UNIVERSIDAD DEL VALLE DE GUATEMALA

Inteligencia Artificial - CC3085

Sección 11

Ing. Luis Alberto Suriano Saravia

Hoja de Trabajo No.2

José Pablo Orellana 21970

Diego Alberto Leiva 21752

Gabriel Estuardo García 21352

GUATEMALA, 23 de febrero 2024

Task 1 - Preguntas teóricas

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

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

  El Proceso de Decisión de Markov (MDP, por sus siglas en inglés) es un modelo matemático utilizado en teoría de control y aprendizaje por refuerzo para representar situaciones de toma de decisiones secuenciales bajo incertidumbre. Los MDP son especialmente útiles en problemas donde las acciones tomadas en un estado dado afectan no solo las recompensas inmediatas, sino también los estados futuros y las recompensas asociadas. Los componentes clave de un MDP son los siguientes:

  - **Un conjunto de estados S:** Cada estado representa una situación concreta del problema a resolver.
  - **Un conjunto de acciones A:** Las acciones permiten transitar entre estados. Cada estado está asociado a un subconjunto de acciones que son las que se pueden realizar desde ese estado.
  - **Una función de probabilidad p:** La probabilidad pij(a) denota la probabilidad de transitar al estado j si aplicamos la acción a desde el estado i.
  - **Una función de coste c:** El coste ci(a) especifica el coste asociado a realizar la acción a desde el estado i. (Ruiz Valverde, 2022).

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

  - **Explosión de estados (State Space Explosion):** A medida que el número de estados posibles aumenta, el espacio de estados puede volverse prohibitivamente grande. Esto hace que sea difícil almacenar y procesar toda la información necesaria para la toma de decisiones.
    - **Enfoque:** Métodos de aproximación, como la aproximación de funciones de valor o la aproximación de políticas, pueden ayudar a manejar el espacio de estados grandes.

- **Explosión de acciones (Action Space Explosion):** Similar al problema de la explosión de estados, el número de acciones posibles puede volverse grande, especialmente en entornos continuos.
    - **Enfoque:** Reducción de dimensionalidad mediante técnicas como la selección de características relevantes o el uso de técnicas específicas para entornos continuos.

- **Complejidad computacional:** La resolución exacta de MDP a gran escala puede ser computacionalmente costosa y a menudo requiere mucho tiempo de procesamiento.
    - **Enfoque:** Uso de métodos de aproximación, paralelización de cálculos, y técnicas de muestreo para estimar soluciones de manera eficiente.

- **Modelo desconocido:** En algunos casos, el modelo de transición o las recompensas pueden ser desconocidos o difíciles de modelar con precisión.
    - **Enfoque:** Enfoques basados en aprendizaje por refuerzo, donde el agente interactúa con el entorno para aprender directamente de la experiencia, pueden ser útiles.

- **Exploración y explotación:** En entornos complejos, encontrar un equilibrio entre explorar nuevas acciones y explotar el conocimiento existente puede ser desafiante.
    - **Enfoque:** Estrategias de exploración cuidadosamente diseñadas, como epsilon-greedy o métodos basados en la incertidumbre, pueden ayudar a abordar este problema.

- **Horizonte temporal largo:** En problemas con horizontes temporales largos, es más difícil tomar decisiones óptimas debido a la incertidumbre acumulativa.
    - **Enfoque:** Uso de métodos de aprendizaje por refuerzo con descuento para priorizar recompensas inmediatas, y técnicas de planificación con horizontes temporales recortados para hacer el problema manejable.

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

  En el contexto de los Procesos de Decisión de Markov (MDP), los conceptos de política, evaluación de políticas, mejora de políticas e iteración de políticas son fundamentales para comprender cómo se abordan los problemas de toma de decisiones.

  - **Política (Policy):**
    Una política en un MDP es una estrategia que especifica la acción a tomar en cada estado del sistema. Puede ser determinista, donde se elige una acción específica en cada estado, o estocástica, donde se define una distribución de probabilidad sobre las acciones posibles.

  - **Evaluación de Políticas (Policy Evaluation):**
    Es el proceso de determinar la calidad o el valor de una política dada en un MDP. Implica calcular la función de valor de esa política, que mide la recompensa esperada a lo largo del tiempo al seguir esa política y puede expresarse mediante funciones de valor como la función de valor de estado o la función de valor de acción.

  - **Mejora de Políticas (Policy Improvement):**
    Este proceso implica modificar o actualizar una política con el objetivo de hacerla más efectiva, es decir, aumentar la recompensa esperada. La mejora de políticas se basa en la información obtenida durante la evaluación de políticas. Una política se considera mejor si para al menos un estado, la acción sugerida por la nueva política es preferible a la acción sugerida por la política anterior.

  - **Iteración de Políticas (Policy Iteration):**
    La iteración de políticas es un proceso iterativo que combina la evaluación y la mejora de políticas. Comienza con una política inicial, luego alterna entre la evaluación de esa política y la mejora de la misma. Este proceso continúa hasta que se alcanza una política óptima, es decir, una política que no se puede mejorar más.

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

  El factor de descuento, denotado como γ γ (gamma), es un parámetro fundamental en los Procesos de Decisión de Markov (MDP). Este factor influye en la toma de decisiones al modelar la preferencia del agente por recompensas inmediatas frente a recompensas futuras y al introducir la noción de a importancia del tiempo en el proceso de toma de decisione El factor de descuento varía en el rango de 0 a 1, y su impacto en la toma de decisiones se puede entender de la siguiente manera.

