SOREO: A System for Safe and Autonomous Drones Fleet Navigation with Reinforcement Learning

Reda Alami, Hakim Hacid, Lorenzo Bellone, Michał Barciś, Enrico Natalizio

Technology Innovation Institute, 9639 Masdar City, Abu Dhabi, United Arab Emirates
reda.alami@tii.ae, hakim.hacid@tii.ae, lorenzo.bellone@tii.ae, michal.barcis@tii.ae, enrico.natalizio@tii.ae

Abstract

This demonstration introduces SOREO, a system that explores the possibility of extending UAVs autonomy through machine learning. It brings a contribution to the following problem: *Having a fleet of drones and a geographic area, how to learn the shortest paths between any point w.r.t. the base points for optimal and safe package delivery?* Starting from a set of possible actions, a virtual design of a geographic location of interest, e.g., a city, and a reward value, SOREO is capable of learning not only how to prevent collisions with obstacles, e.g., walls and buildings, but also to find the shortest path between any two points, i.e., the base and the target. SOREO exploits based on the Q-learning algorithm.

Introduction

An unmanned aerial vehicle (UAV) is an aircraft without any human pilot, crew, or passengers on board. UAVs are a component of an unmanned aircraft system, which includes adding a ground-based controller and a system of communications with the UAV (Hu et al. 2022). The flight of UAVs may operate under remote control by a human operator, as remotely-piloted aircraft, or with various degrees of autonomy, such as autopilot assistance, up to fully autonomous aircraft that have no provision for human intervention (Hu and Lanzon 2018). UAVs have been under the spotlight recently due mainly to their low cost but also to the capabilities that they may offer in different applications areas. In fact, UAVs allow humans to, e.g., explore remote areas that they could not think of exploring without them. While there is good progress in terms of UAVs automation, this affects mainly sophisticated commercial or military drones. End-users drones remain limited in terms of "autonomous" capabilities, and few can be put in practice.

With the recent developments in artificial intelligence (AI), the opportunity came to add more autonomous capabilities for end-user drones. We explore in this work the possibility of using reinforcement learning for supporting drones autonomous and optimal navigation for various applications such as, e.g., packages delivery. The problem we are addressing is the following: *Having a fleet of drones and a geographic area, how to learn the shortest paths between a base and any other target point for optimal and safe package delivery?*

This paper introduces SOREO, *Safe and autOnomous dRones flEet navigatiOn*, a system designed to control a fleet of autonomous drones combining reinforcement learning with an implementation simplifying the integration of (i) new drones in the fleet and (ii) changes in the ground conditions, e.g., new obstacles. Starting from a set of possible actions and a virtual design of the geographic location of interest, e.g., a city, SOREO is capable of learning not only how to prevent collisions with different obstacles but also to find the shortest path between any two points, i.e., the base and the target. SOREO exploits the Q-learning algorithm (Clifton and Laber 2020), combined with an implementation that supports scaling up to several drones at the same time.

The rest of this paper is organized as follows: in next section we briefly discusses some related work. Then, we discuss the system design and architecture and also the demonstration scenarios. Finally, we conclude and give some future work.

Related Work

With the recent advances in AI, drones became a central focus for different communities (Lee, Mckeever, and Courtney 2021). While most of the efforts target specific tasks (Muñoz et al. 2019), others investigate generic approaches (Hodge, Hawkins, and Alexander 2021) proposing navigation algorithms exploiting data from embarked sensors to guide the drone to the target point. Having UAVs capable of finding the shortest path is an important research topic to be considered as it correlates with the level of autonomy of the UAV and for its resources management (Besada et al. 2019). Furthermore, in modern applications, finding the shortest path implies finding and avoiding obstacles along the way that need to be identified and avoided with no impact on the drones. Indeed, UAVs are deployed in a variety of situations including some (critical) missions in harsh environments, where the obstacles are assumed to be unknown and where no exact formal modeling can be defined a-priori. To achieve the highest autonomy, a good body of research dealing with various reinforcement learning (RL) based algorithms for path planning and obstacles avoidance exists.

From the path planning perspective, (Lin et al. 2020) proposes a deep RL and LSTM based algorithm that allows a UAV to have a continuous interaction with the environment. In (Koch et al. 2019), the authors use classical deep reinforcement learning algorithms such as DDPG, TRPO, and PPO to secure an accurate and stable altitude control. In (Li and Wu 2020), a DDPG algorithm for object avoidance and target tracking via the use of a specific reward function, that allows the learning of actions that achieve a smoother trajectory, is proposed. In relation to obstacle avoidance, existing initiatives use image processing on captured images from drone cameras to avoid collisions, e.g. (San Juan, Santos, and Andújar 2018). (Bouhamed et al. 2020) proposes a DDPG based algorithm to help the UAV reaching the target with zero collisions. For this case, the reward function is designed in order to minimize the distance between the UAV and the target with a penalty for each collision with an obstacle.

Our work intends to complement existing efforts. It differs mainly by the fact that it does not need additional sensors to learn and optimize the paths. Furthermore, this work is a unique combination of several fields, i.e., AI and robotics, showing the good performance of RL in practice. Interestingly enough, it is usually not guaranteed that a UAV realizes transition of states successfully, i.e., diverging to another state. The use of Q-learning shows that it copes properly with these issues.

