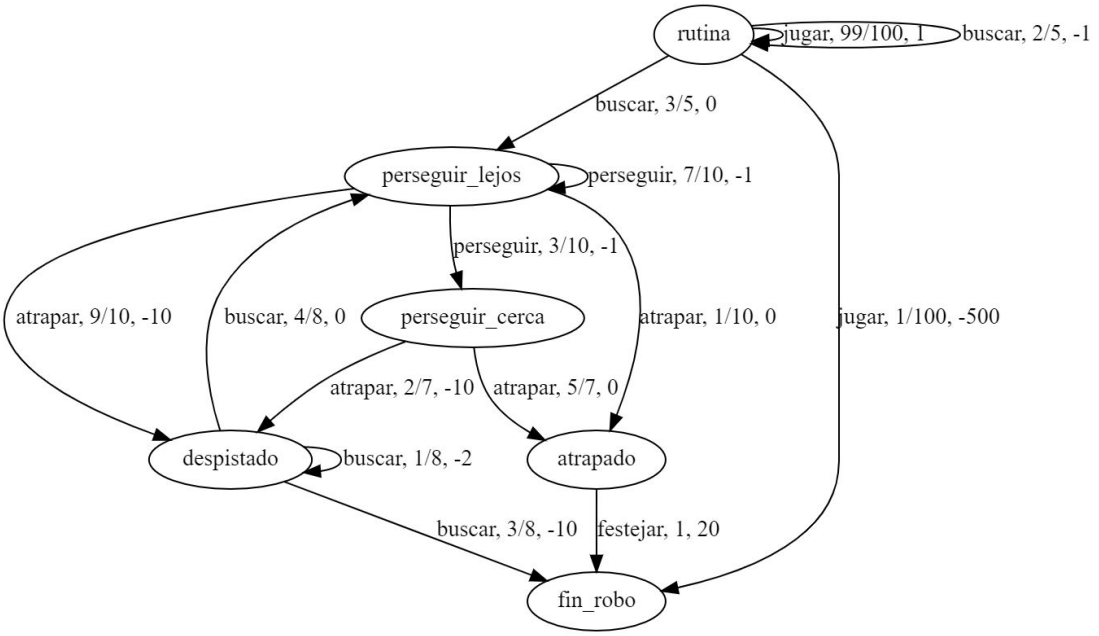
Universidad ORT Uruguay
Inteligencia Artificial
29/04/2020

Tarea 2 - MDP, Monte Carlo, Bellman

El objetivo del siguiente ejercicio es experimentar con Agentes Model Based en particular aquellos que usan MDPs.

MDP dado:



Nota: Todas las acciones que no estén presentes para un estado *s*, se asumen no posibles en dicho estado.
Estado final: *fin_robo*

Se pide:

1. Crear un ambiente basado en la definición de OpenAI Gym que funcione según el MDP dado. Este MDP oculto, dicta las reglas del ambiente. El ambiente debe poder crearse definiendo el estado inicial del agente en este MDP.
2. Crear un agente con *policy random* que explore el ambiente desde un estado inicial aleatorio y genere secuencias episódicas sobre este.
3. Utilizando los episodios generados en el punto 2, aplicar Monte Carlo para estimar el MDP que oculta el ambiente.
    a. Visualizar gráficamente el MDP oculto del ambiente y el MDP estimado por Monte Carlo. Se sugiere el uso de graphviz.
    b. Realizar una comparación tomando en cuenta al menos 3 cantidades de episodios distintas.
4. Sobre uno de los MDPs estimados, se pide implementar Policy Iteration y Value Iteration obteniendo en ambos casos políticas determinísticas. Discuta las mismas entre ellas y según el MDP estimado y el real.
5. Se pide:
    a. Crear un agente para al menos una de las políticas anteriores y probarlo sobre el ambiente.
    b. Comparar y discutir las recompensas obtenidas al ejecutar el agente contra los valores estimados para las política en el punto 4.

Entregables: Código (.py) y un breve informe (.pdf), o una Jupyter Notebook (.ipynb + .py + .html) autocontenida claramente numerada según letra.