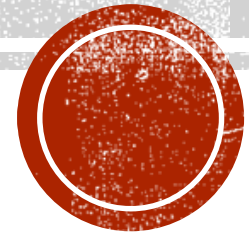# RESULTS OF EXPERIMENTS

Skanda Vaidyanath

Institute of Creative Technologies

6/18/2019

ALGORITHMS

- Q learning

- SARSA

- Expected SARSA

- Monte Carlo (in pipeline)

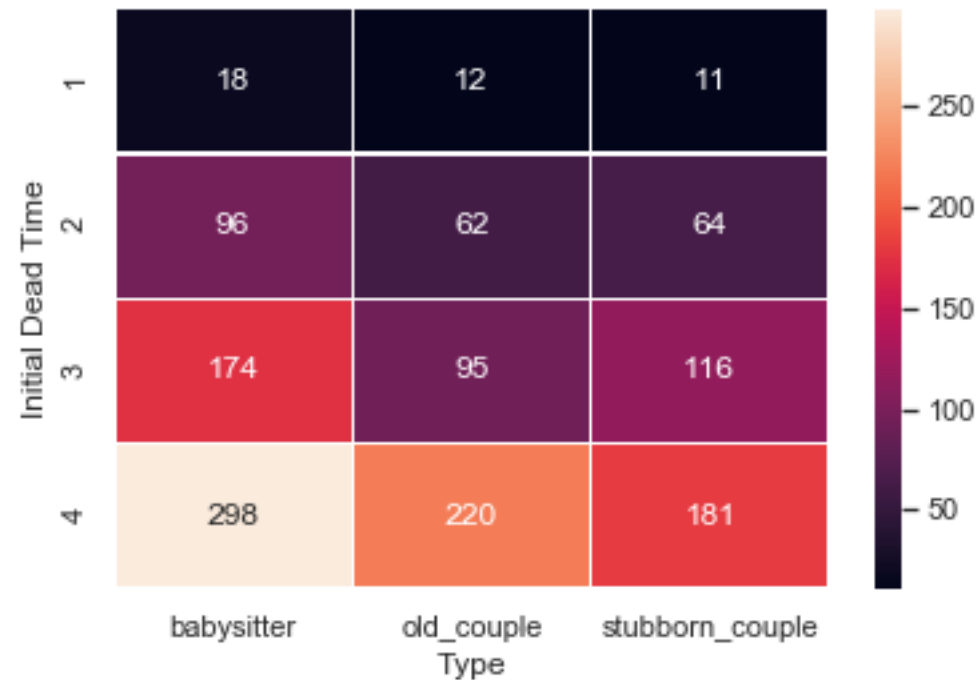- Off Policy Monte Carlo with Weighted Importance Sampling (in pipeline)

Q-LEARNING
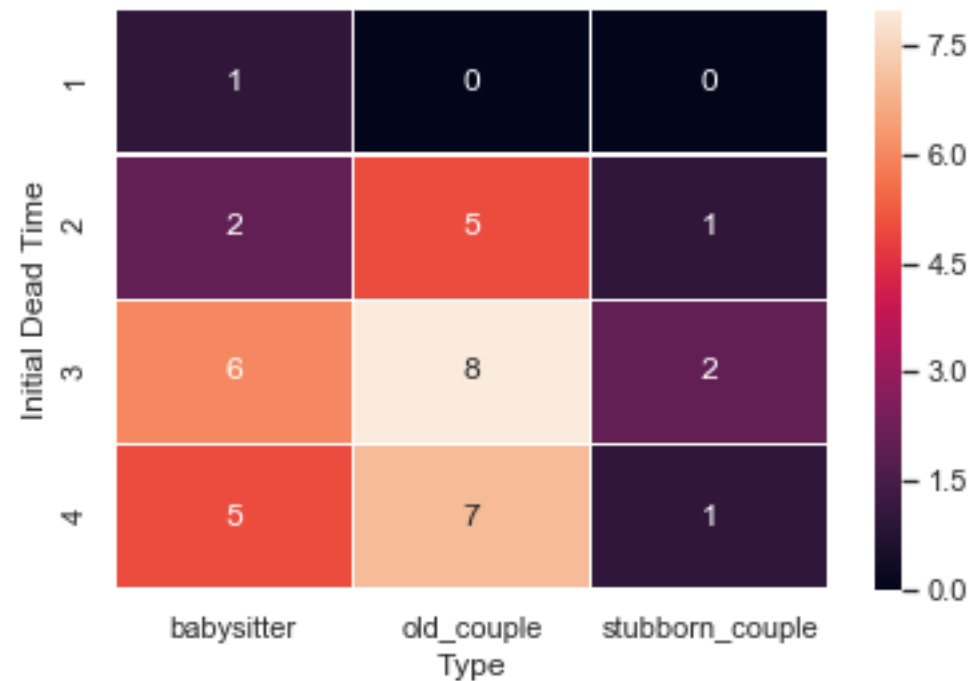
- Training curve for different initializations

Q-LEARNING

- Training statistics for the number of people saved under different initializations

Q-LEARNING

- Testing statistics for the number of people saved under different initializations

Q-LEARNING

- Sample trajectory during testing time (more trajectories in the .txt file)

Episode: 11

old_couple

[OP_BUSY=1, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.]

