Automomous Control of Aircraft for Communications and Electronic Warfare: Insights and Promises of Recent Literature

Sean Carver, Ph.D. at Data Machines Corporation

May 25, 2021

Abstract

We pose an unsolved problem in autonomous control of aircraft for communications and jamming (electronic warfare) and review the literature relevant to this problem. Some work offers approximately optimal solutions to related problems in different domains—promising applicability to the important scenarios considered here. Other work covers methods that we may find useful in extending these relevant solutions.

The problem we address lies within the fields of adversarial Multi-Agent Reinforcement Learning (MARL) and active sensing. In our problem, two opposing factions (labeled "blue" and "red") compete to win a zero-sum/purely adversarial game. The blue side tries to maintain communication links between ground-based assets with a fleet of "comms;" whereas the red side tries to jam this network with a fleet of "jammers." An Unmanned Aerial Vehicle (UAV) becomes a comm or a jammer when fitted for one of these purposes.

Each faction lacks knowledge and access to the state of the opposing side, but benefits from inferring this state probabilisticly through positioning its fleet for best sensory performance and localization (active sensing). This maneuvering should take into account the real possibility of any UAV getting shot down by its adversary's appropriately positioned ground troops. That said, the objective of each side should remain foremost. The blue side aims to simultaneously keep all units in communication while the red side aims to simultaneously jam this communication. Despite best efforts, different units/UAVs can fall in and out of communication with their respective headquarters, making each of the blue and red factions a multi-agent collection, fully cooperating among itself, despite different information, to fight its adversary having opposing goals. Our contribution poses this problem while pointing to literature for possible ideas for moving the field forward towards a successful implementation for the full adversarial problem in real-world combat.

1 Introduction

If unfortunate circumstances compel our leaders to order our armed forces to take a city from an adversary, the command headquarters on the ground would benefit from constant two-way communication with all its other units during the conflict.

In the fog that accompanies such struggles, our forces cannot rely on our enemy's network of cell towers to keep in touch. Instead, two way radios, linked by a network of "comms" (UAVs for communication) will hopefully allow our friendlies to stay connected.

While vastly better than cell phones, such a network has it own set of challenges. Indeed, our adversaries clearly prefer to keep us out of communication. To pursue this preference, they may send up jammers (UAVs for blocking communication). Thus begins a delicate dance of each side positioning its fleet to best find the other's birds and in so doing best keep or block communications.

We study the question of how each side can control its fleet by autonomously ordering and carrying out flight and communications- electronics operation instructions (CEOI) to optimally achieve its objectives. We are interested in the strategies for both sides, because to defeat our enemy, we must understand the intelligent countermeasures they may take. Moreover, in a real war, our side—as well as theirs—may choose to fly both comms and jammers, requiring strategies for both roles.

Recent literature has tackled the problem of near-optimal search and rescue [1] and other related search and localization paradigms [2, 3, 4]. Other work has considered different applications requiring similar tools—notably cyber-security [5] and precision farming [6]. Search and rescue, for example, clearly relates to the problem at hand because, as with rescue, each side in our conflict benefits from successfully inferring the positions of targets on the other side. But electronic warfare differs from search and rescue. People being rescued presumably want to be found and will presumably cooperate with this effort. In electronic warfare, on the other hand, targets aim to conceal their true locations from their adversaries. As a result, while search and rescue can succeed with a purely active sensing and optimal control solution, in our scenario, we need to learn to counter an opponent's strategy. To this end, we propose to apply artificial intelligence: specifically, adversarial multi-agent reinforcement learning. This paper reviews the literature relevant to this approach to victory.

2 Optimal and sub-optimal filtering

The filtering problem takes measurements of a stochastic system—possibly transformed measurements and possibly with noise—and produces estimates of the state of the system. This section reviews classical work on this problem, citing textbooks instead of recent papers.

