Data-driven trajectory imitation with reinforcement learning

Summary

Objective: The objective of this project is to present algorithms for data-driven imitation of trajectories planned or flown, following deep reinforcement learning techniques towards enhancing our trajectory planning abilities.

Challenges:

1. Imitating experts on planning trajectories following a data-driven approach.
2. Learning airspace users' rewards on planning trajectories.
3. Identifying the features that drive planning operations.
4. Produce optimal plans for trajectories.

Approach: Reinforcement Learning, and particularly Q-learning has been studied in the context of predicting trajectories, exploiting historical data about trajectories, enhanced with aircraft intent information. This is a recently-proposed approach whose potential and limitations have been explored by the DART project. However, exploiting aircraft intent has two major shortcomings: One is that it needs a model-based trajectory prediction method in the loop to predict the next aircraft position given a set of commands (requiring at least 500ms at each call to predict), while the other is that combinations of instructions capturing basic commands – if learned independently - may not be flyable, requiring learning "constraints" on the valid commands, or performing joint learning tasks. In addition to these, that approach discretises the continuous state-action space, while it considers as reward the distance to the destination. Building on the knowledge gained from DART, we aim at building a more straightforward approach in which the learning process is (a) an imitation process, where the algorithm tries to imitate "experts" planning trajectories, (b) exploiting raw trajectory data, and (c) it is based on reward models that are learned during imitation. Models learnt by this learning process can then be used for planning and predicting trajectories, although the focus of this project is on planning. Towards this goal we aim to formulate the trajectory imitation problem as a Markov Decision Process in a continuous state-action space and apply deep reinforcement learning (DRL) methods.

Contributions expected:

1. Approaching the trajectory planning problem as an imitation problem, using models learnt from historical data.
2. Identify the features relevant to airspace users for planning trajectories, towards learning their reward models. Tactical interactions and conflicts between trajectories will not be explicitly addressed.
3. Contributions to the ATM Master Plan:
   - Increased abilities for trajectory planning, by means of learning models of trajectories planned and flown by airspace users.
   - Improved operations productivity via contributions to improved trajectory planning.