5-7-2019

Reinforcement Learning in Aviation, Either Unmanned or Manned, with an Injection of AI

Krishnamurthy V. Vemuru

Steven D. Harbour

Jeffery D. Clark

REINFORCEMENT LEARNING IN AVIATION, EITHER UNMANNED OR MANNED, WITH AN INJECTION OF AI

Dr. Krishnamurthy V. Vemuru
Riverside Research
2900 Crystal Dr., Arlington, Virginia 22202
Dr. Steven D. Harbour and Dr. Jeffrey D. Clark
Riverside Research,
2640 Hibiscus Way, Beavercreek, Ohio 45431

We propose a novel theme of aviation with the injection of AI in the form of a reinforcement learning (RL) agent that learns flying skills by observing the pilot's psychological reaction and flight path in a simulator. The pilot and the RL agent learn flying skills simultaneously, forming a symbiotic relationship. The episodes for training the reinforcement learning agent can be simulated by a pilot flying in a simulator, or unmanned using a game on a computer. In a typical episode, the reinforcement learning agent provides a sequence of actions for the pilot to follow. These instructions produce one of the two types of results, either success or failure. The agent observes the psychological reaction of the pilot as well as the flight environment and receives a positive or negative reward. The trained RL agent represents a novel form of AI that assists the pilot for various phases of flight.

Human error is causal to most aircraft accidents; consequently, technologies have emerged to issue alerts when the aircraft's travel trajectory is irregular (Chang et al., (2018)). For example, detecting the aircraft's behavior is one approach to measure the safety of the aircraft. Continuous monitoring and analysis of flight operations is another approach to detect hazardous behavior from a pre-defined list. Li et al., (2016) have reported data mining methods such as cluster analysis of digital flight data using Gaussian Mixture Model (GMM) that are employed by safety analysts to identify unusual data patterns or anomaly detection and latent risks from daily operations. With the advent of Artificial Intelligence (AI), human-autonomy teaming can be an efficient way to minimize human error and further increase aviation safety records. Zhao et al. (2018) have used Reinforcement Learning (RL) as an adaptive online learning model to identify common patterns in flight data and to update the clusters for GMM using recursive expectation-maximization algorithm. The resurgence of interest in AI has attracted applications in aviation systems, in particular, air-traffic management (ATM), air traffic flow management (ATFM) and unmanned aerial systems traffic management (UTM). Kistan et al., (2018) have explored a cognitive human-machine interface (HMI), configured via machine learning, and examined the requirements. They postulated that increased automation and autonomy through AI will lead to certification requirements and discussed how ground-based ATM systems can be accommodated into the existing certification framework for aviation systems. The recent developments in AI open up possibilities in autonomous aviation for introducing a high level of safety by replacing pilot's actions with robotic functions and further research on how AI can be incorporated into autonomous aviation is highly desirable. Our motivation is to show that AI frameworks can be developed by incorporating RL into pilot training simulators.

In this work, we propose a novel theme of aviation with the injection of AI in the form of an RL agent that learns flying skills by observing the pilot's psychological reaction and flight path in a simulator. A unique feature of this AI framework is that the pilot and the RL agent learn flying skills simultaneously, forming a symbiotic relationship. The proposed approach is somewhat similar to how two non-experts comprising of a trainee pilot and an RL agent may learn to play a game by using their joint score as a metric. It is expected that the RL agent will learn a value system i.e., which combinations of states and actions are more rewarding and which ones are not. As the number of game episodes increases, the agent will balance between the exploration of new state-action pairs and the exploitation of known high rewarding state-action pairs until an optimal solution is achieved. RL algorithms usually slow in learning and typically require longer training times, which increase with the size of the state space.

In this context, identifying suitable methods for detecting pilot behavior is the key to developing an AI based on reinforcement learning. Pilot modeling technologies have played a crucial role in manned aviation and control models of human pilot behavior have been developed. Control models are used to analyze the characteristics of the pilot-aircraft system for guidance in the flight control system. Anthropomorphic models of a human operator, which covers the central nervous system, neuromuscular system, visual system and the vestibular system, can represent a pilot's behavior. Recently, Xu et al., (2017) have reviewed control models of human pilot behavior. These models reflect the dynamics of a human sensory and control effectors. AI in the form of computer vision can be coupled with these models to detect non-linear characteristics of human pilot behavior for training the RL agent.

