Large Project Proposal: Spectrum Utilization Through Reinforcement Learning

By Shane Flandermeyer and Geoffrey Dolinger

Why the topic is interesting:

Shane: Cognitive radar is a rapidly growing area of research in the radio frequency (RF) community. I have spent much of my time as an undergraduate researcher implementing radar signal processing algorithms on software-defined radio (SDR) systems, which have been a key enabler in the evolution of cognitive radar due to their flexibility and ability to adapt in real-time. Many adaptive waveform selection and resource management algorithms I have studied (e.g., sense-and-avoid) adapt to the system's surroundings but lack a learning component, resulting in degraded performance when they fail to recognize patterns in the data. Reinforcement learning techniques like Deep Q-Networks and actor-critic networks are promising alternatives to traditional adaptive algorithms that I would like to explore in this project.

Geoff: During my career with the Airforce, I spent the last 4 years as a section chief for an innovation group that focused on Command and Control (C2) applications. This domain has experienced major challenges in growing complexity of information gathering, data processing, abstraction of data into usable forms, and effective decision recommendations and/or actions. I completed my Masters in 2012 with a focus in neural networks as applied to control systems and have followed the field during my career. Through my studies I believe that the C2 domain would benefit immensely from cognitive radar and explainability. I have a personal interest in all aspects of the cognitive radar problem but specifically the exploration of Reinforcement Learning to train deep intelligent radar agents to improve performance and predictions given contested and noisy environments.

Questions/problems this project addresses:

Wireless devices utilize the electromagnetic spectrum to function, and the rise of commercial telecommunications technology such as the 4G/5G and the internet of things has caused the spectrum to become overly crowded. In most cases, these commercial devices are primary users of the spectrum, and secondary users such as radar can only access the spectrum if they do not interfere with other users. It has become increasingly necessary to develop cognitive radar systems which adapt their transmitted waveform to "conform" with their surroundings. At the same time, the development of multi-function antenna arrays has made it possible to perform multiple radar tasks on a single device. However, the resources needed to complete these tasks (e.g., computing power, antenna beams, time) are limited, so efficient resource management is required to fully utilize the available hardware.

How we will address the problem:

We plan to use reinforcement learning techniques to train a cognitive agent that optimizes its waveform on transmit to address the spectrum sharing problem. Our primary goal is to model this agent using policy iteration, which is a Markov Decision Process (MDP)-based technique. We will implement this agent in a simplified RF channel environment using Keras with several different types of interference. As a stretch goal, we will implement the Actor-Critic network developed in [6] and compare it to the MDP approach. These approaches are described in more detail below.

Project Activities:

- Primary goals:
    - Develop a simplified time-frequency state space representing the RF channel. The channel will be divided into a small number of frequency bins, where interference is either present or absent. This space must be relatively small since non-deep RL methods scale poorly with increasing numbers of states.
    - Develop simple interference models that the RL methods will use for training and testing. In the simplest case, this will include a stationary interference source that occupies a static frequency band while turning on and off at a constant rate. This behavior is analogous to a non-adaptive radar. The other baseline interference case we will consider is a waveform that sweeps through the frequency range one bin at a time, reversing direction when it reaches either end of the spectrum (also known as a triangle-sweep waveform).
    - Implement the technique described in [1], which treats the environment as a fully observable MDP and determines its bandwidth and center frequency based on an RL reward structure. Once implemented, we will experiment with different reward structures and metrics (e.g., punishing collisions rather than rewarding SINR). We will compare the MDP to two baselines: a radar occupying the full bandwidth and a Markovian radar that transitions to a new band with some user-defined probability.
    - We will evaluate the MDP itself using statistics such as the mean and variance of the reward, along with the average SINR and bandwidth utilized. We will account for multiple training sessions with our system and record our results to ensure consistency. Once we have settled on reward structure, we will compare results across the different explored architectures.
- Stretch goals:

- Implement an Actor Critic technique as described in [6] and evaluate the radar spectrum sharing performance. This reference utilizes the Matlab Reinforcement Learning toolkit, but a Keras based tutorial is provided at [7] to support implementation and testing of the stretch goal.
- Develop more complex interference models. The interference model presented above is not representative of most real radar scenarios, where interference can take on a continuum of values and is not merely present or absent. We could obtain more practical results with a realistic model of the environment, considering RF propagation effects such as power predicted by the radar range equation, multipath, and clutter. We could also model interference that is also a cognitive agent, which has not been explored extensively in the literature and could lead to some interesting game theoretic analysis.

Suggested changes and how we addressed them:

1. As you move forward, you'll want to firm up which RL techniques you'll use and which libraries you'll use to implement and test them.

