Procesos de decisión de Márkov

1. Quien fue Markov

   Matemático ruso que desarrolló la moderna teoría de procesos estocásticos. Trabajó en la casi totalidad de los campos de la matemática. En el campo de la la teoría de la probabilidad, profundizó en las consecuencias del teorema central del límite y en la ley de los grandes números. En su honor, lleva su nombre un tipo muy especial de procesos estocásticos.
   Sus primeras investigaciones versaron sobre análisis y teoría de números, en particular sobre las fracciones continuas, límites de integrales, teoría de aproximaciones y convergencia de series. En 1900 estudió la teoría de probabilidades. Demostró a partir de supuestos muy generales el llamado teorema central del límite, que establece que la suma de un número grande de variables aleatorias independientes se aproxima a una distribución gaussiana.
   Tras este trabajo, estudió las variables dependientes e introdujo el concepto de sucesos encadenados. Markov extendió los resultados clásicos de sucesos independientes a cierto tipo de sucesos encadenados, conocidos como sucesos markovianos, que son aquellos cuyo estado en un instante de tiempo depende de uno o varios estados cronológicamente anteriores.

2. Procesos de Decisión de Markov

   - Procesos de decisión secuenciales.
   - Procesos de decisión de Markov (MDPs).
   - Tecnicas de solución:
     - Iteración de Valor
     - Iteración de política
   - MDPs Parcialmente observables (POMDPs)
   - Extensiones
     - MDPs factorizados
     - Abstracción, descomposición
   - Aplicaciones
   - Problema de decisión que involucra un conjunto de decisiones cuyo resultado (utilidad) se conoce hasta el final.
   - Se considera que se tiene una serie de estados y decisiones asociadas en el tiempo.
   - Se tiene incertidumbre asociada con los resultados de las acciones (MDP), y posiblemente también con los estados (POMDP).

3. Técnicas que existen para resolver estos procesos
   - **Los métodos principales para resolver MDPs son:**
     γ = factor de descuento

     - Iteración de valor (Bellman, 57)
       $V*(s) = \max_a \{ R(s, a) + \gamma\, \Sigma_{s'} P(s' \mid s, a)\, V*(s') \}$

     - Iteración de política (Howards, 60)
       $\pi*(s) = \arg\max_a \{ R(s, a) + \gamma\, \Sigma_{s'} P(s' \mid s, a)\, V*(s') \}$

     - Programación lineal (Puterman, 94)

- **Soluciones aproximadas para POMDP**
    - Representar un POMDP de horizonte finito a través de un “policy tree” (acciones | observaciones | acciones…), resolviéndolo en forma recursiva a través de planes condicionales.
    - Para POMDP de horizonte finito la función de valor es convexa y lineal a pedazos (vectores α). Se puede aproximar mediante un subconjunto de vectores que dominan a los demás, y mediante esto encontrar una política aprox. óptima.
    - Considerar un número finito de pasos y modelar el problema como una red de decisión dinámica – la aproximación depende del número de estados que se “ven” hacia delante (lookahead).

4. Aplicaciones
    - Manejo de inventarios
    - Mantenimiento de equipos y carreteras
    - Control de sistemas de comunicaciones
    - Modelado de procesos biológicos
    - Planeación en robótica móvil
    - Construcción de mapas / localización
    - Control de procesos industriales
    - Control de aviones
5. Referencias
https://www.ugr.es/~eaznar/markov.htm