WARN [OP_BUSY=0, STATUS=WARNING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] VEHICLE GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] SP-NEGOTIATE [OP_BUSY=0, STATUS=SPN, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).] UAV GUIDE [OP_BUSY=0, STATUS=SPN, DEAD_TIME=2, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).] QUERY-FOR-GUIDE-DETAILS [OP_BUSY=0, STATUS=SPN, DEAD_TIME=2, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).] QUERY-FOR-GUIDE-DETAILS [OP_BUSY=0, STATUS=READY_TO_MOVE, DEAD_TIME=2, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).] QUERY-FOR-GUIDE-DETAILS [OP_BUSY=0, STATUS=READY_TO_MOVE, DEAD_TIME=2, GUIDE_TYPE=VEHICLE, NEGOTIATIONS=SPN(s).] VEHICLE GUIDE [OP_BUSY=0, STATUS=SAVED, DEAD_TIME=2, GUIDE_TYPE=VEHICLE, NEGOTIATIONS=SPN(s).]
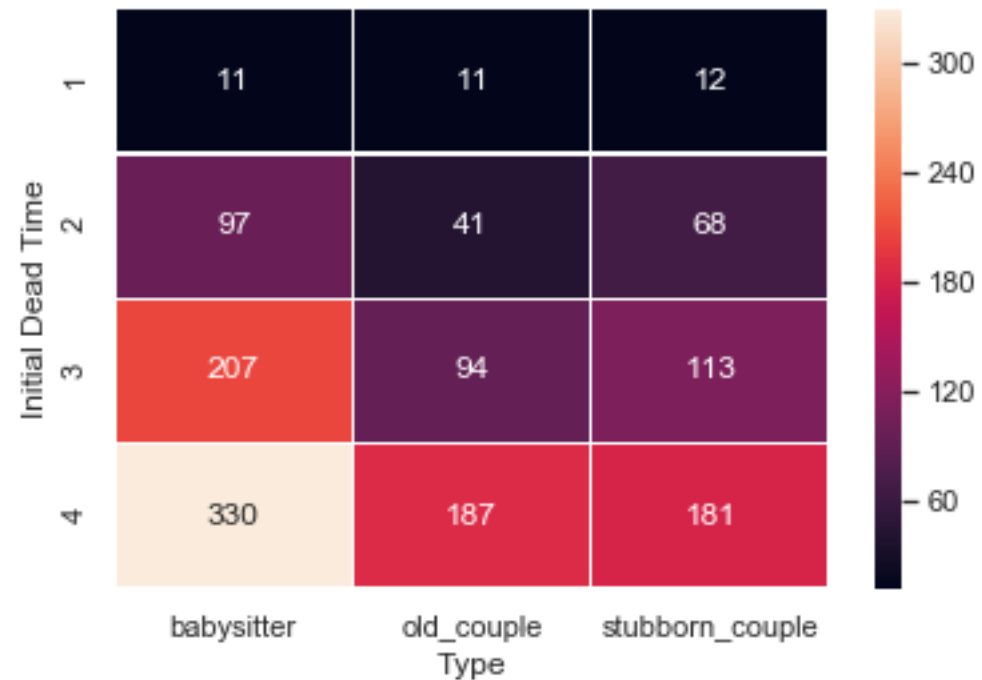
Episode Return is 5000