Readers will find the optimal solution to this problem in the first pages of many textbooks on nonlinear filtering [7, 8, 9, 10]. The solution implements a recursion consisting of alternating applications of the Chapman-

Kolmogorov Equation and Bayes Rule.

Unfortunately, solving each of these equations demands an integration remaining provably intractable in most cases—indeed in all but two cases that researches have already identified. In all other cases, a researcher must settle for an approximation—a sub-optimal (but hopefully still *approximately* optimal) filter. Readers will find that the rest of the nonlinear filtering textbooks (the rest beyond the first few pages devoted to the optimal exposition) develop these sub-optimal approximations.

First, we list the two truly optimal solutions to the filtering equation as (1) the Kalman filter, and (2) the finite hidden Markov filter. The Kalman filter uses a linear model of the process, a quadratic objective function measuring optimality, and Gaussian noise corruption (LQG problem). The Gaussian assumption must hold for both the process noise and the measurement noise. The LQG problem divides into the linear quadratic estimator (LQE) problem for the optimal state estimates and the linear quadratic regulator (LQR) for the optimal control of such a system. As we will see below, the solutions of these problems decouple in a surprising and useful way (the separation principle, see the last section below) but unfortunately, this decoupling makes use of the specific assumptions we have imposed here.

Likewise, a finite hidden Markov filter uses a finite-state model of the process. These restrictions unacceptably constrain usable models for our application area, and therefore we will focus on approximately optimal alternatives.

Several specific approximations merit mention. An extended Kalman filter linearizes the state space around each sample point allowing the calculations behind the Kalman filter to proceed approximately, even when the assumptions for using a Kalman filter do not hold exactly. The approximation works well when the optimal probability distributions for state remain close to Gaussian. If they do not remain approximately Gaussian, the Extended Kalman Filter can perform poorly, leading to poor state estimates and impoverished inference.

A second approximation, a grid-based filter, approximates the state space with a finite grid of points allowing the calculations behind a hidden (finite) Markov filter to proceed, even for infinite state spaces. A grid-based filter works well for the lowest dimensional state spaces, but becomes computationally intractable when the dimensions become even slightly higher. In preliminary investigations the electronic warfare problem that motivated this review [Carver, research paper in preparation], one and two targets worked well (each adding two dimensions, longitude and latitude, to the state), whereas three simultaneous targets remained expensive beyond reach. In this work, we aimed to find jammers ("targets") without bearing information from observing successful or unsuccessful radio connections to friendlies.

The field calls the last class of filtering approximations that deserves our attention "particle filters." In short, the idea approximates evolving distributions with a finite swarm of Monte Carlo sample points called particles. These methods possess great generality and flexibility, but many researchers find particle methods more difficult to understand, and to successfully implement, than their simpler and more straightforward cousins.

Note that there exist many different ways to implement particle filters, each with its own benefits and limitations. We will discuss these methods further in the next section, as several papers concerning Active Sensing use particle filters.

3 Active sensing

Active sensing solves a control problem, and as such has considerable overlap with reinforcement learning. Both use observations to select actions with some notion of how to make that selection. Whereas reinforement learning tries to optimize cumulative reward, active sensing chooses its action (they call it control) to maximize information or, equivalently, minimize entropy. For example, in active sensing for search and rescue, we want to control the sensors (choose the action) to best reduce the uncertainty in the target locations.

The electronic warfare application that we consider does not have this form. We aim to keep or block communication between units landing this problem squarely in the purvue of reinforcement learning. That said, the objective benefits from knowing, even with uncertainty, the locations of other side's birds. Moreover, like with active sensing, each side can move its own birds to localize the others'. Therefore, we look to the literature on active sensing for inspriation.

4 Reinforcement learning and its extensions

This section spans several disciplines, including reinforcement learning, deep reinforcement learning, distributional reinforcement learning, Bayesian reinforcement learning, and multi-agent reinforcement learning (see citations below).

Let us start by defining the terms above. Reinforcement learning (RL) [11, 12] extends machine learning to sequential problems where an agent or agents learn to interact with an environment to maximize cumulative reward.

