Decentralized Anomaly Detection via Deep Multi-Agent Reinforcement Learning

Hadar Szostak and Kobi Cohen (Senior Member, IEEE)
School of Electrical and Computer Engineering,
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

Abstract—We consider a decentralized anomaly detection problem, where multiple agents collaborate to localize a single anomalous process among a finite number $M$ of processes. At each time, a subset of the processes can be observed by each agent, and the observations from each chosen process follow two different distributions, depending on whether the process is normal or abnormal. The communication channel between agents is rate-limited. The objective is a sequential search strategy that minimizes the Bayes risk, consisting of the sampling cost, and the joint terminal cost among the agents. This problem generalizes previous studies that considered anomaly detection by a single detector. We develop a novel algorithm based on deep multi-agent reinforcement learning optimization to minimize the Bayes risk. Numerical experiments demonstrate the ability of the algorithm to learn good policies in this challenging problem, and improve the single-agent performance by applying the proposed multi-agent collaborative learning method.

I. INTRODUCTION

We consider a decentralized sequential anomaly detection problem, where multiple agents collaborate to localize a single anomalous process among a finite number $M$ of processes. The processes can represent components (such as routers and paths) in a cyber system, channels in a communication network, potential locations of targets, and sensors or UAVs monitoring certain events. Due to resource constraints, only a subset of the processes can be probed by the agents at a time. Each normal process generates a noisy observation drawn from a typical distribution corresponding to a normal behavior. When the anomalous process is sampled, then a noisy observation drawn from a distribution that deviates from the typical distribution is obtained. The anomaly detection problem is active in the sense that the agent can control the observation distribution by choosing which process to sample at each given time. Thus, the problem has a control aspect on the distribution of the generated observation. The communication channel between agents is rate-limited. The objective is a sequential search strategy that minimizes the Bayes risk, consisting of the sampling cost, and the joint terminal cost among the agents.

The active anomaly detection problem was studied in recent years under different assumptions and variations, including linear search, sublinear search, correlated processes, and various cost functions [1]–[19]. Since solving the optimal solution is mathematically intractable, even for the single-agent case, analytic solutions mainly focused on developing asymptotically optimal algorithms in terms of minimizing the required sample size as the error rate approaches zero. In this paper, we develop the first solution for this problem in the multi-agent setting based on deep reinforcement learning (DRL).

The decentralized active anomaly detection problem can be considered as a variation of the (centralized single-agent) sequential design of experiments problem first studied by Chernoff in 1959 [20], in which the agent chooses and dynamically changes the experiment (thus the observation model) among a set of available experiments. Chernoff focused on the case of binary hypotheses and showed that a randomized strategy (referred to as the Chernoff test) is asymptotically optimal as the maximum error probability diminishes. Specifically, the Chernoff test chooses the current experiment based on a distribution that depends on the past actions and observations. Variations and extensions of the problem and the Chernoff test were studied in [21]–[25], where the problem was referred to as controlled sensing for hypothesis testing in [22], [25] and active hypothesis testing in [23], [24].

The decentralized active anomaly detection problem can also be considered as a variation of the decentralized Wald problem pioneered by Teneketzis in 1987 [26]. Teneketzis considered a similar setting of Wald’s sequential hypothesis testing problem, where the observation model under each hypothesis is predetermined. The new aspect in [26] is the decentralization of the problem, where two agents make decisions, and the Bayes risk consists of the sampling cost and the joint terminal cost among agents. Subsequent studies considered variations and extensions of the problem, where heuristic solutions were proposed. Here, the new aspect is that agents control the observation distributions via the available action set.

