NML Research Group
Internet-Draft
Intended status: Informational
Expires: September 14, 2017

M-S. Kim
Y-G. Hong
ETRI

March 13, 2017

Collaborative Intelligent Multi-agent Reinforcement Learning over a Network
draft-kim-nmlrg-network-00

Abstract

This document describes agent reinforcement learning (RL) in a distributed environment to transfer or share information for autonomous shortest path-planning over a communication network. The centralized node, which is the main node to manage agent workflow in hybrid peer-to-peer environment, provides a cumulative reward for each action that a given agent takes with respect to an optimal path based on a to-be-learned policy over the learning process. A reward from the centralized node is reflected when an agent explores to reach its destination for autonomous shortest path-planning in distributed nodes.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

In large surveilling applications, information of Critical Key Infrastructures and Resources (CKIR) to protect and share is necessary in larger ground, maritime and airborne areas, where there is a special need for collaborative intelligent distributed systems with intelligent learning schemes. These applications also need the development of computational multi-agents learning systems in large distributed networking nodes, where the agents have limited, incomplete knowledge, and only access to local information in distributed computing nodes over a communication network.

Reinforcement Learning (RL) is one effective technique to transfer and share information among agents for autonomous shortest agent path planning, as it does not require a-priori-knowledge of the agent’s behavior or environment to accomplish its tasks [Megherbi]. Such a knowledge is usually acquired/learned automatically and autonomously by trial and error.

Reinforcement Learning (RL) actions involve interacting with a given environment, so the environment provides an agent learning process with the elements as followings:

- Starting agent state, one or more obstacles, and agent destinations

- Initially, agent explores randomly in a given node

- Agents’ actions to avoid an obstacle and move to one or more available positions to reach its goal(s)

- After an agent reaches its goal, it can use the information collected in initial random path-planning work to improve its learning speed

- Optimal ways in the following phase and exploratory learning trials

Reinforcement Learning (RL) is one of the Machine Learning techniques that will be adapted to the various networking environments for automatic networks [I-D.jiang-nmlrg-network-machine-learning]. Thus, this document provides motivation, learning technique, and use case for network machine learning.

2. Conventions and Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Motivation

3.1. General Motivation for Reinforcement Learning (RL)

Reinforcement Learning (RL) is a system capable of autonomous acquirement and incorporation of knowledge. It can continuously self-improve learning speed with experience and attempts to maximize cumulative reward for a faster optimal path used in used in multi-agents-based monitoring systems [Teiralbar].

3.2. Reinforcement Learning (RL) in networks

In large surveilling applications, it is necessary to protect and share information in many Infrastructure and Resource area. In wireless networking layers, Reinforcement Learning (RL) is an emerging technology to monitor dynamics of the network to achieve fair resource allocation for nodes within the wireless mesh setting. Monitoring parameters of the network and adjusts based on the network dynamics can demonstrate to improve fairness in wireless environment Infrastructures and Resources [Nasim].

3.3. Motivation in our work

There are many different networking issues such as latency, traffic, management and etc. Reinforcement learning [RL] is one of the Machine Learning mechanisms that will be applied with multiple cases to solve diverse networking problems against human operating capacities. It can be a challenge-able due to a multitude of reasons such as large state space search, complexity in giving reward, difficulty in agent action selection, and difficulty in sharing/merging learned information among the agents in a distributed memory nodes to be transferred over a communication network [Minsuk].

4. Related Works

4.1. Autonomous Driving System

Autonomous vehicle is capable of self-automotive driving without human supervision depending on optimized trust region policy by reinforcement learning (RL) that enables learning of more complex and special Neural Network. Such a vehicle provides a comfortable user experience safely and reliably on interactive communication network [April][Markus].

4.2. Game Theory

The adaptive multi-agent system, which is combined with complexities from interacting game player, has developed in a field of reinforcement learning (RL). In the early game theory, the interdisciplinary work was only focused on competitive games, but

Reinforcement Learning (RL) has developed into a general framework for analyzing strategic interaction and has been attracted field as diverse as psychology, economics and biology [Ann].

4.3. Wireless Sensor Network (WSN)

Wireless sensor network (WSN) consists of a large number of sensors and sink nodes for monitoring systems with event parameters such as temperature, humidity, air conditioning, etc. Reinforcement learning (RL) in WSNs has been applied in a wide range of schemes such as cooperative communication, routing and rate control. The sensors and sink nodes are able to observe and carry out optimal actions on their respective operating environment for network and application performance enhancements [Kok-Lim].

4.4. Routing Enhancement

Reinforcement Learning (RL) is used to enhance multicast routing protocol in wireless ad hoc networks, where each node has different capability. Routers in the multicast routing protocol are determined to discover optimal route with a predicted reward, and then the routers create the optimal path with multicast transmissions to reduce the overhead in Reinforcement Learning (RL) [Kok-Lim].

5. Multi-agent Reinforcement Learning (RL) Technologies

5.1. Reinforcement Learning (RL)

Reinforcement Learning (RL) is one of the machine learning algorithms based on an agent learning process. Reinforcement Learning (RL) is normally used with a reward from the centralized node, and capable of autonomous acquirement and incorporation of knowledge. It is continuously self-improving and becoming more efficient as the learning process from an agent’s experience to increase an agent learning speed for autonomous shortest path-planning [Sutton][Madera].

