Research Paper Presentation

Umesh Kalvakuntla

CS20BTECH11024

Title

Avoiding Jammers: A Reinforcement Learning Approach

Authors

- Serkan Ak
- Stefan Bruggenwirth

Date of Conference

28-30 April, 2020

Avoiding Jammers: A Reinforcement Learning Approach

Abstract

1. This paper investigates the anti-jamming performance of a cognitive radar under a partially observable Markov decision process (POMDP) model.
2. First, we obtain an explicit expression for the uncertainty of jammer dynamics, which enables us to discover new insights into the performance metric of the probability of being jammed for the radar beyond a conventional signal-to-noise ratio (SNR) based analysis.
3. Considering two frequency hopping strategies, this performance metric is analyzed with deep Q-network(DQN) and long short term memory(LSTM) networks under various uncertainty values.
4. Simulation results show that this operator improves upon the performance of the traditional target network.

Definitions

Radio Detection and Ranging(RADAR)

Radar is a detection system that uses radio waves to determine the distance (range), angle, or velocity of objects. It can be used to detect aircraft, ships, spacecraft, guided missiles, motor vehicles, weather formations, and terrain.

Cognitive RADAR

Cognitive radars are systems based on the perception-action cycle of cognition that sense the environment, learn from it relevant information about the target and the background, then adapt the radar sensor to optimally satisfy the needs of their mission according to a desired goal.

Markov Decision Process

Markov decision process provides a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker.

Introduction

- A cognitive radar envisioned as an intelligent radar can optimize its operational parameters with respect to data gathered through a feedback loop as a result of the interaction with the surrounding environment.
- Deep reinforcement learning (DRL) has recently been used for developing frequency hopping strategies.
- In this paper, we focus on the probability of being jammed performance of a radar under two different strategies based on RL algorithm, named as KARAA strategy and LARA strategy.
- Utilizing Bellman's optimality in DRL, we show that proposed strategies are considerably better than a purely random hopping strategy in terms of the probability of being jammed.
- We replace the target network with a softmax operator.

System Model

Jammer Dynamics

- We consider a jammer and radar both of which operate in the same set of N channels.
- At each time slot, the radar desires to intelligently select an unoccupied channel to transmit successfully.
- The signal model of the system is comprised of two phases: (1). training phase, (2). implementation phase.
- In the training phase, the radar does not transmit any signal but observes the all channels to collect data, which will be used to train the employed neural network via the state-action value function.
- The received signal in channel k at each time slot at the radar can be written as

$$y_k = \begin{cases} w_k & \text{under } \mathcal{H}_0 \\ g_k w_k + x_k & \text{under } \mathcal{H}_1 \end{cases}$$

System Model

where $k \in \{1, 2, ...\mathrm{N}\}$ and $x_k \in \mathrm{C}$

- The implementation phase consists of two steps, i.e., theinitial step and the operation step, respectively.
- The first step is used to determine the channel occupied by the jammer at time slot t and utilizes this information to take an action in the next time slot.
- In the latter step, according to the chosen strategy, the radar transmits at time slot $t+1$ and at all future time slots.
- In this step, we assume that the jammer and radar transmit simultaneously at the beginning of each time slot.
- This assumption takes cares of not letting either the radar or the jammer take advantage of observing the transmission of the other in the same time slot.

Problem Formulation

1. **Uncertainity of Jammer Dynamics** are studied under two assumptions in this paper.
   - **Assumption 1 :** For each state $s_i \in \mathrm{S}, s_j \in \mathrm{S}$ can be reached in one step for s_i i.e., $p_{ij} > 0$ and $i, j \in \{1, 2, ...\mathrm{N}\}$.
     The jammer should be able to hop to any one of the available channels in the next time slot for not leaving a safe channel for the radar.
   - **Assumption 2 :** Each state $s_k \in \mathrm{S}$ is associated with a distinct label.
     We will consider orthogonal channels for frequency bands.
2. **Partially Observable Markov Decision Process :**
   We model the received signal in as a discounted POMDP defined by the tuple $(\mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma)$ where $\mathcal{S}$ is state, $\mathcal{O}$ is noisy observation in channel k at time slot t, $\mathcal{A}$ is action of the radar, $\mathcal{R}$ is reward function.
3. **Reinforcement Learning Algorithm :**
   - Jammer Dynamics is such that function that takes the noisy observation for channel k at time slot t as an input and determines whether channel k is occupied or not as an output.