A. Main Results

We now summarize the main results of this paper. First, we formulate a novel decentralized anomaly detection problem, where multiple agents collect noisy observations sequentially from the environment and need to locate the anomalous process. This problem generalizes previous studies that considered anomaly detection by a single detector. Second, we develop a novel algorithm based on deep multi-agent reinforcement learning framework to solve the decentralized active anomaly detection problem,
dubbed DRL for Decentralized Active Anomaly Detection (DDAAD). Using DDAAD, the agents learn a policy that maps a state to an action. The policy is learned online by exploring actions and observing the received reward, with the goal of minimizing the Bayes risk. Due to communication constraints, only short-messages that specify the selected probing location are shared. The agents need to learn the best detection strategy from their own observations, and the shared messages. Existing DRL-based algorithms do not incorporate multi-agent learning for active anomaly detection problems with observation control, as considered here. This is the first paper that tackles this problem. Specifically, DDAAD uses an Actor-Critic DRL method to train the agents, with Proximal Policy Optimization (PPO), so as to minimize the Bayes risk. Simulation results demonstrate the ability of the agents to learn how to collaborate to improve performance, and show that DDAAD is superior to single-agent learning.

B. Other Related Work

DRL has been investigated in recent years to solve active statistical inference problems [27]–[31]. Furthermore, it allows efficient implementations in real-world tasks and cheap-hardware agent devices [32], [33]. In [27], [29], [31] active hypothesis testing was studied using data-driven DRL approaches, used to approximate the optimal solution. In [27], a Q-learning method was used to design an experiment selection strategy for the agent such that the confidence level on the true hypothesis increases as quickly as possible. In [29], [31], Actor-Critic DRL algorithms were used.

Multi-agent tasks have been studied in different domains using online learning and deep learning, and can be broadly divided into two groups: Competitive and cooperative multi-agent tasks [34]–[38]. In competitive multi-agent tasks, each agent (or a group of agents) aims to maximize its own reward regardless of the performance of other agents. In cooperative multi-agent tasks, agents collaborate to maximize a common reward, which can be coupled among agents as well. Also, the reward for each agent might trade-off between individual and social rewards. Here, we consider a cooperative setting.

II. SYSTEM MODEL AND PROBLEM STATEMENT

Consider the following decentralized anomaly detection problem. $K$ agents are required to detect the location of a single anomalous object (referred to as the target) located in one of $M$ cells. Let $H \in \mathcal{H}$ be the true hypothesis. If the target is in cell $m$, we say that hypothesis $H_m$ is true, i.e., $H = H_m$. The a priori probability that $H_m$ is true, i.e., $H = H_m$, is denoted by $p_m$, where $\sum_{m=1}^M p_m = 1$. To avoid trivial solutions, it is assumed that $0 < p_m < 1$ for all $m$. The set of actions is denoted by $\mathcal{A}$, including sampling actions that control the observation distribution of the sample, and a stopping action that any agent can take to finalize its test and declare the location of the anomalous process (while other agents may continue the test). The stopping rule is denoted by $\tau_k$, which is the time when agent $k$ finalizes the test by declaring the location of the anomalous process. Let $\delta_k \in \{1, 2, ..., M\}$ be a decision rule, where $\delta_k = m$ if the agent declares that $H_m$ is true. At each time $n$, each agent $k$ broadcasts a short-message $m_k(n)$ (which can be a null message as well) to all other agents. At each time $n \in \{1, 2, ..., \}$, agent $k$ can take a sampling action $a_k \in \mathcal{A}$ based on the history of the task it has (i.e., all its past sampled observations, received messages, and actions) and obtain observation $o_k(n)$. The collected observation $o_k(n)$ is an independent random variable drawn from probability density $p_k(a_k)$ under the true hypothesis $H$ (a typical distribution for normal processes or a distribution that deviates from the typical distribution for the anomalous process). The time series vector of sampling actions is denoted by $a_k(n)$, $n = 1, 2, ..., $. A policy for agent $k$ is given by the tuple $\Pi_k \equiv (\delta_k(n), a_k(n))$. The multi-agent policy for all $K$ agents is thus given by $\Gamma \equiv \{\Pi_k\}_{k=1}^K$.