Reinforcement Learning

Reinforcement learning is a type of semi-supervised learning inspired by the way animals learn. It relies on the definition of state space, actions for transitions between states and an associated reward structure in a Markov decision process. In a typical application of RL, an agent makes multiple attempts at a goal and learns from its failures and successes based on a reward structure that has both negative and positive types of rewards. In some of the simple forms of RL, an agent learns the optimal policy by evaluating the value functions $V(s)$ or by $Q(s,a)$ learning, where is s is the current state and a is the action taken in state (s), from episodes, which are attempts by the agent for reaching the goal (Watkins and Dayan (1992). In a game setting, the episodes can be either successful or unsuccessful attempts of playing the game. The game-like situations are realized in many daily life examples, including attempts of a pilot in flying an aircraft.

Flight Simulation Game Framework

Flight simulators are used in pilot training and research on the relationship between emotional intelligence and simulated flight performance to understand how emotional factors affect flight-training performance. Pour et al., (2018) have used a human-robot facial expression reciprocal interaction platform to study social interaction abilities of children with autism.

In this framework, a computer vision system captures the psychological reaction of a pilot undergoing training in a simulator for finding out the result of a pilot's reaction on a flight path following the pilot's operational action. To train the RL agent, we design a flight simulator framework, which is like a game for the pilot to play using his actions, *a* and express his/her gesture, which is representative of the result of his/her actions in the simulator. We represent the gesture, *g* as a two state variable; with values '*happy*' (☺) or '*unhappy*' (☹). The state space of the flight simulator, s consists of five variables: altitude, A, speed, S, heading, H, turn, U and roll, R. Table 1 lists the range of these five action variables.

Table 1.
The five variables that define the state s and their ranges.

State variable	Minimum	Maximum
Altitude, A	0 ft	35,0000 ft
Speed, S	0 mph	550 mph
Heading, H	0˚	360˚
Turn, U	0˚	360˚
Roll, R	0˚	360˚

Flight Path Analysis

A reliable flight path analysis can be obtained by real-time computation of the gradients of state space variables, namely altitude gradient dA/dt, speed gradient dS/dt, heading gradient dH/dt, turn gradient dU/dt and roll gradient dR/dt. A rule-based model compares the gradients with a predefined range to determine, if the maneuver is safe or risky and calculate a dynamic reward. Table 2 shows the gradients and the initial guess values of their ranges. The minimum and maximum of the range can be set as tunable parameters to improve the values iteratively.

Table 2.
Ranges of gradients of the state variables that define the safe operational zone. These ranges will be used in a rule model to dynamically determine the reward for the flight maneuvers.

Gradient of State Variable	Minimum,	Maximum
Altitude gradient, dA/dt	0 ft/s	1,000 ft/s
Speed gradient, dS/dt	0 mph/s	20 mph/s
Heading gradient, dH/dt	0˚/s	3˚/s
Turn gradient, dU/dt	0˚/s	3˚/s
Roll gradient, dR/dt	0˚/s	2˚/s

Pilot's Gesture Assignment

An example computer vision system can consist of a digital video camera, a neural processing unit such as Myriad 2, and a single board computer can be integrated for reading pilot's gesture. The computer vision system can be trained using a face detection machine learning algorithm for real-time monitoring of the 'happy' or 'unhappy' facial expression of the pilot. The agents who play the game of flying the plane in the simulator are particularly advised to show a happy gesture (☺) while their actions result in a safe operation and to show a unhappy gesture (☹) when their actions result in a risky or unsafe or catastrophic operation. The computer vision system can be as simple as a Google AIY kit, which operates with a Tensorflow machine learning model to detect a smile.

