Universidad del Valle de Guatemala
CC3085 – Inteligencia Artificial
Sección 10
3 de Marzo de 2025

Nombres: Daniela Navas **Carné:** 211000

HOJA DE TRABAJO 2

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.

Un Proceso de Decisión de Markov es un modelo matemático utilizado para representar problemas de toma de decisiones secuenciales en entornos dinámicos, da resultados probabilísticos y no determinísticos, a diferencia de modelos vistos previamente en clase. Los componentes clave son:

- **Conjunto finito de estados (S):** Representa todas las posibles situaciones en las que puede encontrarse el sistema.
- **Conjunto finito de acciones (A):** Son las decisiones que puede tomar el agente en cada estado.
- **Matriz de transición de estados (P):** Define la probabilidad de pasar de un estado a otro dado una acción, denotada como P(s'Is,a).
- **Función de recompensa (R):** Asigna una recompensa a cada transición de estado y acción, denotada como R(s,a,s').
- **Factor de descuento (γ):** Un valor entre 0 y 1 que determina cómo se valoran las recompensas futuras respecto a las inmediatas

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

- **Política (π)**: Estrategia que define la acción que debe tomar el agente en cada estado.
- **Evaluación de políticas**: Proceso de determinar el valor de una política dada, es decir, calcular el valor esperado de las recompensas futuras siguiendo esa política.
- **Mejora de políticas**: Proceso de mejorar una política basada en la evaluación de políticas, seleccionando acciones que maximicen el valor esperado.
- **Iteración de políticas**: Método que alterna entre la evaluación de políticas y la mejora de políticas hasta que se encuentra una política óptima.

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

El factor de descuento γ es un valor entre 0 y 1 que determina la importancia de las recompensas futuras. Un valor de γ cercano a 0 hace que el agente prefiera recompensas inmediatas, mientras que un valor cercano a 1 hace que el agente valore más las recompensas futuras. Influye en la toma de decisiones al ajustar el peso de las recompensas futuras en el cálculo del valor esperado de una política. Como ya se mencionó, Markov valora más maximizar recompensas que minimizar costo.

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

- **Iteración de valores:** Un algoritmo que busca encontrar el valor óptimo de cada estado sin necesidad de una política explícita. Actualiza los valores de los estados iterativamente hasta converger. Se basa en la ecuación de Bellman para actualizar iterativamente la función de valor V(s). Utiliza la mejora de políticas implícitamente seleccionando la acción con el mayor valor esperado en cada iteración. Es más estable, pero puede ser más lento porque actualiza valores en cada iteración antes de cambiar la política.
- **Iteración de políticas:** Combina la evaluación de una política con su mejora iterativa. Primero evalúa la política actual y luego mejora la política basándose en esa evaluación. Alterna entre evaluación de políticas y mejora de políticas.Puede converger más rápido que la iteración de valores en algunos casos, ya que cambia la política explícitamente después de cada mejora.

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.

- **Dimensionalidad del estado:** A medida que el número de estados aumenta, el problema se vuelve computacionalmente intratable. Enfoques como la descomposición del estado y la abstracción pueden ayudar a reducir la complejidad.
- **Requerimientos computacionales:** Los algoritmos pueden requerir una gran cantidad de recursos computacionales. El uso de técnicas de aproximación y algoritmos de aprendizaje profundo puede ayudar a manejar estos requerimientos.
    - **Aprendizaje por refuerzo profundo (DRL):** Utiliza redes neuronales para aproximar funciones de valor o políticas, lo que permite manejar espacios de estados grandes.
    - **Métodos de aproximación**: Como la aproximación lineal o la aproximación mediante funciones, que simplifican el cálculo de valores y políticas.
    - **Algoritmos de muestreo:** Que reducen la cantidad de datos necesarios explorando selectivamente el espacio de estados y acciones

Task 2 - Pregunta 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 establece que el estado actual de un sistema contiene toda la información relevante para predecir el futuro, sin necesidad de conocer el historial pasado. Es decir, el próximo estado, s', solo depende del estado actual, s, y la acción tomada a, y no de la secuencia completa de eventos previos; a esto se le conoce como independencia del pasado. También hay que considerar la estacionariedad, es decir, que las probabilidades de transición y las recompensas no cambian con el

tiempo. De la misma forma, se considera su uso en sistemas que son completamente observables, es decir, el agente tiene acceso a toda la información de las variables del sistema.

