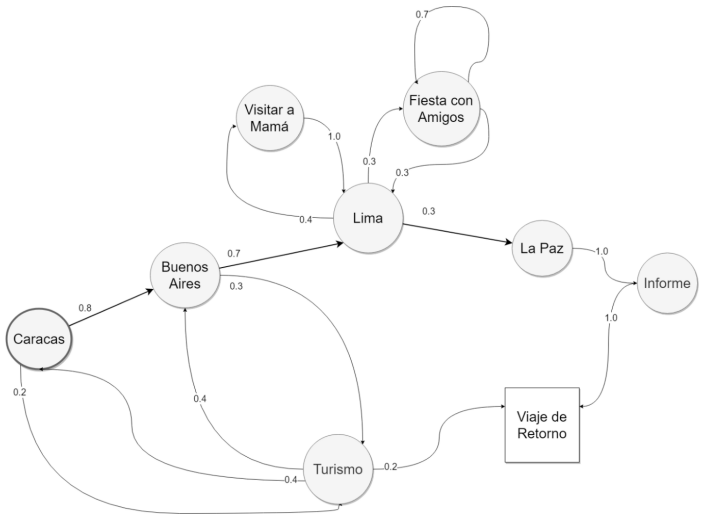
Andres Bermeo

Procesos de Decisión de Markov

- Problema de obtener la política óptima en un ambiente observable – MDP.
- El método clásico para resolver estos problemas se conoce como “iteración de valor” (value iteration).
- La idea básica es calcular la utilidad de cada posible estado y usar éstas para seleccionar la acción óptima en cada estado.
- Otros métodos de solución son “iteración de política” (policy iteration) y programación lineal (al transformar el problema a un problema de optimización lineal).
- Los métodos principales para resolver MDPs son:

– Iteración de valor:

En el caso de horizonte infinito, se puede obtener la util idad de los estados y la política óptima, mediante un mét odo iterativo.
En cada iteración (t+1), se estima la utilidad de cada est ado basada en los valores de la iteración anterior (t):



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

- S es una lista de estados a los cuales puede pertenecer.
- P es una matriz de transición de estado.
- R es la recompensa inmediata en el estado donde nos encontraríamos, se puede expresar de la siguiente manera.

$$R_s = \mathbb{E}\left[R_{t+1} \middle| S_t = s \right]$$

- Y es un valor de descuento que va entre 0 y 1.