Reinforcement Learning Agent – Learning from Pilot's Actions

We first consider a human in the loop approach to develop an RL agent that can use artificial intelligence to determine a human pilot's gesture and calculate rewards. This type of RL agent is trained with the episodes that are generated when a pilot is flying an aircraft in a simulator i.e., on a computer. In a typical episode, the RL agent provides a sequence of actions for the pilot to follow. These instructions produce a result, which is either success or failure. The agent receives two types for rewards: one reward depends on the observation of the psychological reaction of the pilot and the other reward depends on the flight dynamics. The RL agent receives a first reward of +1 when the pilot's gesture is 'happy' or a reward of -1 when the pilot's gesture is 'unhappy.' The RL agent receives a second reward of +1 when the flight state variables and their gradients are in the safe range or a reward of -1 otherwise. The episodes can be used to train the RL agent with different reward structures to select the most suitable reward structure for *Q*-learning. The training process is repeated until convergence of the learning process. After training the agent using a sufficiently large number of episodes, the knowledge acquired by the RL agent is expected to represent a novel form of AI that directs the pilot with accurate instructions for various phases of flight. Fig. 1 shows a learning framework of a RL agent along with its interactions with the flight simulator and the computer vision system that detects the pilot's gesture for receiving rewards to update *Q(s,a)* function and the policy *π(s,a)*.

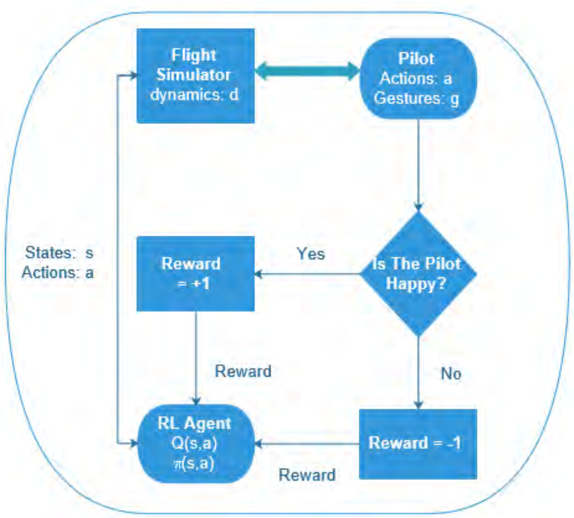

Figure 1.
A framework for the Reinforcement Learning (RL) Agent and its interactions with its environment consisting of the flight simulator and the pilot's gesture recognition system.

Flight Simulator Game Framework

Fig. 2 shows a game framework in a flight simulator for generating the states, actions, rewards, the Q function and the policy. The flight simulator game framework has an additional local reward and a long-term reward compared to the reward structure of the RL agent. The Game RL agent in the flight simulator game framework receives as additional reward of -1 for each instance of state variable's gradient falling outside the safe range. An optional long-term reward of +2 is also awarded to the Game RL agent when the total time taken to reach the destination is below a preset value. The RL agent will receive a reward of +1 when all of the state variable's derivatives are within the safe range. The choice of rewards is arbitrary and can evolve to a more realistic structure based on episodes. A game simulator module initiates the game by extracting actions using the current policy to simulate the flight dynamics. Then, two other modules evaluate the flight dynamics and the gesture of the pilot to identify the rewards. Then, the Q(s,a) function is calculated and updates for each state-action pair and the associated reward. The policy π(s,a) is then recalculated from the Q(s,a) values and updated.