RL shares a sizable overlap with control theory in engineering, particularly adaptive control [13, 14]. The terminology in engineering differs from the terminology in machine learning, but the terms map to each other. The "controller" (agent) interacts with "the plant" (the environment) by selecting a "control" (an action) that "minimizes cost" (maximizes reward). Engineers typically deal with continuous systems (ie robotics), whereas many, but not all, reinforcement solutions have finite action and state spaces. Adaptive control, moreover, aims more to maintain control when the plant changes, rather than to learn to control the plant *de novo*. Despite cosmetic differences, clearly if both systems solve the same problem, and to the extent that they solve it optimally, with the same optimality criteria, the solutions must coincide. But the approximations to optimality made in each discipline may differ, of course. We expect to

find the same pattern when we fold active sensing in with reinforcement learning.

Deep RL [15] uses neural networks to represent the functions learned by the agent(s). Classically, RL implementations deal with inevitable uncertainty in represented quantities by maintaining best point estimates for these quantities. Distributional RL [16] departs from this tradition by maintaining full probability distributions for the uncertain quantities. If the actor(s) also perform Bayesian inference on these distributions (as they generally do), the actor implements Bayesian RL [17]. Finally multi-agent RL [18] extends RL to environments that include other interacting agents cooperating or competing for reward.

A lot of good work exists in classical RL and classical deep RL, but we will not review any papers in these fields. Instead, we would like to direct the reader's attention to distributional RL, its slightly smaller subset, Bayesian RL. Consider that the target localization and active sensing literature of interest represent uncertain quantities with distributions, just like distributional RL.

While an RL algorithm could derive point state estimates from distributional state information to decide on an action, such an approach seems wasteful. Achieving an optimal solution to the electronic warfare problem stands as a worthy ambition for a machine learning engineer. That said, to not using the full distribution returned by filtering amounts to giving up on this ambition, or so it seems.

However, the mathematics can sometimes work out so that throwing away the distribution for a point estimate succeeds as the optimal solution [19]. For example, to control a linear regulator, where a quadratic cost function determines optimality, corrupted by Gaussian noise, (the LQR problem) the optimal solution uses a Kalman filter to produce optimal state estimates, then applies the optimal control for a determinisitic regulator with known state equal to those optimal state estimates. In this case, throwing away the distribution allowed a simpler but just as good—indeed optimal—solution.

A system satisfies the so-called *separation principle* if such a statement holds for the system. But systems do not always satisfy the separation principle—those that do stand as the exceptions. If an analyst can get away with invoking the separation principle, the calculations greatly simplify. Indeed, for the LQR problem, the solution exists in closed form.

We do not know how well the separation principle applies in our problem and we do not know how computationally expensive it will become to use purely distributional methods. The approach we suggest applies distributional RL (or more precisely Bayesian RL) to simple but related problems first then push the envelope both in terms of the complexity of the scenario, and in terms of the approximations used, such as the separation principle assuming it applies approximately.

In the LQR problem we have a precise notion of optimality (the quadratic cost function) and a proof that the solution presented optimizes this conditon. In our scenario we have neither. We can say, however, that we will apply principles, such as Bayes Rule, that researchers in other contexts have shown lead to provably optimal solutions. We can say that "optimal" principles underlie our methods, rather than that they achieve

optimality. Approaching the problem in this way, we hope that our solutions will stand as very good, indeed good enough, and perhaps the very best.

Finally, a multi-agent enviroment such as the one needed for electronic warfare consists of actors other than self making decisions based on other hidden information. This becomes an additional wrinkle which we must deal with. Reinforcement learning algorithms divide into *model-based* algorithms and *model-free* algorithms. The two methods differ in that a model-based algorithm maintains a model of the environment and uses this model to select its actions.