Problem Formulation

- The output of above function is unity if channel k is occupied at time slot t; otherwise it is 0.
- In the training phase, the radar collects the sample at each time slot from the environment and stores them into the replay memory M.
- The state-action value function $\mathrm{Q}^{\pi} : \mathcal{S} \times \mathcal{A} \rightarrow \mathrm{R}$ is defined as the aggregated is counted reward obtained by when policy π is used to take actions, that is, where E is the expectation operator.
- The objective of reinforcement learning algorithm is to find the optimal policy π^*,which obtains the largest aggregate discounted reward, that is,

$$\pi^*(s_t) = \underset{\pi}{sup}\, Q^{\pi}(s_t, a_t) \tag{1}$$

Strategies for avoiding Jammers

Two strategies employed by the RADAR to avoid jamming are

(1) KARAA Strategy - $\pi_K(s)$

- The radar searches for the most frequent hopping sequence of the jammer by the RL algorithm.
- It avoids this sequence in a random fashion; thus, it may reduce probability of being jammed, effectively.
- The characteristic of actions taken in the first step is deterministic.
- In the following step, the radar generates a random sequence of actions $\pi_K(s)$.
- Note that it is not necessary to satisfy unifilar property in the second step since the uncertainty of the jammer's Markov source is independent of $\pi_K(s)$.

Strategies for avoiding Jammers

(2) LARA Strategy - $\pi_L(s)$

- The intuition behind this strategy is to exploit the fact that there is at least one hopping sequence of the jammer, which yields the smallest aggregate discounted reward under Bellman's optimality.
- Formally, given the reward function in (7), the least frequent hopping sequence of the jammer corresponds to the optimal policy $\pi_L(s)$ under Bellman's optimality.

$$\pi_L(s) = \underset{\pi}{inf}\ Q^{\pi}(s_t, a_t) \tag{2}$$

where the infimum operator takes all policies into account.

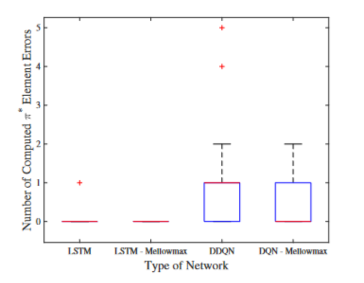
Strategies for avoiding Jammers

Advantages of LARA strategy over KARAA strategy

1. In the case of full observation, i.e.,y_k without noise, it yields the optimum result for the probability of being jammer performance under the Bellman's optimality.
2. Taking actions in a random fashion is not necessary. In fact, this would lead to a sub-optimal result.

Simulation Results

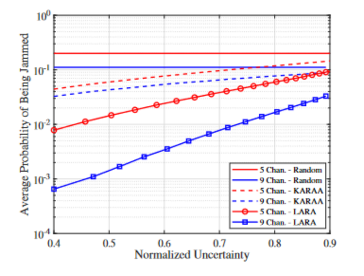
Compound π^* elements in error for normalized uncertainty of values 0.85.



Figure

Simulation Results

Average probability of being jammed for a variety of values of normalized uncertainty. In the training,SNR ranges between 5dB and 10dB.



Figure

Conclusion

- In this paper, we have studied the probability of being jammed performance of a radar under a POMDP model.
- Inspired by Shannon's landmark paper, we have proposed a novel approach to analyze the probability of being jammed in terms of the extent of uncertainty of jammer dynamics under two specific strategies - KARAA strategy and LARA strategy.
- Beyond a traditional SNR based analysis approach,the proposed analysis is of the prime importance for shedding light on the performance achievable by general strategies.
- Simulation results have confirmed the potential and success of the proposed strategies.