UNIVERSIDAD DEL VALLE DE GUATEMALA

Inteligencia Artificial

Sección 10

Catedrático: Alberto Suriano

Laboratorio 4

Astrid Marie Glauser Oliva 21299

Gonzalo Enrique Santizo vega 21504

ARTURO HEBERTO ARGUETA AVILA 21527

BRYAN CARLOS ROBERTO ESPANA MACHORRO 21550

Task 1:

1. **Defina el proceso de decisión de Markov (MDP) y explique sus componentes.**
    a. **Definicion:**

Un proceso de decisión de Markov (MDP) es un marco matemático para modelar la toma de decisiones secuencial en situaciones donde:

- El estado del sistema es incierto.
- Las acciones tomadas por un agente afectan la evolución del sistema.
- El objetivo es encontrar una política que maximice la recompensa esperada a largo plazo.

    b. **Componentes:**

Un MDP se define por los siguientes elementos:

Conjunto de estados (S): Es el conjunto de todos los estados posibles del sistema.

Conjunto de acciones (A): Es el conjunto de acciones que el agente puede tomar en cada estado.

Función de transición de probabilidad (P): Define la probabilidad de que el sistema pase de un estado a otro después de tomar una acción.

Función de recompensa (R): Define la recompensa que el agente recibe al pasar de un estado a otro después de tomar una acción.

Política (π): Es una función que mapea cada estado a una acción. La política define cómo el agente se va a comportar en cada estado.

Explicación de los componentes:

Conjunto de estados: El conjunto de estados puede ser finito o infinito. En algunos casos, el estado del sistema puede ser completamente observable, mientras que en otros casos solo se puede observar parcialmente.

Conjunto de acciones: El conjunto de acciones puede ser finito o infinito. Las acciones pueden ser deterministas o probabilísticas.

Función de transición de probabilidad: La función de transición de probabilidad define la dinámica del sistema. Es importante tener en cuenta que la función de transición de probabilidad puede depender de la política que esté siguiendo el agente.

Función de recompensa: La función de recompensa define el objetivo del agente. El objetivo del agente es encontrar una política que maximice la recompensa esperada a largo plazo.

Política: La política define cómo el agente se va a comportar en cada estado. La política puede ser determinista o probabilística.

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

1. Política:

Una política es una función que mapea cada estado a una acción. En otras palabras, la política define cómo el agente se va a comportar en cada estado. La política puede ser determinista o probabilística.

2. Evaluación de políticas:

La evaluación de políticas consiste en calcular el valor esperado de una política. El valor esperado de una política es la recompensa esperada a largo plazo que el agente recibe al seguir esa política.

3. Mejora de políticas:

La mejora de políticas consiste en encontrar una política que sea mejor que la política actual. Esto se puede hacer mediante la búsqueda local o mediante la programación lineal.

4. Iteración de políticas:

La iteración de políticas es un método para encontrar la política óptima. Este método consiste en alternar entre la evaluación de políticas y la mejora de políticas hasta que se encuentra una política óptima.

En resumen:

La política define cómo el agente se va a comportar en cada estado.
La evaluación de políticas calcula el valor esperado de una política.
La mejora de políticas encuentra una política que sea mejor que la política actual.
La iteración de políticas es un método para encontrar la política óptima.

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

1. Definición:

El factor de descuento es un valor real entre 0 y 1. Un valor de gamma cercano a 1 indica que el agente da mucha importancia a las recompensas futuras, mientras que un valor de gamma cercano a 0 indica que el agente solo le importa las recompensas inmediatas.

2. Influencia en la toma de decisiones:

El factor de descuento influye en la toma de decisiones del agente de la siguiente manera:

Valores altos de gamma: Cuando el valor de gamma es alto, el agente tendrá en cuenta las recompensas futuras a largo plazo. Esto significa que el agente estará dispuesto a sacrificar recompensas inmediatas para obtener mayores recompensas a largo plazo.

Valores bajos de gamma: Cuando el valor de gamma es bajo, el agente solo le importará las recompensas inmediatas. Esto significa que el agente no estará dispuesto a sacrificar recompensas inmediatas para obtener mayores recompensas a largo plazo.

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

1. Iteración de valores:

La iteración de valores se centra en calcular la función de valor óptima, que es la función que mapea cada estado a su valor esperado máximo. El algoritmo comienza con una función de valor arbitraria y luego la actualiza iterativamente hasta que converge a la función de valor óptima.

2. Iteración de políticas:

La iteración de políticas se centra en encontrar una política óptima, que es la política que maximiza la recompensa esperada a largo plazo. El algoritmo comienza con una política arbitraria y luego la mejora iterativamente evaluando la política actual y luego buscando una mejor política.

Diferencias clave:

Objetivo: La iteración de valores se centra en calcular la función de valor óptima, mientras que la iteración de políticas se centra en encontrar una política óptima.

Aproximación: La iteración de valores aproxima la función de valor óptima directamente, mientras que la iteración de políticas aproxima la función de valor óptima indirectamente a través de la evaluación y mejora de políticas.

Convergencia: La iteración de valores siempre converge a la función de valor óptima, mientras que la iteración de políticas solo converge a una política óptima si la política inicial es "greedy".