Model-based reinforcement learning would fail catastrophically with our scenario in the real world. The model in the model-based algorithm would have to include models of all other agents, including those of our adversary. But we must expect that our adversaries will exploit any modeling assumptions we make—and posing a model of our adversaries requires such assumptions.

The model-free alternative presents itself as an option. A recent comprehensive survey of BRL [17] discusses and develops only two such classes of algorithms in that context: *Bayesian Policy Gradient Algorithms* [20, 21] and *Bayesian Actor-Critic Algorithms* [22, 21]. These observations together with our preference for BRL, narrow considerably the initially wide field of RL algorithms of interest. That said, it seems that just one set of authors has produced most of the work on model-free Bayesian RL, with the rest of the community slow to adopt their ideas. This observation gives us cause for concern, but no cause to dismiss their work without further investigation. The path forward appears clear: test BRL with simple models, then push the envelope on the complexity of the models, finally examine the performance/computational cost trade offs with less expensive suboptimal techniques.

References

[1] G. M. Hoffmann and C. J. Tomlin, "Mobile sensor network control using mutual information methods and particle filters," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 32–47, 2009.

[2] A. Ryan and J. K. Hedrick, "Particle filter based information-theoretic active sensing," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 574–584, 2010.

[3] J. Tisdale, A. Ryan, Z. Kim, D. Tornqvist, and J. K. Hedrick, "A multiple UAV system for vision-based search and localization," in *2008 American Control Conference*, pp. 1985–1990, IEEE, 2008.

[4] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 425–437, 2002.

[5] D. Nicholson, S. D. Ramchurn, and A. Rogers, "Information-based control of decentralised sensor networks," in *Defense Industry Appli-*

cations of Autonomous Agents and Multi-Agent Systems, pp. 15–32, Springer, 2007.

[6] E. Testi, E. Favarelli, and A. Giorgetti, "Reinforcement learning for connected autonomous vehicle localization via UAVs," in *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*, pp. 13–17, IEEE, 2020.

[7] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications*. Artech house, 2003.

[8] D. Crisan and B. Rozovskii, *The Oxford handbook of nonlinear filtering*. Oxford University Press, 2011.

[9] A. Smith, *Sequential Monte Carlo methods in practice*. Springer Science & Business Media, 2013.

[10] N. Chopin and O. Papaspiliopoulos, *An introduction to sequential Monte Carlo*. Springer, 2020.

[11] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[12] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.

[13] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2013.

[14] S. G. Khan, G. Herrmann, F. L. Lewis, T. Pipe, and C. Melhuish, "Reinforcement learning and optimal adaptive control: An overview and implementation examples," *Annual reviews in control*, vol. 36, no. 1, pp. 42–59, 2012.

[15] Y. Li, "Deep reinforcement learning: An overview," *arXiv preprint arXiv:1701.07274*, 2017.

[16] I. Osband, J. Aslanides, and A. Cassirer, "Randomized prior functions for deep reinforcement learning," *arXiv preprint arXiv:1806.03335*, 2018.

[17] M. Ghavamzadeh, S. Mannor, J. Pineau, and A. Tamar, "Bayesian reinforcement learning: A survey," *arXiv preprint arXiv:1609.04436*, 2016.

[18] L. Buşoniu, R. Babuška, and B. De Schutter, "Multi-agent reinforcement learning: An overview," *Innovations in multi-agent systems and applications-1*, pp. 183–221, 2010.

[19] K. J. Åström, *Introduction to stochastic control theory*. Courier Corporation, 2012.

[20] Y. Engel and M. Ghavamzadeh, "Bayesian policy gradient algorithms," *Advances in neural information processing systems*, vol. 19, p. 457, 2007.

[21] M. Ghavamzadeh, Y. Engel, and M. Valko, "Bayesian policy gradient and actor-critic algorithms," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2319–2371, 2016.

[22] M. Ghavamzadeh and Y. Engel, "Bayesian actor-critic algorithms," in *Proceedings of the 24th international conference on Machine learning*, pp. 297–304, 2007.