Programming Assignment 3 : CS 747

Bhaskar Gharu, 180050023

November 2020

Direction-Action mapping with no king moves

- up - 0, right - 1, down - 2, left - 3

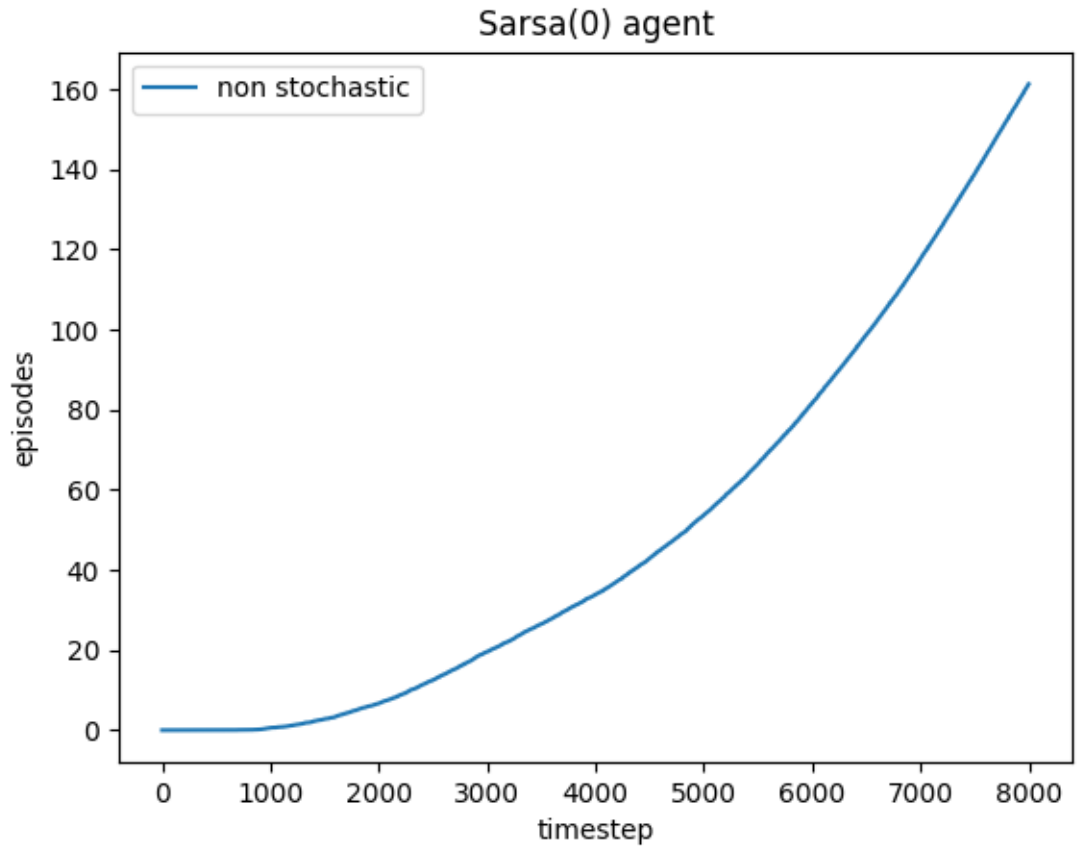
Direction-Action mapping with king moves

- up - 0, right - 1, down - 2, left - 3
- up-right - 4, down-right - 5, down-left - 6, up-left - 7

Parameters and Assumption

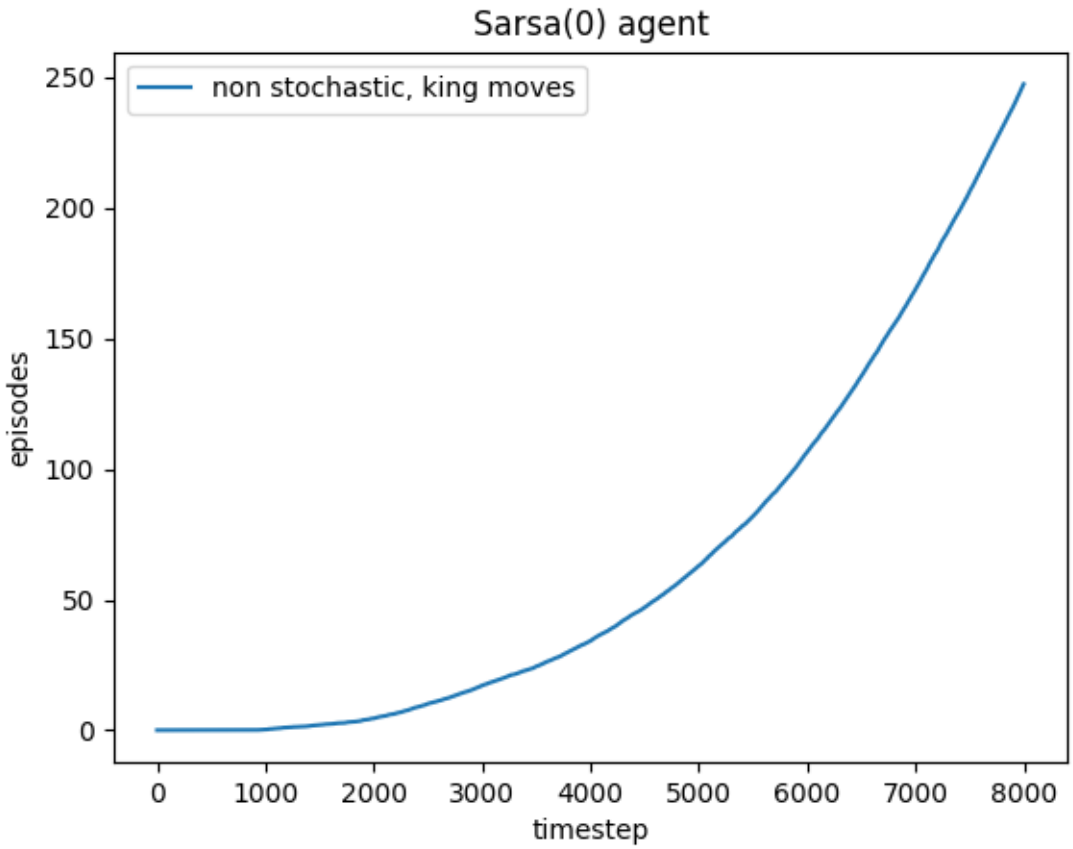
- epsilon - 0.1, alpha(learning rate) - 0.5, timesteps - 8000
- used epsilon greedy policy for choosing actions in all algorithms