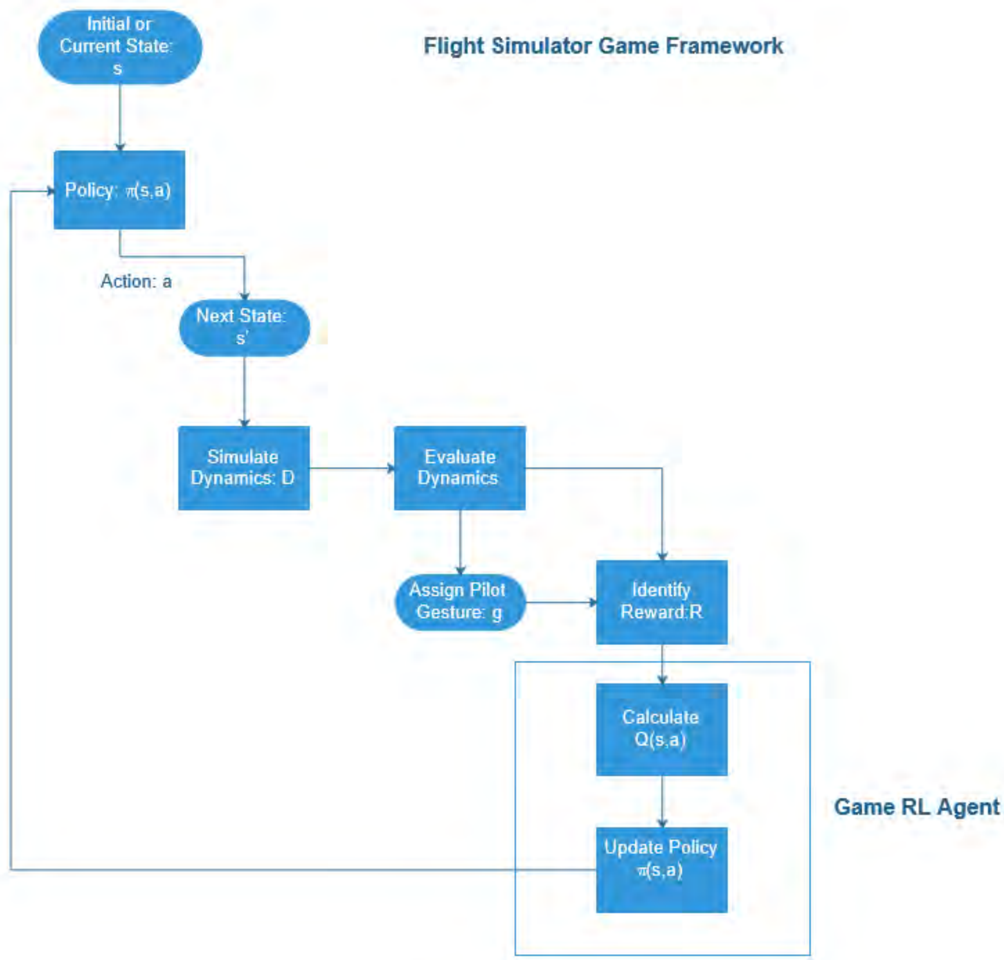

Figure 2.
The framework for flight simulator as a game for obtaining the states, actions, rewards, Q(s,a) function and policy.

Summary

In summary, we have proposed a novel framework of autonomous aviation with the application of artificial intelligence in the form of a reinforcement learning agent which learns flying skills by observing a pilot's psychological reaction and flight path in a flight simulator. The framework consists of a gaming module that works as a flight simulator, a computer vision system that detected pilot's gesture and a flight dynamics analyzer for verifying the safety limits of the state space variables during a simulated flight and a module to calculate the Q-function and the learned policy. With sufficient training within the proposed framework, the RL agent is expected to learn to fly the aircraft as well as to guide the pilot for safe aviation. It would be interesting if present work can attract the attention of game programmers and training tools developers in the AI domain for exploring prototypes based on the proposed frameworks. Finally, an alternate approach to RL is Inverse Reinforcement Learning (IRL) from expert pilot's operations and behavior. This method will require a significant amount of training data in the form of expert pilot's simulator data.

References

Chang, T. H., Hsu, C. S., Wang, C., and Yang, L. –K., (2008). On board measurement and warning module for measurement and irregular behavior, *IEEE Transactions on Intelligent Transportation Systems*, 9(3), 501-513.

Li, L. S., Hansman, R. J., Palacios, R., and Welsch, R., (2016). Anomaly detection via Gaussian mixture model for flight operation and safety monitoring, *Transportation Technologies, Part C: Emerging Technologies*, 64, 45-57.

Zhao, W. Z., He, F., Li, L. S., and Xiao, G., (2018). An adaptive online learning model for flight data cluster analysis, In *Proc. of 2018 IEEE/AIAA 37th Digital Avionics Systems Conference*, *IEEE-AIAA Avionics Systems Conference (*pp.1-7). London, England, UK.

Kistan, T., Gardi, A., and Sabatini, R., (2018). Machine learning and cognitive ergonomics in air traffic management: Recent developments and considerations for certification, *Aerospace*, 5(4), Article Number 103.

Xu, S. T., Tan, W. Q., Efremov, A. V., Sun, L. G., and Qu, X., (2017). Review of control models for human pilot behavior, *Annual Review in Control*, 44, 274-291.

Watkins, C. J. C. H. and Dayan, P., (1992). Q-learning, *Machine Learning*, 8(3-4), 279-292.

Pour, A. G., Taheri, A., Alemi, M., and Meghdari, A., (2018). Human-Robot facial expression reciprocal interaction platform: Case studies on children with autism, *International Journal of Social Robotics*, 10(2), 179-198.