System Design and Architecture

SOREO is a platform that learns states transitions and avoiding obstacles through RL. To achieve this, SOREO works with two main components: (i) offline and (ii) online. Figure 1 illustrates the architecture of the system. The offline component is mainly responsible for learning the different states transitions through the execution of the Q-learning algorithm.To do so, the system takes three inputs: (1) a floor map of the area of interest, (2) a list of permitted actions, and (3) a reward value. The Q-learning processor generates successful state transitions, i.e., without collisions with other objects such as walls, and stores them in the states transition database. The online component is the one exploiting the results of the offline part and puts them into application. It searches for the best states transition depending on the base and the target points of the drones and executes the navigation operations. The navigation is realized using a conjunction of two controllers. The orientation of the UAV is adjusted using a simple p-controller, which rotates the drone towards the desired point. The speed of the UAV is regulated with a non-linear controller, which decelerates when the UAV is facing away from the goal.

Demonstration Overview[1]

In order to illustrate the capabilities of SOREO, we propose to demonstrate two main scenarios illustrating the learning phase which is performed offline, and the execution of the mission, e.g., package delivery, performed online. The first scenario will demonstrate how the learning process is operated. For this, we provide input to the system consisting of

[1]https://youtu.be/M5vimmkp2JY

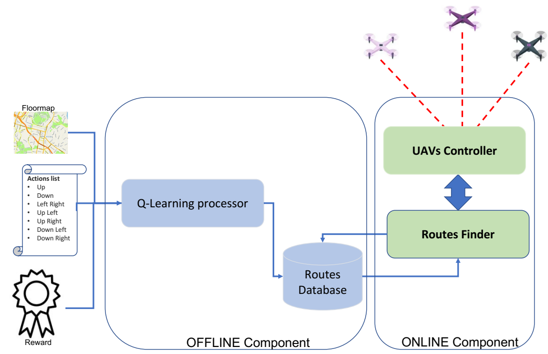

Figure 1: Architecture of SOREO

a virtual floor plan, the set of instructions, and the reward values. The system will run the algorithm and generates the paths. This is visually displayed for a good user-experience. The participants will appreciate the execution time.

The execution of the mission supposes is translated by the exploitation of the learned paths. For this, we vary the complexity of the situations over two main aspects: (i) the number of drones and (ii) the number of obstacles. For the number of drones, we showcase situations with either 2 or 3 drones (and the corresponding number of base and target points, naturally). For the number of obstacles, we use also 3 configurations: 12, 18, and 24 obstacles. Figure 2 illustrates the layout of the current flying area with 18 obstacles, 3 drones, 3 base points, and 3 targets. The situations will demonstrate the speed reaching the target and the selection of the paths depending on the base point. For this, statistics and a tracking visualization are shown through the visual interface.

Conclusion and Future Work

This paper contributes to the problem of equipping UAVs with more autonomy. It introduces the use of Q-learning algorithm to support a system that controls a fleet of drones, illustrated by three drones in this demonstration. As future work, the learning efficiency in terms of learning time should be improved. Also, another track of research is to investigate the case of moving targets.

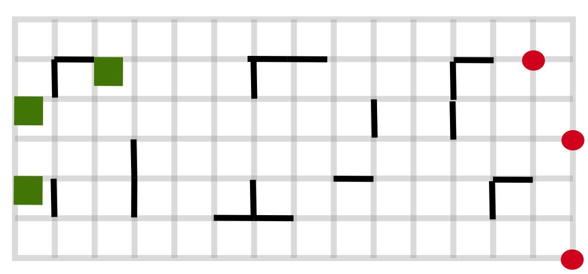

Figure 2: A medium size flying area with 18 obstacles (Black lines: unknown obstacles, Red dots: targets, Green squares: base positions.

References

Besada, J. A.; Bernardos, A. M.; Bergesio, L.; Vaquero, D.; Campaña, I.; and Casar, J. R. 2019. Drones-as-a-service: A management architecture to provide mission planning, resource brokerage and operation support for fleets of drones. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 931–936.

Bouhamed, O.; Ghazzai, H.; Besbes, H.; and Massoud, Y. 2020. Autonomous UAV navigation: A DDPG-based deep reinforcement learning approach. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5. IEEE.

Clifton, J.; and Laber, E. 2020. Q-learning: theory and applications. *Annual Review of Statistics and Its Application*, 7: 279–301.

Hodge, V. J.; Hawkins, R.; and Alexander, R. 2021. Deep reinforcement learning for drone navigation using sensor data. *Neural Computing and Applications*, 33(6): 2015–2033.

Hu, J.; and Lanzon, A. 2018. An innovative tri-rotor drone and associated distributed aerial drone swarm control. *Robotics and Autonomous Systems*, 103: 162–174.

Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; and Arvin, F. 2022. Fault-tolerant cooperative navigation of networked UAV swarms for forest fire monitoring. *Aerospace Science and Technology*, 123: 107494.

Koch, W.; Mancuso, R.; West, R.; and Bestavros, A. 2019. Reinforcement learning for UAV attitude control. *ACM Transactions on Cyber-Physical Systems*, 3(2): 1–21.

Lee, T.; Mckeever, S.; and Courtney, J. 2021. Flying Free: A Research Overview of Deep Learning in Drone Navigation Autonomy. *Drones*, 5: 52.

Li, B.; and Wu, Y. 2020. Path planning for UAV ground target tracking via deep reinforcement learning. *IEEE Access*, 8: 29064–29074.

Lin, Y.; Wang, M.; Zhou, X.; Ding, G.; and Mao, S. 2020. Dynamic spectrum interaction of UAV flight formation communication with priority: A deep reinforcement learning approach. *IEEE Transactions on Cognitive Communications and Networking*, 6(3): 892–903.

Muñoz, G.; Barrado, C.; Çetin, E.; and Salami, E. 2019. Deep Reinforcement Learning for Drone Delivery. *Drones*, 3(3).

San Juan, V.; Santos, M.; and Andújar, J. M. 2018. Intelligent UAV map generation and discrete path planning for search and rescue operations. *Complexity*, 2018.