Eficiencia: La iteración de valores suele ser más eficiente que la iteración de políticas para MDP con un gran número de estados.

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.**

1. Explosión del estado: El número de estados en un MDP puede crecer exponencialmente con el tamaño del problema. Esto puede hacer que la resolución del MDP sea computacionalmente intratable.

2. Maldición de la dimensionalidad: La calidad de la solución a un MDP puede disminuir a medida que aumenta la dimensionalidad del problema.

3. Recompensas diferidas: En algunos MDP, las recompensas pueden ser diferidas por muchos pasos, lo que dificulta la búsqueda de la política óptima.

4. Incertidumbre: En muchos casos, la dinámica del sistema y las recompensas no son completamente conocidas, lo que introduce incertidumbre en el proceso de toma de decisiones.

Enfoques para abordar estos desafíos:

1. Aproximaciones: Existen varias aproximaciones que se pueden utilizar para reducir la complejidad computacional de la resolución de MDP. Estas aproximaciones incluyen:

Agrupamiento de estados: Agrupar estados similares en un solo estado para reducir el número de estados.
Muestreo: Usar técnicas de muestreo para obtener una estimación de la función de valor óptima.
Redes neuronales: Usar redes neuronales para aproximar la función de valor óptima.

2. Planificación jerárquica: Descomponer el problema en subproblemas más pequeños y manejables.

3. Aprendizaje por refuerzo: Usar técnicas de aprendizaje por refuerzo para encontrar una política óptima a través de la prueba y error.

4. Incorporación de información a priori: Incorporar información previa sobre el problema para mejorar la calidad de la solución.

5. Computación paralela: Usar computación paralela para acelerar la resolución del MDP.

Task 2:

6. **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 pueden no ser válidos y sus implicaciones para la toma de decisiones.**

1. Supuestos de la propiedad de Markov:

La propiedad de Markov en los MDP se basa en dos supuestos principales:

Memoria limitada: El estado futuro del sistema solo depende del estado actual, no de los estados pasados.
Independencia de las acciones: La probabilidad de transición de un estado a otro solo depende del estado actual y la acción tomada, no de las acciones pasadas.

2. Escenarios donde los supuestos pueden no ser válidos:

Memoria no limitada:

Dependencia de estados pasados: En algunos casos, el estado futuro del sistema puede depender de estados pasados. Por ejemplo, el estado de un paciente en un hospital puede depender de su historial médico.

Información incompleta: Si el agente no tiene información completa sobre el estado actual del sistema, la propiedad de Markov puede no ser válida.

Dependencia de las acciones:

Efectos persistentes de las acciones: Algunas acciones pueden tener efectos persistentes que influyen en la probabilidad de transición de un estado a otro.

Interacciones entre agentes: En entornos con múltiples agentes, las acciones de un agente pueden afectar la probabilidad de transición de otro agente.

3. Implicaciones para la toma de decisiones:

Si los supuestos de la propiedad de Markov no son válidos, las técnicas tradicionales para resolver MDP pueden no ser precisas. Esto puede llevar a decisiones subóptimas e incluso a resultados indeseables.

4. Enfoques para abordar las limitaciones de la propiedad de Markov:

Modelos de memoria extendida: Incorporar información sobre estados pasados en el modelo del MDP.

Aprendizaje por refuerzo: Usar técnicas de aprendizaje por refuerzo para aprender una política óptima en un entorno sin memoria.

Planificación jerárquica: Descomponer el problema en subproblemas más pequeños donde la propiedad de Markov puede ser válida.

7. **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.**

1. Tipos de incertidumbre:

Incertidumbre de la dinámica del sistema: La probabilidad de transición de un estado a otro no se conoce con precisión.

Incertidumbre de las recompensas: La recompensa asociada a cada estado y acción no se conoce con precisión.

Incertidumbre del modelo: El modelo del MDP no es una representación perfecta del mundo real.

2. Implicaciones de la incertidumbre:

Dificultad para encontrar la política óptima: La incertidumbre dificulta la identificación de la política que maximiza la recompensa esperada.

Riesgo de tomar decisiones subóptimas: Las decisiones tomadas bajo incertidumbre pueden no ser las mejores decisiones posibles.

Necesidad de estrategias para la toma de decisiones robusta: Es necesario desarrollar estrategias para tomar decisiones que sean robustas a la incertidumbre.

3. Estrategias para la toma de decisiones robusta:

Enfoque conservador: Elegir la política que minimiza el peor escenario posible.

Enfoque optimista: Elegir la política que maximiza el mejor escenario posible.

Enfoque bayesiano: Incorporar información probabilística sobre la incertidumbre en el proceso de toma de decisiones.

Enfoque de robustez a la incertidumbre: Buscar políticas que sean robustas a diferentes tipos de incertidumbre.

4. Técnicas para modelar la incertidumbre:

Teoría de la probabilidad: Usar la teoría de la probabilidad para representar la incertidumbre en las transiciones y recompensas.

Lógica difusa: Usar lógica difusa para representar la incertidumbre en los estados y acciones.

Programación estocástica: Usar programación estocástica para optimizar la política bajo incertidumbre.

Task 3:

Link repositorio:

https://github.com/GonzaloSantizo/HT2AI