Emerson Hall

Ramana M. Pidaparti

2 September 2025

Machine Learning Model Selection and Design

Introduction

The goal of this week's assignment is to select and design a machine learning model for UAV trajectory optimization. Building on the baseline trajectory planner implemented previously, the ML model should be capable of learning from data and adapting to complex environments that may contain dynamic or unpredictable elements. Unlike deterministic search algorithms such as A*, a learning-based model offers flexibility, scalability and the potential for online adaptation during flight.

Algorithm Selection

Several algorithms were considered before making a final choice. Deep Q Networks are effective when the action space is discrete; for example, when a UAV chooses from a small set of moves such as moving north, south, east or west. Proximal Policy Optimization is a reinforcement learning method that is better suited for continuous action spaces where the UAV must adjust its speed or position in a smooth manner. Long Short-Term Memory networks are a type of recurrent neural network that can capture time dependencies and are useful when past states influence future predictions. For this project, Proximal Policy Optimization was chosen because it works well with continuous control and provides stable learning. Deep Q Networks may be used as a baseline for comparison and Long Short-Term Memory layers may be introduced later if sequential decision making becomes important.

Input Features

The model will take several features as input to represent both the UAV's current state and the environment. These features include the position of the UAV relative to the goal, the velocity components in the x, y and z directions and the distance to obstacles. A simplified local occupancy grid or sensor scan can provide additional awareness of the surrounding space. These features together allow the model to understand both the global direction it should move toward and the local hazards it must avoid.

Output Actions

The outputs of the model will be continuous control signals. The UAV will generate changes in velocity or small waypoint offsets in the x, y and z directions. This approach produces smooth motion and ensure that the UAV can maintain safe and efficient flight paths.

Neural Network Architecture

The network architecture for the Proximal Policy Optimization model includes an input layer that receives the normalized feature vector. This is followed by two dense hidden layers with nonlinear activation functions to capture patterns in the data. An optional Long Short Term Memory layer can be added if temporal information is needed. The network then splits into two parts. The policy head outputs a distribution of possible actions while the value head estimates the quality of the current state. The actor critic structure allows the model to balance exploration of new strategies with exploitation of learned strategies.

Training

Training will require careful design of the reward function. The UAV will receive a positive reward for moving closer to the goal and a negative reward if it collides with an obstacle. Smaller penalties will discourage sharp changes in direction so that the trajectory remains smooth. Proximal Policy Optimization already samples actions from a probability distribution but additional entropy regularization can be added to prevent premature convergence and to encourage broader exploration. Training episodes will be generated using the existing simulation environment with randomized obstacle placements to promote generalization.

Conclusion

Proximal Policy Optimization was selected as the machine learning algorithm for UAV trajectory optimization. The design specifies the input features, output actions, and network architecture which provides a foundation for future implementation. This choice ensures that the model can handle continuous control tasks and produce smooth collision free trajectories. The work completed in week 5 establishes the groundwork for training and evaluating the model in later stages of the project.