Enhanced Multiagent Reinforcement Learning for Flight Planning of Aerial Vehicles

Weichang Wang[1] Yongming Liu[2] Lei Ying [3]
ECEE, Electrical Engineering, Arizona State University[1]
School for Engineering of Matter, Transport and Energy, Arizona State University[2]
Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor[3]

Motivations

- The increase of air traffic volume makes the flight planning problem much more complicated.
- Changing environment requires reacting policy
- Efficient Exploration: less data, faster training

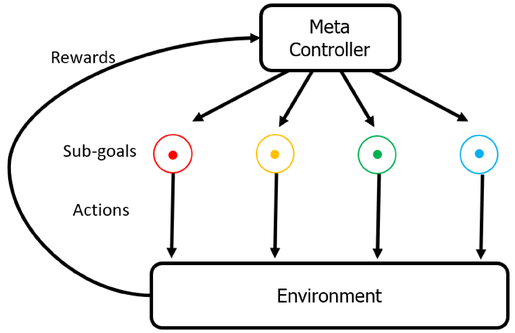
Methodology:

- Hierarchical Training:
  - Meta Controller: assign a task for the vehicles

$$\max_{\pi_g} \mathbb{E}\left[\sum_{t'=t}^{\infty} \gamma^{t'-t} f_{t'}(s,g) | s, \pi_g\right]$$

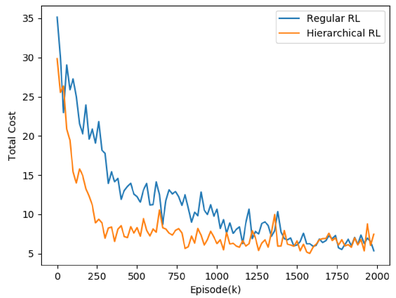
  - Sub Controller: finish assigned task

$$\max_{\pi_{a|g}} \mathbb{E}\left[\sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'}(s,a|g) \ | s, g, \pi_{a|g}\right]$$



Efficient Exploration in Maze:

- Main Problem: Reach destination
- Sub task: Approach next way point



Efficient Exploration, Faster Training

Algorithm 1: Hierarchical Tabular Q Learning

Input: Initialized tabular $Q_{sub}(s,a|g)$ and $Q_{meta}(s,g)$
foreach *Episode* do
  $s \leftarrow$ Start Points, $g \leftarrow epsGreedy(s, \varepsilon_1, Q_{meta})$;
  while *Not Arrived* do
    $R = 0, S_{meta} = s, T = 0$;
    while $s \neq g, d$ *or* $U(0,1) < \beta$ do
      $a \leftarrow epsGreedy(s, \varepsilon_2, Q_{sub,g})$;
      Execute Action a;
      Obtain reward r, next state s' from environment;
      $R \leftarrow R + \gamma^T R$;
      $f \leftarrow -dist(s,g) - \alpha_1 \mathbb{1}\{\text{collision}\}$;
      $Library_g \leftarrow (s, a, f, s')$;
      $s \leftarrow s'$;
      Update $Q_{sub}(s,a|g)$ from $Library_g$
    end
    $S_{meta} \leftarrow s$;
    $g \leftarrow epsGreedy(s, \varepsilon_1, Q_{meta})$;
    $Library_{meta} \leftarrow (S_{meta}, g, R, T, s)$;
    Update $Q_{meta}(s,a|g)$ from $Library_{meta}$;
  end
end

Mean-Field Multiagent Reinforcement Learning in Continuous State Space

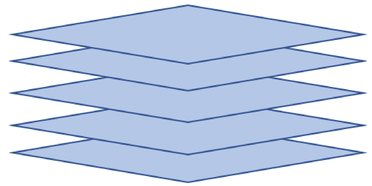
Motivations

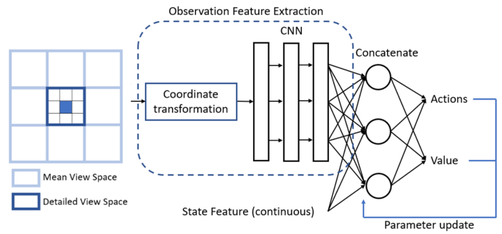
- Discrete space is not able to handle the complex movement of aerial vehicles
- Make the designed flight plan easily implemented.
- Improve the accuracy of the flight plan.

- Advantage:
  - Distributed Algorithms, Centralized Learning with Decentralized Execution
  - Low Complexity
  - Safe Trajectories
- Multi-Resolution Mean-Field
  - Detailed observation on the intruders in close area to the agent, to avoid collision.
  - Mean-field observation only require average number of intruders in further area, to avoid potential congestion.

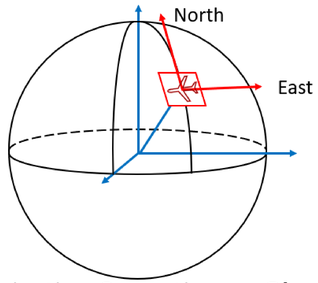
- Implementation in GNATS System
  - Provide a training process for predefined flights
  - Works in cruise phase
  - The longitude and latitude of intruders are corrected on Local tangent plane.
  - The positions updated and decisions made every minute.
  - Control of aircraft: {left-left turn, left turn, remain, right turn, right-right turn}

- Input: Four Layer Matrix
  - 1st layer: mark the number of intruders in the view space
  - 2nd layer: intruder's distance towards destination
  - 3rd layer: intruder's heading direction in the view space
  - 4th layer: intruder's speed in the view space
  - 5th layer: mean view space





Network Structure of MFCA in GNATS



Planar Projection Correction on Observation

Acknowledgments

The research reported in this paper was supported by funds from NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Bajwa, Principal Investigator: Dr. Yongming Liu). The support is gratefully acknowledged.