Minicurso: Introducción al Planeamiento Probabilistico

Dra. Karina Valdivia Delgado – Universidad de Sao Paulo, Brasil

Mausam y Andrey Kolobov - Planning with Markov Decision Processes
Objetivo: Introducción extensa a la teoría y algoritmos en planeamiento probabilístico.

Introducción al Planeamiento Probabilístico en IA

Definición: El planeamiento probabilístico aborda la toma de decisiones en entornos inciertos donde las acciones no siempre tienen resultados predecibles.
Modelo Fundamental: Los Procesos de Decisión de Markov (MDPs) son el modelo principal utilizado, propuesto por Bellman y Howard en los años 50, popularizados en IA en los 90s.

Outline

- Fundamentals of MDPs

Un Proceso de Decisión Markoviano es un conjunto de estados, acciones, transiciones y recompensas. La solución optima es la mejor política en terminos de la recompensa final.

- Uninformed Algorithms.

Algoritmos que no requieren conocimiento previo del dominio.

Planeamiento en IA → Planeamiento Probabilístico

- MDP (Markov Decision Process): Modelo matemático utilizado para tomar decisiones en situaciones inciertas. Se utiliza cuando:
    - El entorno es estático o dinámico.
    - El ambiente puede ser completamente o parcialmente observable.
    - El conocimiento del estado del sistema es limitado o completo.
    - Las acciones pueden ser deterministas (100% predecibles) o estocásticas (no deterministas).

- Problema de Planeamiento: El objetivo es determinar la mejor secuencia de acciones para que un agente (robótico o de software) realice en un entorno dado. La solución clásica es un plan secuencial de acciones (mundo determinista), pero en el planeamiento probabilístico se utilizan MDPs.

Conceptos Clave en el Planeamiento Probabilístico

- Recompensas: El objetivo es maximizar la suma de recompensas a lo largo de una secuencia de acciones.
- Función de Utilidad: Método para evaluar la calidad de una política basada en las recompensas acumuladas.
- Política: Secuencia de decisiones o acciones que un agente toma para optimizar su desempeño en el entorno.
- Política Óptima: La política que maximiza el valor esperado de las recompensas.
- Desafíos:
    - Agentes en entornos parcialmente observables.
    - Procesos estocásticos donde los resultados de las acciones no son predecibles.
- Aplicaciones:
    - Economía, teoría de grafos, juegos, control, inteligencia artificial, psicología.

Aprendizaje por Refuerzo

Cuando no se conocen las probabilidades del modelo, el problema se resuelve mediante aprendizaje por refuerzo.

- MDP también está detrás del aprendizaje por refuerzo, donde el objetivo es encontrar la mejor secuencia de acciones para maximizar recompensas.
- El aprendizaje por refuerzo permite enfrentar decisiones cíclicas y complejas a largo plazo.

Fundamentos de la Evaluación de Políticas

1. Utilidad Aditiva Lineal Esperada (ELA):

Es el método más utilizado en MDPs para calcular la suma de recompensas esperadas, el uso de un factor de descuento asegura que la suma sea finita, incluso si hay un número infinito de pasos de decisión.

2. Principio de Optimalidad:

Si la calidad de todas las políticas se puede medir a través de la ELAU, existe una política que es óptima en cada paso.

Decisiones de un Solo Paso (Ejemplo)

Problema:

- 100% de ganar 1M contra 50% de ganar 2M.

Cálculo de la esperanza:

- 1
- (0.5 * 2) + (0.5 * 0)

Modelos clave a comparar

Con ELAU, es posible evaluar y comparar distintos modelos de MDPs con políticas bien definidas.

- Horizonte finito.
- Horizonte infinito con recompensas descontadas.
- Camino más corto estocástico.

Métodos para encontrar política óptima

Aseguran decisiones precisas y eficientes.
1. Evaluación iterativa: Se refina el valor de cada política hasta hallar la mejor.
2. Sistema de ecuaciones: Se resuelven ecuaciones para identificar la política que maximiza las recompensas.

Resumen

- MDP: Un proceso que consiste en estados, acciones, transiciones y recompensas.
- Política: La solución es una política que define la mejor acción en cada estado.
- Factor de Descuento: Limita las recompensas acumuladas y permite comparar políticas.
- Política Óptima: Es la política que maximiza el valor esperado de las recompensas.