

Investigacion sobre los procesos de decision de Markov

Procesos de Markov

un proceso de Markov es un proceso sin memoria y aleatorio; en otras palabras, es una secuencia de estados aleatorios que posee la propiedad de Markov. Los Procesos de Decision de Markov son la forma idealizada matemáticamente del problema de aprendizaje por refuerzo, para el cual se podría encontrar un enunciado teórico preciso que pueda describirla. los MDP describen formalmente el medio ambiente en el cual se desarrolla el RL. Con los MDP se introducen varios elementos clave para la descripción matemática del problema, como el retorno, funciones de valor y las ecuaciones de Bellman.

Propiedad de Markov

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

$$\mathbb{P}[S_{t+1}|S_t] = \mathbb{P}[S_{t+1}|S_1, \dots, S_t]$$

La cual significa que el estado actual (representado por S_t) contiene toda la información relevante de los estados pasados (S_1, \dots, S_t)

Matriz de transición de estados

nos muestra cual sería la probabilidad de transición desde un estado S a un estado S' y en donde cada fila sumaría uno.

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

Procesos de Recompensa de Markov

Los procesos de recompensa de Markov se pueden ver como procesos de Markov normales con valores que juzgan que tan positivo es estar en un estado, esto traería un cambio a la definición de procesos de Markov que tuvimos previamente agregándole dos nuevas variables.

Función de valor

Se representa como $v(s)$ y nos muestra cual es la recompensa esperada desde ese estado hasta el final de la secuencia.

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.