Nombre: Andrés Zeas

Materia: Inteligencia Artificial I

Procesos de decisión de Markov

Introducción

Un proceso de decisión de Markov es un proceso estocástico de decisiones secuenciales el cual se caracteriza por sus funciones de recompensa y transición que dependen únicamente dele estado actual del sistema y la acción actual. Estos procesos modelas sistemas dinámicos bajo el control de un decisor, cuyo objetivo es escoger una secuencia de acciones que llamaremos política que optimiza el desempeño del sistema sobre el horizonte de toma de decisiones. Las acciones tomadas por el decisor en el presente afecta el estado del futuro del sistema, haciendo así que las decisiones posteriores se vean afectadas también.

Propiedad de Markov

La propiedad de Markov nos muestra que el futuro es independiente del pasado, dado el presente, lo cual se expresa en la siguiente formula:

$$\mathbb{P}\left[S_{t+1} \middle| S_t\right] = \mathbb{P}\left[S_{t+1} \middle| S_1, \ldots, S_t\right]$$

La cual significa que el estado actual (representado por S_t) contiene toda la información relevante de los estados pasados (S_1,…. S_t), por lo tanto ya no nos serviría tener la mayor información de los estados pasados.

Matriz de transición de estados

Esta matriz nos muestra cual sería la probabilidad de transición desde un estado S a un estado S’ y en donde cada fila sumaria uno, se vería de siguiente manera:

$$\begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & & \\ P_{n1} & & P_{nn} \end{bmatrix}$$

Definición

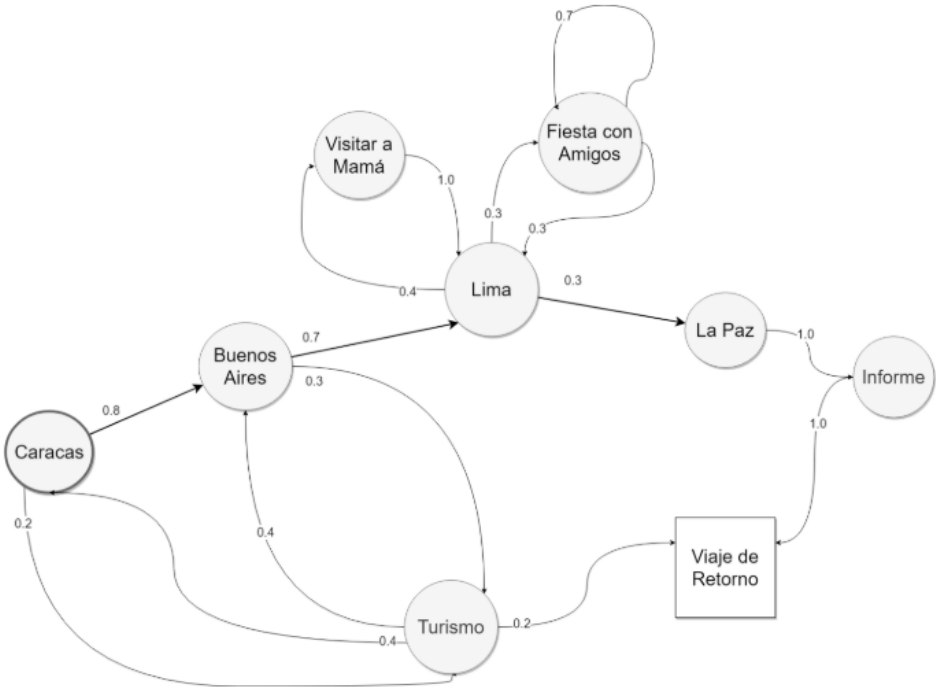
Un procesos de Markov es un procesos sin memoria y aleatorio; en otras palabras es una secuencia de estados aleatorios que posee la propiedad de Markov.

Se podría definir el proceso de Markov como una tupla <S, P>

- S es una lista de estaos a los cuales puede pertenecer.
- P es una matriz de transición de estado.

Ejemplo

En este ejemplo se muestra el caso de un empleado del área de TI en una empresa de maquinarias que reside en México, se le ha encomendado la labor de visitar varias ciudades de Sudamérica donde conversará con varios consultores para encontrar la mejor oferta sobre una consultoría de optimización de procesos de producción, el objetivo de este empleado será escribir un informe en el cual dará su opinión de cada proveedor luego de visitar estas ciudades, en este viaje se verá tentado por hacer turismo en las bellas ciudades que visitará o en permanecer un tiempo en Lima, su ciudad natal, en la cual puede distraerse con amigos o con familia.



Los números representan la probabilidad de ir al siguiente estado.

Ahora veremos cómo podemos sacar muestras de la cadena de Markov propuesta donde se iniciará desde nuestro primer destino, Caracas (S_1= Caracas).

