MORE RESULTS

SKANDA VAIDYANATH

INSTITUTE OF CREATIVE TECHNOLOGIES

7/1/2019

CHANGES IN THE ENVIRONMENT

- Minor changes in the probabilities of convincing each group with each negotiation strategy
- Variable SPN and ON end time depending on dead time and the nature of the group
- Added Reward shaping for ‘bad moves’ like querying for guide details at the wrong time

MONTE CARLO 200K

- Observations:
  1. Dead time 4 reaches near optimal behavior
  2. Other dead times are making roughly one ‘bad move’ (calling the guide at the wrong time, etc.)
  3. But all dead time are able to save the civilian group consistently.
- Problems:
  1. Could use a little bit more training to iron out the creases for lower dead times.
  2. The ‘optimal behavior’ is to interrupt the operator under all circumstances.

Q-LEARNING 200K

- Observations & Problems:
  1. Needs a lot, lot more training, although some curves are close to reaching optimal behavior.
  2. Lots of problems like playing WAIT as the first move, interrupting itself often, etc. (for the suboptimal cases)
  3. If we do end up running this for 1 million episodes, will we need to change the way we define the learning rate? (1/(1+n))

MONTE CARLO 500K

- Observations & Problems:
  1. For some odd reason, all the training curves look very similar to the 200k episode case (something to do with the exploration rate maybe?).
  2. Hence, the observations and problems are the same as before.

MONTE CARLO 500K HEAVY PENALTY

- In all the previous models, the environment had penalties of 0, -100, -200, -300 for interrupting the operator at op_busy 0, 1, 2, 3 respectively.
- This one has -300, -800, -1300, -1800 respectively for op_busy 0, 1, 2, 3.
- Observations:
    1. Some cases still need more training while the others have reached near optimal behavior.
    2. The Policy no longer just interrupts the operator blindly. So definite improvement here.
- Problems:
    1. Maybe could use a little bit more training but slightly worried about how MC 200k and MC 500k did not have any difference in the previous environment. So may need to play around with the exploration rate decay before running this one for longer.