Sarsa(0) agent in Non stochastic no king moves environment



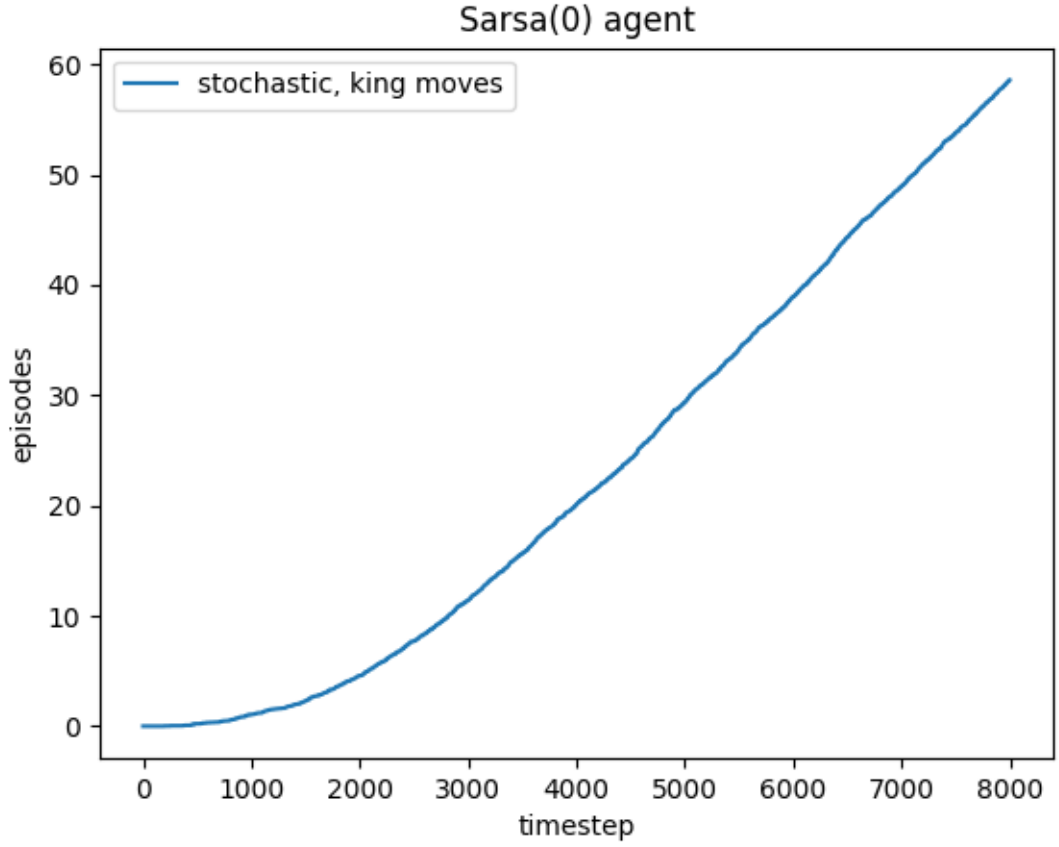
the graph is as expected and converges to the optimal action value function but due to epsilon some randomness in action taken is observed

Sarsa(0) agent in Non stochastic, king moves



With additional moves path is shorter, so more episodes are observed

Sarsa(0) agent in stochastic, king moves



Since there is no fixed path we observe very less episodes for given timesteps

Algorithm comparison for non stochastic and non king moves

Q learning converges the fastest, and sarsa is slowest to converge, for expected sarsa used the custom probability distribution of epsilon/(num actions) for action with sub optimal(non max) value and (1 - epsilon + epsilon/(num actions)) for action with optimal(max) value