Let $c$ be the sampling cost for each observation for each agent. Let $J(\delta^1, \delta^2, ..., \delta^K, H)$ be the terminal cost incurred by the final decisions $\delta^1, \delta^2, ..., \delta^K$ of the agents, where $H$ is the true hypothesis. The terminal cost is coupled in general, thus, $J(\delta^1, \delta^2, ..., \delta^K, H) \neq \sum_{k=1}^K J(\delta^k, H)$. Otherwise, the problem decomposes into $K$ independent classic AHT problems [26]. Furthermore, we assume that the terminal cost increases with the number of agents (say $i$) that declare $\delta^i \neq H$ (i.e., experiencing more mistakes incurs a higher terminal cost). The Bayes risk under multi-agent strategy $\Gamma$ when hypothesis $H$ is true is given by:

$$R_H(\Gamma) \equiv \mathbb{E}_H \left\{ \sum_{i=1}^K c \tau^i \right\} + J(\delta^1, \delta^2, ..., \delta^K, H),$$

(1)

where $\mathbb{E}_H$ denotes the operator of expectation with respect to the probability measure under hypothesis $H$. The average Bayes risk is given by:

$$R(\Gamma) \equiv \sum_{m=1}^M p_m R_{H_m}(\Gamma).$$

(2)

The objective of the decentralized anomaly detection problem is to find a multi-agent policy $\Gamma$ that minimizes the average Bayes risk $R(\Gamma)$:

$$\inf_{\Gamma} R(\Gamma).$$

(3)

III. THE DRL FOR DECENTRALIZED ACTIVE ANOMALY DETECTION (DDAAD) ALGORITHM

In this section we present the structure of the proposed DDAAD algorithm to solve (3) based on DRL optimization. We then present preliminary simulation results. The detailed algorithm and extensive simulation results can be found in the extended version of this paper [39].
A. The Structure of DDAAD Algorithm

DDAAD algorithm uses DRL to provide a good approximation of the objective value by combining a deep neural network (DNN) with reinforcement learning. The algorithm trains a DNN in the multi-agent setting, such that each agent would map its current observation and communication signals to sampling actions or stopping action and decision rule based on the trained DNN. This is done by maximizing the accumulated (discounted) reward (or the negative Bayes risk in (3)). DDAAD uses Actor-Critic DRL method to train the agents, with Proximal Policy Optimization (PPO) [40] to ensure stable updates.

Specifically, the Q-value for each action-state pair for agent $k$ is computed via TD error updates. Then, PPO is used to the policy updates to ensure subtle updates. The Actor-Critic algorithm with PPO updates uses both clipping and adaptive penalty coefficient. The Policy and Value function networks do not share parameters between them. The learning process follows the Centralized Training and Decentralized Execution (CTDE) structure to provide the value network of each agent an additional information about other agents dynamics. Input signals are allowed following the a rate-limited communication channel between the agents.

B. Simulation Results

We simulated an anomaly detection task by two agents with the goal of locating a single anomalous process among 5 processes. Normal processes generate, when sampled, a Gaussian noisy observation with unit variance. An anomalous process generates, when sampled, a Gaussian noisy observation with unit variance plus unit DC signal. An agent can control the observation distribution by choosing which process to sample at each given time. We used the information regarding the selected action as the short-message signal shared between the agents.

We present the empirical error probability as a function of the average sample size (i.e., the average detection delay). The results are presented in Fig. 1. In addition to the multi-agent setting, we tested the single-agent setting as well as a benchmark for performance assessment. It can be seen that by collaborating, the agents reduce the detection delay for a given error probability value by roughly 15% as compared to the single-agent setting. A significant reduction is observed in terms of the Bayes risk as well, as can be seen in Fig. 2.

IV. CONCLUSION

We considered a novel formulation of decentralized anomaly detection problem, where multiple agents collaborate to localize a single anomalous process among a finite number $M$ of processes. Each agent can observe a subset of the processes at each given time, and the observations from each chosen process follow two different distributions, depending on whether the process is normal or abnormal. The agents communicate via a rate-limited channel. The objective is a sequential search strategy that minimizes a collaborative Bayes risk, consisting of the sampling cost, and the joint terminal cost among the agents. We developed a novel algorithm based on deep multi-agent reinforcement learning optimization to minimize the Bayes risk, and demonstrated the ability of the agents to collaborate to reduce the Bayes risk and improve performance as compared to the single-agent detection performance.

V. ACKNOWLEDGEMENT

This research was supported by the Israel Science Foundation (grant No. 2640/20)