç, 2023)

5. ¿Cuáles son algunos desafíos o limitaciones comunes asociados con la resolución de MDP a gran escala? Discuta los enfoques potenciales para abordar estos desafíos.

Uno de los desafíos más grandes que presenta el Markov Decision Process es el **tiempo computacional y memoria**. Conforme vaya incrementando la complejidad del problema, así va aumentando la cantidad de memoria utilizada y de capacidad de computación.

Otro problema que también se afronta es que el MDP **asume completa observación**. Quiere decir que asume cierta perfección en cómo se presenta el problema y esto no siempre va a ser fiel a la realidad, lo que puede llevar a inexactitudes.

(Lark Editorial Team, 2023)

Task 2 - Preguntas Analíticas

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible.

1. Analice críticamente los supuestos subyacentes a la propiedad de Markov en los Procesos de Decisión de Markov (MDP). Analice escenarios en los que estos supuestos puedan no ser válidos y sus implicaciones para la toma de decisiones.

La propiedad de Markov nos indica que el futuro es independiente al pasado y se puede determinar solo con el presente que encapsula toda la información necesaria del pasado. Lo cual significa que S_t que significa el estado actual, contiene la información necesaria del pasado, por lo cual no serviría tener mayor información de los estados pasados.

$$\mathbb{P}[S_{t+1} | S_t] = \mathbb{P}[S_{t+1} | S_1, \dots, S_t]$$

(Silva, 2021)

Los supuestos subyacentes de la propiedad de Markov son los siguientes:

Independencia temporal a largo plazo: Este supuesto asume que el futuro es independiente del pasado, dado el presente. Sin embargo, en situaciones donde la dependencia temporal a largo plazo es crucial, como en el procesamiento del lenguaje natural o la predicción del comportamiento humano, este supuesto puede no ser válido.

Estacionariedad de probabilidades de transición y recompensas: La propiedad de Markov asume que las probabilidades de transición y las recompensas son conocidas y estacionarias. Sin embargo, en entornos inciertos o dinámicos, donde estas probabilidades pueden ser desconocidas o pueden cambiar con el tiempo, este supuesto puede no ser válido.

Discretización de espacios de estados y acciones: En entornos con espacios de estados y acciones continuos o grandes, el supuesto de discretización necesario para aplicar métodos basados en MDP puede ser poco práctico y puede introducir errores significativos.

2. Explore los desafíos de modelar la incertidumbre en los procesos de decisión de Markov (MDP) y analice estrategias para una toma de decisiones sólida en entornos inciertos.

La incertidumbre se modela mediante la probabilidad de transición entre estados y la función de

recompensa asociada a cada acción y estado. La incertidumbre es un aspecto clave en los procesos de decisión de Markov ya que capturan situaciones en las que las decisiones deben tomarse en un entorno estocástico. Existen desafíos a la hora de modelar la incertidumbre:

Algunos desafíos se pueden dar cuando la **información está incompleta** generando dificultad en la construcción de un modelo preciso. También la **complejidad del modelo** depende en cuantas incertidumbre desea capturar en el modelo, dado que aumenta su complejidad haciendo que el modelado sea más difícil y que los algoritmos de resolución sean más costosos computacionalmente. En la evaluación de decisiones bajo incertidumbre, se pueden aplicar tres criterios principales: el **Criterio MAXIMIN**, que busca maximizar el peor resultado posible y se asocia con una perspectiva pesimista; el **Criterio MAXIMAX**, que busca maximizar el mejor resultado posible y refleja una actitud optimista; y el **Criterio de la frustración mínima**, que busca minimizar la diferencia entre el resultado obtenido y el mejor resultado posible en cada situación, representando una perspectiva más equilibrada hacia el riesgo y la incertidumbre.

(Hernández-Lerma, s. f.)

Task 3 - Preguntas Prácticas

Desarrolle un agente básico capaz de resolver un problema simplificado del Proceso de Decisión de Markov (MDP). Considere utilizar un ejemplo bien conocido como el entorno 'Frozen Lake'. Proporcione el código Python para el proceso de toma de decisiones del agente basándose únicamente en los principios de los procesos de decisión de Markov. Recuerde que para este tipo de problema, el ambiente es una matriz de 4x4, y las acciones, pueden ser moverse hacia arriba, abajo, derecha, izquierda. Considere que el punto inicial siempre estará en la esquina opuesta del punto de meta. Es decir, puede tener hasta 4 configuraciones diferentes. Por ejemplo, el punto inicial puede estar en la coordenada (0, 0) y el punto de meta en la coordenada (3, 3). Además, la posición de los hoyos debe ser determinada aleatoriamente y no debe superar el ser más de 3. Es decir, si aleatoriamente se decide que sean 2 posiciones de hoyo, las coordenadas de estas deben ser determinadas de forma aleatoria. Asegúrese de usar "seed" para que sus resultados sean consistentes.

La manera de cómo implementar el MDP en el entorno FrozenLake será:

1. Definir las acciones que se pueden hacer. Lo de moverse arriba, abajo, izquierda, derecha.
2. Definir los estados. Si es matriz de 4x4 implica que hay 16 estados, 1 es de inicio y 1 es final.
3. Definir la función de recompensa, donde digamos caer en un hoyo sea visto como negativo y moverse a un punto que no lo sea sea como positivo.
4. Definir las transiciones. En caso sea determinístico o estocástico.
5. Estructurar las funciones para ir realizando los movimientos y cómo van a tomar la recompensa dada por cada movimiento
6. Definir las políticas, si vienen de algoritmos como Policy Iteration o Value Iteration. También implementar el algoritmo de aprendizaje.

Entregas en Canvas

1. Documento PDF con las respuestas a cada task

CC3045 - Inteligencia Artificial

Inteligencia Artificial

- Hoja de Trabajo 2 -

2. Jupyter Notebook o script resolviendo el task 3

- [por Refuerzo: Procesos de Decisión de Markov — Parte 1.](#)
- Hernández-Lerma, O. (s. f.). *Toma de decisiones de agentes racionales con procesos markovianos. Avances recientes en economía y finanzas.* [Toma de decisiones de agentes racionales con procesos markovianos. Avances recientes en economía y finanzas.](#)

Evaluación

1. [0.5 pt] Task 1
 - a. [0.1 pts] Cada pregunta
 2. [1 pts.] Task 2
 - a. [0.5 pts] Cada pregunta
 3. [1 pts.] Task 3
- Total 2.5pts

Semestre I - 2024

Referencias:

- Lark Editorial Team (2023). *Markov Decision Process* from [Markov Decision Process?](#)
- Kanade, Vijay (December 20, 2022). *What is the Markov Decision Process? Definition, Working, Examples* from [Markov Decision Process Definition, Working, and Examples - Spiceworks](#)
- Tokuç, A. Aylin (March 24, 2023). *Value Iteration vs. Policy iteration in Reinforcement Learning* from [Value Iteration vs. Policy Iteration in Reinforcement Learning | Baeldung on Computer Science](#)
- Silva, M. (2021, 10 diciembre). Aprendizaje por Refuerzo: Procesos de Decisión de Markov — Parte 1. *Medium*. [Aprendizaje](#)