- CARACAS, TURISMO, VIAJE DE RETORNO
- CARACAS, BUENOS AIRES, LIMA, FIESTA CON AMIGOS, FIESTA CON AMIGOS, LIMA, VISITAR A MAMÁ, LIMA, LA PAZ, INFORME, VIAJE DE RETORNO
- CARACAS, BUENOS AIRES, LIMA, VISITAR A MAMÁ, LIMA, LA PAZ, INFORME, VIAJE DE RETORNO
- CARACAS, BUENOS AIRES, LIMA, LA PAZ, INFORME, VIAJE DE RETORNO

Ahora observaremos la matriz de transición de estados para este caso:

	CARACAS	TURISMO	BUENOS AIRES	LIMA	VISITAR A MAMÁ	FIESTA CON AMIGOS	LA PAZ	INFORME	VIAJE DE RETORNO
CARACAS		0.2	0.8						
TURISMO	0.4		0.4						0.2
BUENOS AIRES		0.3		0.7					
LIMA					0.4	0.3	0.3		
VISITAR A MAMÁ				1.0					
FIESTA CON AMIGOS				0.3		0.7			
LA PAZ								1.0	
INFORME									1.0
VIAJE DE RETORNO									1

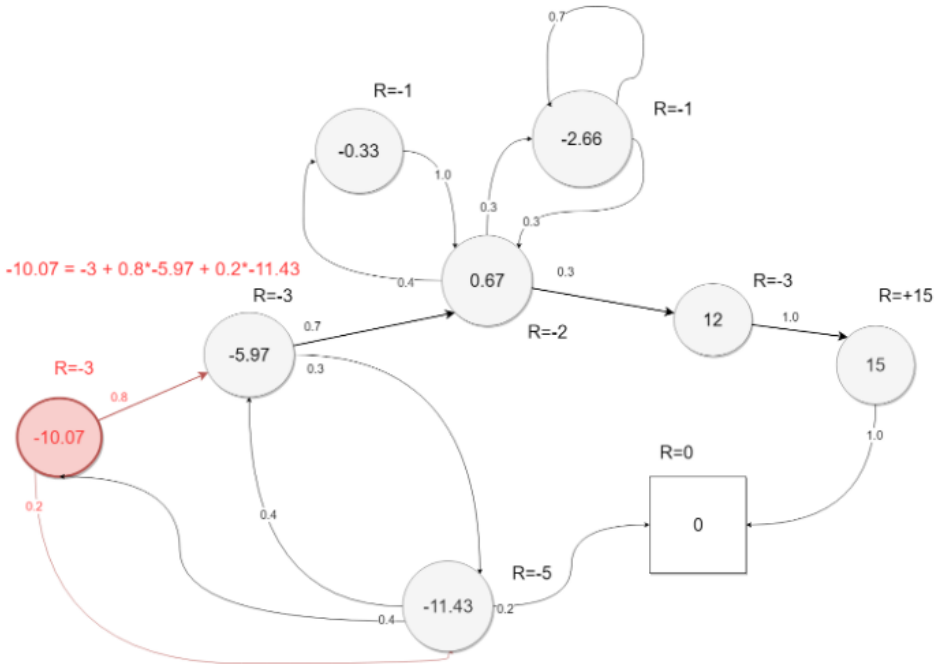
Ecuación de Bellman

La ecuación de Bellman lo que hace es partir la función de valor en dos, en la recompensa inmediata de ese estado y el valor que vas a obtener luego de ese estado en adelante.

$$
\begin{aligned}
v(s) &= \mathbb{E}\left[G_t \middle| S_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots \middle| S_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma\left(R_{t+2} + \gamma R_{t+3} + \cdots\right) \middle| S_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma G_{t+1} \middle| S_t = s\right] \\
&= \mathbb{E}\left[R_{t+1} + \gamma\left(S_{t+1}\right) \middle| S_t = s\right]
\end{aligned}
$$

Básicamente para hallar el valor de un estado se mira en los siguientes y en valor de cada uno de estos, luego se suman todos estos valores para representar valor del estado inicial.

Ahora mostraremos la ecuación de Bellman aplicado al MRP de ejemplo, antes explicado.



Como se ve en nuestro gráfico explicaremos el valor de un estado específico (el que se encuentra de color rojo), se toma los valores de los dos posibles estados en los que puede terminar el agente desde el estado inicial, estos dos estados tienen un valor y una probabilidad de terminar en cada uno, estos se multiplican y se suman para hallar el valor del estado inicial.

Resolviendo la ecuación de Bellman

La ecuación de Bellman es lineal y puede ser resuelta directamente aunque su complejidad es $O(n^3)$ para n estados, por lo tanto solo puede ser resuelta de esta manera para MRPs pequeños, para otros casos es recomendado otras técnicas como programación dinámica u otras técnicas.