  - **γ = 0:** Cuando γ es igual a cero, el agente solo valora las recompensas inmediatas y no tiene en cuenta las recompensas futuras. Esto implica que el agente es "miopemente" enfocado en maximizar la recompensa inmediata y no considera las consecuencias a largo plazo de sus acciones.

  - **γ = 1:** Cuando γ es igual a uno, el agente valora todas las recompensas, independientemente de cuándo ocurran en el tiempo. En este caso, el agente toma decisiones considerando tanto las recompensas inmediatas como las futuras, sin importar cuán distante esté en el futuro.

  - **0 < γ < 1:** En la mayoría de los casos, el factor de descuento se elige en este rango. Un valor de γ cercano a 1 indica que el agente valora las recompensas futuras, pero da menos importancia relativa a medida que se van alejando en el tiempo. Esto refleja la noción de impaciencia del agente, que tiende a preferir recompensas inmediatas, pero aún tiene en cuenta las recompensas futuras.

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

  La Iteración de Valores (Value Iteration) y la Iteración de Políticas (Policy Iteration) son dos enfoques fundamentales para resolver Procesos de Decisión de Markov (MDP). Ambos métodos están diseñados para encontrar la política óptima, es decir, la estrategia de toma de decisiones que maximiza la recompensa esperada a lo largo del tiempo.

  - Iteración de Valores: En este enfoque, el algoritmo itera sobre las funciones de valor de los estados. Durante cada iteración, actualiza las estimaciones de los valores de los estados para converger hacia la función de valor óptima.
  - Iteración de Políticas: Este método, en cambio, itera sobre las políticas directamente. Alternando entre evaluación y mejora de políticas, busca encontrar la política óptima al ajustar continuamente las decisiones tomadas en cada estado.

Task 2 - Preguntas Analiticas

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 en los Procesos de Decisión por sus siglas en inglés MDP, establece que el futuro depende únicamente del estado presente y no de los estados pasados. Esto implica que la transición de estados y las recompensas son independientes del historial de estados y acciones. Aunque esta propiedad es fundamental para simplificar la modelización y el análisis matemático, hay escenarios en los que estos supuestos pueden no ser válidos como:

- Dependencia de estados pasados: En algunas situaciones, las decisiones futuras pueden depender no solo del estado presente sino también de los estados pasados. Por ejemplo en los juegos competitivos donde la estrategia del oponente se basa en el historial de acciones del agente.
- Cambios de entorno: Si el entorno es dinámico y sufre cambios que no pueden ser capturados por el estado actual, la propiedad de Markov puede no ser válida. Por ejemplo, en entornos donde las condiciones cambian abruptamente sin señales claras en el estado actual.
- Información incompleta: Si el agente no tiene acceso a toda la información relevante del entorno en el estado actual, la propiedad de Markov puede no mantenerse. Por ejemplo, en entornos donde ciertas variables clave están ocultas o no son observables directamente.
- Persistencia de la memoria: En algunos casos, la persistencia de la memoria puede ser crucial para la toma de decisiones óptimas, lo cual no se cumple bajo la propiedad de Markov. Por ejemplo, en tareas donde la experiencia pasada tiene un impacto significativo en las decisiones futuras.

Las implicaciones de estos casos pueden ser la necesidad de modelos más complejos, como los Procesos de Decisión de Markov Parcialmente Observables, los cuales permiten el manejo de la incertidumbre y la dependencia de estados pasados (Surton & Barto, 2018).

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.

Desafíos:

- Estimación de modelos: La incertidumbre en las transiciones del estado y las recompensas puede ser difícil de modelar con precisión, especialmente en los entornos complejos y dinámicos.
- Exploración contra Explotación: En los entornos inciertos, es importante equilibrar la exploración de nuevas acciones con la explotación de acciones conocidas para maximizar las recompensas a largo plazo.
- Ruido estocástico: la presencia de ruido estocástico en el entorno puede dificultar la distinción entre acciones efectivas y no efectivas.
- Incertidumbre parcial: En entornos parcialmente observables, la incertidumbre sobre el estado actual del entorno puede complicar la toma de decisiones.

Estrategias:

- Exploración inteligente: Utilizar estrategias de exploración que equilibren la búsqueda de la información con la explotación de conocimientos actuales.
- Actualización de creencias: Mantener una distribución de creencias sobre las posibles configuraciones del entorno y actualizarla en función de las observaciones.
- Planificación robusta: Utilizar algoritmos de planificación que sean robustos frente a la incertidumbre y capaces de adaptarse a diferentes escenarios.
- Aprendizaje adaptable: Incorporar capacidad de aprendizaje para ajustar el modelo y las políticas de decisión en función de la experiencia y las nuevas observaciones.

(Bongard, 2008).

Referencias

García Hernández. (2009, diciembre). Acceleration of association-rule based markov decision processes. https://www.redalyc.org/pdf/474/47413020008.pdf

Ruiz Valverde, A. (2022). Algoritmos heurísticos para procesos de decisión de Markov (p. 69). Universidad de Málaga. https://riuma.uma.es/xmlui/bitstream/handle/10630/25975/Ruiz%20Valverde%20Antonio%20Memoria.pdf?sequence=1&isAllowed=y

Ruiz-Loza, S., & Hernandez, B. (2014). Markov Decision Process and Micro scenarios for Crowd Navigation and Collision Avoidance (spanish language). En Research in Computing Science (p. 116).

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.

Bongard, J. (2008). Probabilistic Robotics. Sebastian Thrun, Wolfram Burgard, and

Dieter Fox. (2005, MIT Press.) 647 pages. *Artificial Life*, *14*(2), 227-229.

https://doi.org/10.1162/artl.2008.14.2.227