SARSA

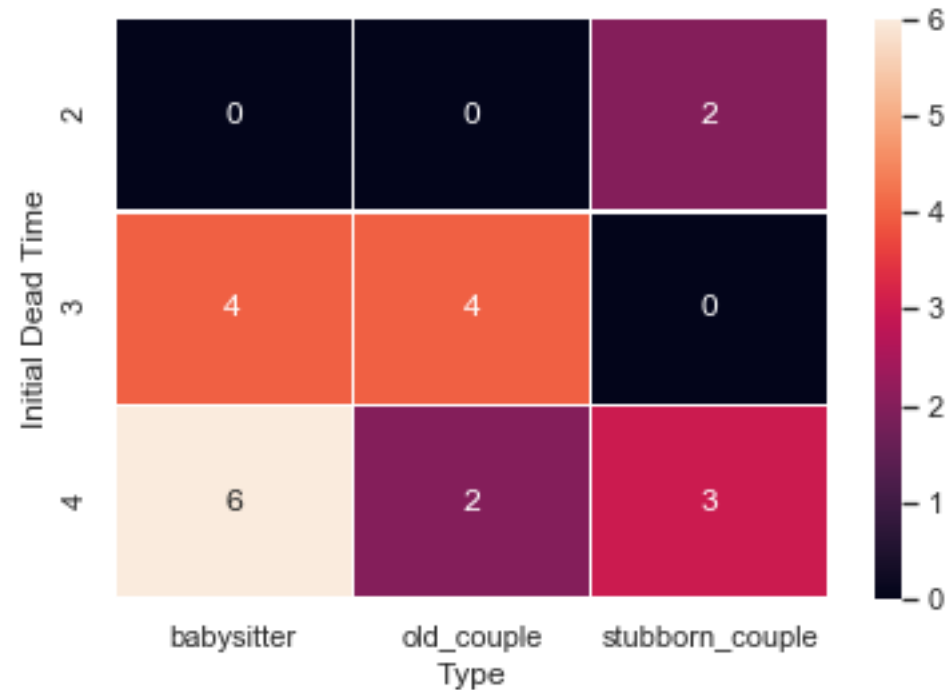- Training curve for different initializations

SARSA

- Training statistics for the number of people saved under different initializations

SARSA

- Testing statistics for the number of people saved under different initializations

SARSA

- Sample trajectories during testing time (more trajectories in the .txt file)

Episode: 16

stubborn_couple

[OP_BUSY=2, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.]

UAV GUIDE [OP_BUSY=1, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.] UAV GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.] VEHICLE GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.] VEHICLE GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.] VEHICLE GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=4, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.] VEHICLE GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.] WARN [OP_BUSY=0, STATUS=WARNING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] UAV GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] INTERRUPT-OPERATOR [OP_BUSY=3, STATUS=ON, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] UAV GUIDE [OP_BUSY=3, STATUS=ON, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] UAV GUIDE [OP_BUSY=0, STATUS=READY_TO_MOVE, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] QUERY-FOR-GUIDE-DETAILS [OP_BUSY=0, STATUS=SAVED, DEAD_TIME=3, GUIDE_TYPE=SELF, NEGOTIATIONS=WARN(s).]
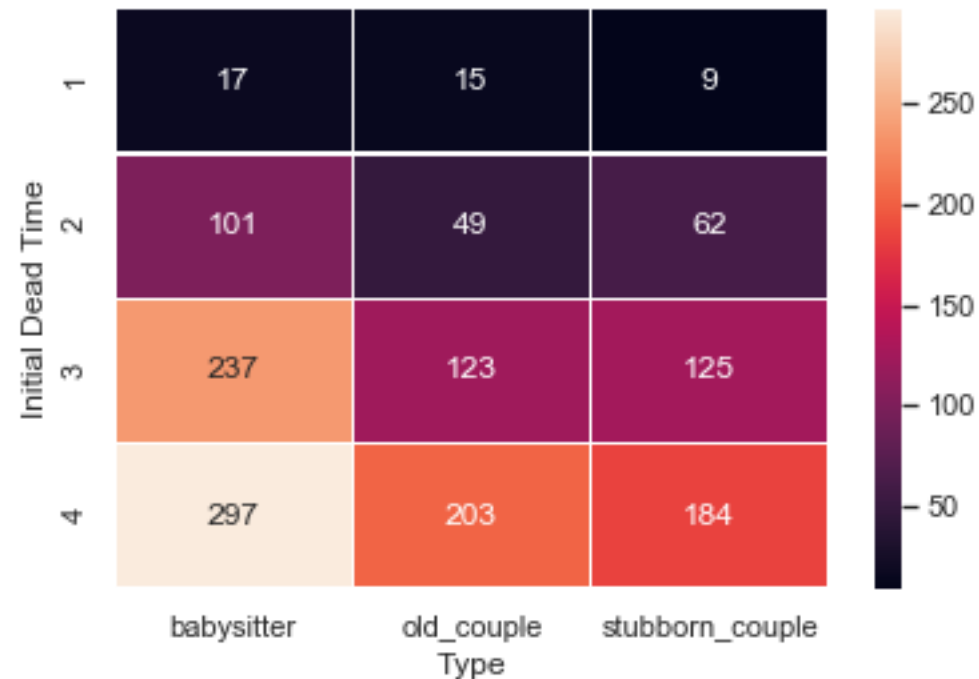
Episode Return is 5000