En algunos escenarios estos supuestos pueden dar pie a errores y no ser del todo válido, por ejemplo, en muchos sistemas reales, el estado futuro puede depender de una secuencia de estados anteriores, no solo del estado actual. Por ejemplo, en la gestión de la salud de un paciente, el estado de salud futuro puede depender de la historia completa de tratamientos y enfermedades pasadas.

Con relación a la segunda suposición, en algunos casos, las probabilidades de transición y las recompensas pueden cambiar con el tiempo. Por ejemplo, en mercados financieros, las condiciones del mercado y las recompensas pueden variar significativamente con el tiempo. En cuanto a la observabilidad, existen ciertos sistemas donde las variables no pueden saberse o ser medidas directamente, es decir, cuando el agente no tiene acceso completo a la información del estado actual, lo que requiere modelos más complejos como los Procesos de Decisión de Markov Parcialmente Observables (POMDP), por ejemplo, en la planificación de rutas de tráfico, las condiciones del tráfico pueden verse afectadas por patrones históricos y eventos futuros desconocidos.

Esto tiene implicaciones como, inexactitud en las predicciones, ya que, si los supuestos no se cumplen, las predicciones sobre futuras transiciones y recompensas pueden ser inexactas, lo que lleva a decisiones subóptimas. También la necesidad de modelos más complejos, en escenarios donde los supuestos de Markov no son válidos, se requieren modelos más sofisticados que incorporen la historia o la no estacionariedad, como los POMDP o modelos con memoria.

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.

Las diferentes fuentes de donde pueden provenir la incertidumbre incluyen:

- **Incertidumbre en las transiciones**: Las probabilidades de transición pueden no ser conocidas con precisión, lo que dificulta la modelación exacta del sistema.
- **Incertidumbre en las recompensas**: Las recompensas pueden ser inciertas o variar con el tiempo, complicando la evaluación de políticas.
- **Estados parcialmente observables**: En muchos casos, el estado actual del sistema puede no ser completamente observable, lo que introduce incertidumbre adicional.

Estrategias para una toma de decisiones sólida en entornos inciertos

- **Modelos de decisión robustos**: Utilizar enfoques que consideren la variabilidad y la incertidumbre en las probabilidades de transición y recompensas, como los MDP robustos. Proporciona garantías sobre el rendimiento en entornos inciertos, pero puede ser conservador y no aprovechar completamente las oportunidades en entornos más favorables.
- **Aprendizaje por refuerzo**: Implementar algoritmos de aprendizaje por refuerzo que puedan adaptarse y aprender de la incertidumbre en el entorno. Puede adaptarse a entornos inciertos sin necesidad de un modelo explícito, pero requiere una gran cantidad de datos y puede ser lento en converger.

- **Simulación y análisis de escenarios**: Realizar simulaciones y análisis de escenarios para evaluar el impacto de diferentes fuentes de incertidumbre y desarrollar estrategias de mitigación.
- **Modelos parcialmente observables (POMDP)**: Utilizar Procesos de Decisión de Markov Parcialmente Observables (POMDP) para manejar la incertidumbre en la observación de estados
- **Planificación bajo Incertidumbre:** Utiliza técnicas como la planificación de escenarios o la planificación estocástica para considerar múltiples posibilidades y prepararse para diferentes resultados. Permite anticipar y prepararse para diferentes escenarios inciertos, pero puede ser computacionalmente intensivo y requerir un buen entendimiento del espacio de estados.
- **Aproximación de Modelos:** Utiliza técnicas de aproximación para simplificar el modelo y hacerlo más manejable, como la aproximación lineal o mediante funciones. Reduce la complejidad computacional y permite tratar con espacios de estados grandes, pero puee perder precisión si la aproximación no es adecuada.
- **Reducción de la Dimensión del Problema**: técnicas como descomposición de estados o aprendizaje por representación pueden ayudar a manejar MDP de gran escala. Se hace uso de embeddings en NLP para representar estados con menor cantidad de dimensiones.

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 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.

GITHUB:
https://github.com/danielanavas2002/InteligenciaArtificial/tree/main/HDT/HDT02