Autonomous Air Traffic Control: A Deep Multi-Agent Reinforcement Learning Approach

Marc Brittain & Peng Wei
Iowa State University

Introduction

We want to develop an autonomous air traffic control (ATC) system using deep distributed multi-agent reinforcement learning (DD-MARL) that is able to resolve conflicts both at intersections and merging points.

Motivation

- fast growing air traffic complexity in traditional (commercial airliners) and low altitude airspace (drones and eVTOL aircraft)

Challenges

- stochastic environment: aircraft enter the sector at different time intervals
- dynamic environment: there is no fixed number of aircraft in the sector at a given time
- convergence: ensuring that the agents converge to a cooperative policy

Methods

- **A2C + PPO**
  - ➢ incorporate the loss function from PPO to stabilize the learning process
- **Centralized Learning, Decentralized Execution**
  - ➢ train a centralized Actor/Critic and distribute the policy to each agent. Encourages cooperation by improving the joint expected return of all agents
- **Weight Sharing**
  - ➢ Actor/Critic share weights of the same neural network to reduce the total number of trainable parameters

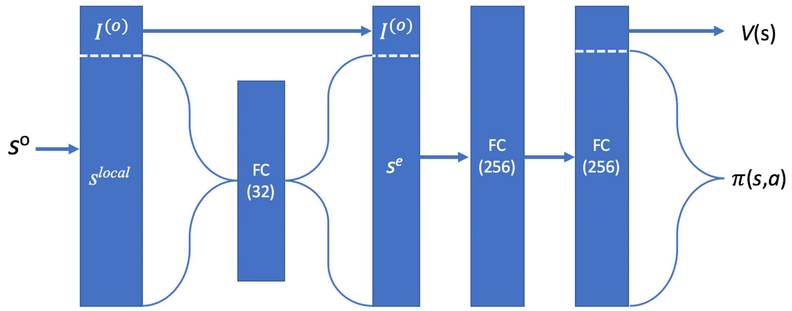


Fig 1. Neural network architecture for A2C with shared layers between the actor and critic.

ATC Environment: BlueSky

- Developed by TU Delft
- Fast-time Air Traffic Control simulator

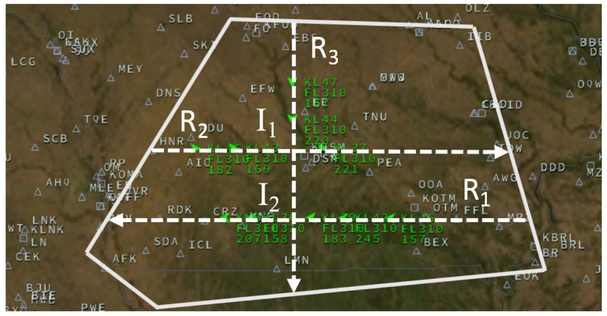


Fig 2. Case study 1: three routes R_1, R_2, and R_3, along with two intersections, I_1 and I_2.

Fig 3. Case study 2: two routes R_1, R_2 merge to a single route at the merging point M_1.

DD-MARL Formulation

- State-Space
  I^i represents position, speed, acceleration, distance to intersection, route ID, and half of the loss of separation (LOS) distance for aircraft *i*

$$s_t^o = (I^o, d^1, \text{LOS}(o,1), I^1, \ldots, d^n, \text{LOS}(o,n), I^n)$$

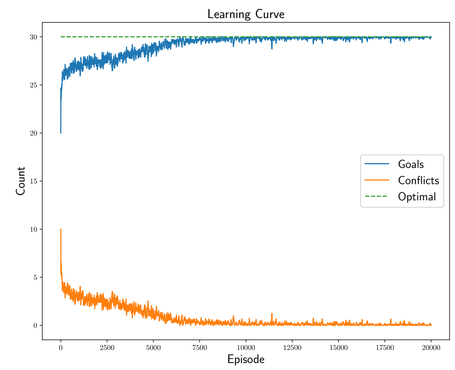
- Action-Space $A_t = [v_{\min}, v_{t-1}, v_{\max}]$
- Reward

$$r_t = \begin{cases} -1 & \text{if } d_o^c < 3 \\ -\alpha + \beta \cdot d_o^c & \text{if } d_o^c < 10 \text{ and } d_o^c \geq 3 \\ 0 & \text{otherwise} \end{cases}$$

Experiment Setup

- Aircraft arrive following a uniform distribution of 4, 5, and 6 minutes
- 30 total aircraft enter the airspace
- Objective is to maintain safe-separation (3 nmi) for all aircraft

Results



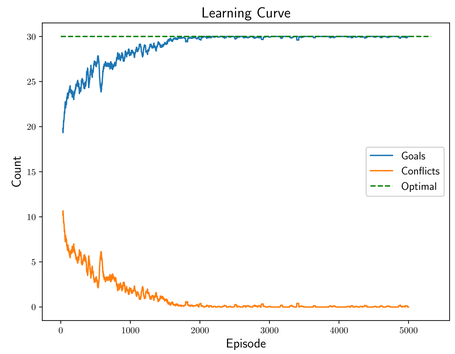


Fig 4. Learning curve for the DD-MARL framework for Case Study 1 (top) and Case Study 2 (bottom). Results are smoothed with a 30 episode rolling average for clarity.

Table 1. Performance of the converged policy tested for 200 episodes.

Case Study	Mean	Median
1	29.99 ± 0.141	30
2	30	30

This research is partially funded by the National Science Foundation under Award No. 1718420, NASA Iowa Space Grant under Award No. NNX16AL88H.

Contact: Marc Brittain (mwb@iastate.edu) **Website**: marcbrittain.github.com