EXPECTED SARSA

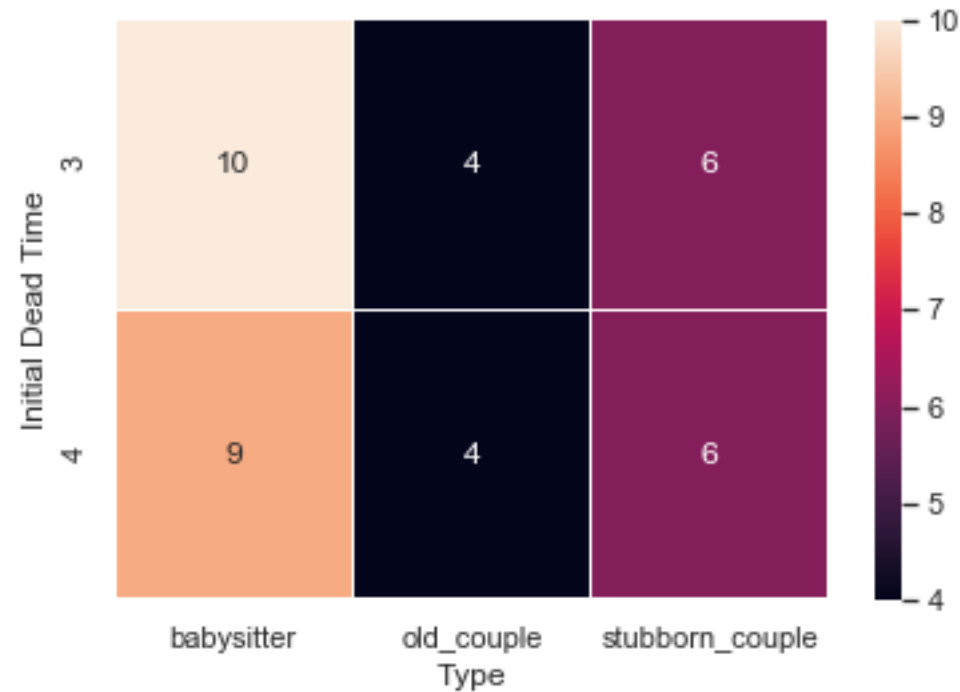- Training curve for different initializations

EXPECTED SARSA

▪ Training statistics for the number of people saved under different initializations

EXPECTED SARSA

- Testing statistics for the number of people saved under different initializations

EXPECTED SARSA

- Sample trajectories during testing time (more trajectories in the .txt file)

Episode: 35

old_couple

[OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=NONE.]

WARN [OP_BUSY=0, STATUS=WARNING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] UAV GUIDE [OP_BUSY=0, STATUS=MONITORING, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=WARN(s).] SP-NEGOTIATE [OP_BUSY=0, STATUS=SPN, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).] WAIT [OP_BUSY=0, STATUS=READY_TO_MOVE, DEAD_TIME=3, GUIDE_TYPE=UNKNOWN, NEGOTIATIONS=SPN(s).] QUERY-FOR-GUIDE-DETAILS [OP_BUSY=0, STATUS=READY_TO_MOVE, DEAD_TIME=2, GUIDE_TYPE=VEHICLE, NEGOTIATIONS=SPN(s).] VEHICLE GUIDE [OP_BUSY=0, STATUS=SAVED, DEAD_TIME=2, GUIDE_TYPE=VEHICLE, NEGOTIATIONS=SPN(s).]

Episode Return is 5000

OBSERVATIONS

- The training curves do not give us a lot of information, (probably the wrong choice of graph from my end) but we can see that the blue and orange lines extend all along the X axis – indicating that people have been killed and saved in early episodes as well as the later ones. Simply put, there is no clear indication of convergence for any of the three algorithms.

- SARSA during test time, was not able to save anyone with a dead time initialization of 1

- Expected SARSA during test time, was not able to save anyone with a dead time initialization of 1 or 2

- The agents seem to have learned when to call the appropriate guide calls (to different extents). They also play them at terrible times, but also make sure that they play them at the right times. They also know when to query for guide information.

- Surprisingly, during a warning or a SP negotiation or a Operator negotiation, the favored actions are one of the guides (that fail) or a query about guiding information (that also fails). They never interrupt a warning with a SPN or an ON etc.

- Definitely can't say that any one algorithm is head and shoulders above the rest but Q-Learning seems to be marginally better in some cases.