- We had not chosen between Keras and the Matlab RL toolbox in our draft proposal, but we have now decided on Keras.
- Decided to use MDP (policy iteration) and potentially Actor Critic techniques

2. Consider thinking of this project in terms of primary goals vs stretch goals

- We completely rewrote the activities section and subdivided it into primary and stretch goals.

3. Consider statistical comparisons to be made between methods

- Added a portion to the MDP goal section describing the statistics used for evaluation.

4. Also consider possible baselines to which RL could be compared

- Added full-band and random action baseline comparisons to the MDP section. In our stretch goals, MDP is itself a baseline for the Actor-Critic approach.

Topic Paragraphs:

Note (Shane): The paragraph below is slightly modified from my initial submission. Here, I am focusing on reinforcement learning rather than deep learning.

Shane: Traditional adaptive radio frequency (RF) systems perform signal processing at the receiver to improve performance and increase efficiency, taking advantage of their surroundings without altering them. Cognitive RF systems, on the other hand, use information from the environment to optimize both receiver processing and the parameters of the transmitted signal/waveform. Thus, cognitive systems operate based on the perception-action cycle of cognition in which the system collects information about its surroundings (perception), then tailors its transmitted waveform to fit the needs of its mission (action). Since the transmission has an impact on the system's surroundings, the perception-action process is repeated, and a feedback loop is formed between the system and its environment. This framework raises several open research questions on how the system should utilize the information it collects, and which parts of the radio processing architecture should be optimized using cognition. Reinforcement learning has been introduced as a candidate solution to these problems, and deep reinforcement learning algorithms have been successfully used to improve the waveform design process, dynamically manage RF system resources, and share the electromagnetic spectrum between competing devices. I plan to explore existing reinforcement learning approaches that have been used to solve problems involving cognitive radio.

Geoff: The Department of Defense and specifically the US Air Force has had an extreme growth in challenges related to the domain of Command and Control (C2). C2 experiences major challenges in growing complexity of information gathering, data processing, abstraction of data into usable forms, and effective decision recommendations and/or actions. I intend to investigate, research, and implement methods for radar cognition to address some of these challenges. Radar Cognition is separated into 3 components: a) radar systems perceiving their environment, b) radar systems capable of taking actions to improve their perception or influence the environment, and c) radar systems action decisions based on their perceptions. Machine Learning (ML) serves as a promising set of tools to address multiple aspects of the radar cognition problem. I intend to explore existing machine learning methods applied to radar cognition as well as ML techniques that show promise for radar cognition problems. The primary topic I plan to utilize spatiotemporal ML analysis of radar data (e.g., CNN/RNN methods) in connection with radar actions (e.g., radar system controls) to train learning agents to improve detection or tracking of targets. Primarily, I will explore deep learning methods to train cognitive agents such as reinforcement learning and artificial evolution.

References:

[1] E. Selvi, R. M. Buehrer, A. Martone and K. Sherbondy, "Reinforcement Learning for Adaptable Bandwidth Tracking Radars," in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 5, pp. 3904-3921, Oct. 2020, doi: 10.1109/TAES.2020.2987443.

[2] C. E. Thornton, M. A. Kozy, R. M. Buehrer, A. F. Martone and K. D. Sherbondy, "Deep Reinforcement Learning Control for Radar Detection and Tracking in Congested Spectral Environments," in IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 4, pp. 1335-1349, Dec. 2020, doi: 10.1109/TCCN.2020.3019605.

[3] M. Kozy, J. Yu, R. M. Buehrer, A. Martone and K. Sherbondy, "Applying Deep-Q Networks to Target Tracking to Improve Cognitive Radar," *2019 IEEE Radar Conference (RadarConf)*, 2019, pp. 1-6, doi: 10.1109/RADAR.2019.8835780.

[4] C. E. Thornton, R. M. Buehrer, A. F. Martone and K. D. Sherbondy, "Experimental Analysis of Reinforcement Learning Techniques for Spectrum Sharing Radar," 2020 IEEE International Radar Conference (RADAR), 2020, pp. 67-72, doi: 10.1109/RADAR42522.2020.9114698.

[5] S. Ak and S. Brüggenwirth, "Avoiding Jammers: A Reinforcement Learning Approach," *2020 IEEE International Radar Conference (RADAR)*, 2020, pp. 321-326, doi: 10.1109/RADAR42522.2020.9114797.

[6] M. Altmann, P. Ott, N. C. Stache, D. Kozlov and C. Waldschmidt, "A Cognitive FMCW Radar to Minimize a Sequence of Range-Doppler Measurements," *2020 17th European Radar Conference (EuRAD)*, 2021, pp. 226-229, doi: 10.1109/EuRAD48048.2021.00065.

[7] Balsys, Rokas “Asynchronous Advantage Actor-Critic (A3C) algorithm” https://pylessons.com/A3C-reinforcement-learning , 2020.