5.2. Reward of Distance and Frequency

In general, an agent takes the return values of its current state and next available state to decide and move an action, but the learning process in Reinforcement Learning (RL) involves lots of limitations since it provides the agents with only a single level of exploratory learning process. The limitation is generated to reduce agent learning speed in an optimal path, so that the Distance-and-Frequency technique based on the Euclidean distance in Reinforcement Learning (RL) was derived to enhance agent’s optimal learning speed. Distance-and-Frequency is based on more levels of agent visibility to enhance learning algorithm by an additional way that uses the state occurrence frequency [Al-Dayaa].

5.3. Distributed Computing Node

Autonomous path-planning for multi-agent environment is related to agent transfer of path information, as the agents require information to achieve efficient path-planning on a given local node or distributed memory nodes over a communication network.

5.4. Agent Sharing Information

The quality of agent decision making often depends on the willingness of agents to share a given learning information with other agents for optimal path-planning. Sharing Information means that an agent would share and communicate the knowledge learned and acquired with / to other agents using Message Passing Interface (MPI). In sharing information, each agent has an attempt of exploring its environment, where all agents explore to reach their destinations via a distributed reinforcement reward-based learning method on the existing local distributed memory nodes. The agents can be running on the same or different nodes over a communication network (via sharing information). The agents have limited resources and incomplete knowledge of their environments. Even if the agents do not share the capabilities and resources to monitor an entire given large terrain, they are able to share the needed information for collaborative path-planning in distributed networking nodes [Chowdappa][Minsuk].

5.5. Sub-goal Selection

A new technical method for agent sub-goal selection in distributed nodes is introduced to reduce the agent initial random exploration with a given selected sub-goal.

[TBD]

5.6. Cluttered-index-based scheme

We propose a learning algorithm to optimize agent sub-goal selection. It is a proposed clutter-index-based technique for a new reinforcement learning scheme with a reward and an improved method to optimize multi-agent learning speed over a communication network.

[TBD]

6. Proposed Architecture for Reinforcement Learning (RL)

The architecture using Reinforcement Learning (RL) describes a collaborative multi-agent-based system in distributed environments as shown in figure 1, where the architecture is combined with a hybrid architecture making use of both a master / slave architecture and a peer-to-peer. The centralized node, assigns each slave computing node a portion of the distributed terrain and an initial number of agents. The network communication handles all communication among components and agents in the distributed networking environment. The components are deployed on different nodes. The communication handler alternatively sends one message from the outgoing queue and distributes one message in the incoming queue to the destination agent or component, and runs in a separate thread on each node with two message queues that consists of the incoming queue and the outgoing queue.

```
                  +--------------------------------------+
     +------------|----------+       |      +------------|----------+
     | Communication Handler |       |      | Communication Handler |
     +-----------------------+       |      +-----------------------+
     |        Terrain        |       |      |        Terrain        |
     +-----------------------+       |      +-----------------------+
                                     |
                  +--------------------------------------+
     +------------|----------+       |      +------------|----------+
     | Communication Handler |       |      | Communication Handler |
     +-----------------------+       |      +-----------------------+
     |        Terrain        |       |      |        Terrain        |
     +-----------------------+       |      +-----------------------+
                                     |
                        +-----------------------+
                        | Communication Handler |
                        +-----------------------+
                        |Centralized Global Node|
                        +-----------------------+
```

Figure 1: Top level components, deployment and agent communication handler

Figure2 shows the relationship of an action, state and reward between an agent and its destination in the environment for reinforcement learning. The agent does an action that leads to a reward from achieving an optimal path toward its destination.

```
                        +-------------------------+
  States & Reward ------| Centralized Global Node |<----------------+
         |              +-------------------------+                 |
         |                                                          |
         |                                                          |
         |                                                        States
         |                                                          |
         |                                                          |
  +-------------+                               +-------------+     |
  | Multi-agent |-------------Action----------->| Destination |-----+
  +-------------+                               +-------------+
```

Figure 2: Architecture Overview

7. Use case of Multi-agent Reinforcement Learning (RL)

7.1. Distributed Multi-agent Reinforcement Learning: Sharing Information

In this section, we deal with case of a collaborative distributed multi-agent, where each agent has same or different individual destination in a distributed environment. Since sharing information scheme among the agents is problematic one, we need to expand on the work described by solving the challenging cases.

Basically, the main proposed algorithm is presented by distributed multi-agent reinforcement learning as below:.

```
+--Proposed Algorithm--------------------------------------------+
|                                                                |
| Let N, A and D denote number of node, agent and destination    |
+----------------------------------------------------------------+
| Place N, A and D in random position(x, y)                      |
+----------------------------------------------------------------+
| Every A agents in N nodes                                      |
+----------------------------------------------------------------+
| Do inital exploration(random) toward D                         |
|  (1) Let S denotes current state                               |
|  (2) Relinguish S so other agent can occupy the positions      |
|  (3) Assign the agent's new position                           |
|  (4) Update the current state S <- Sn                          |
+----------------------------------------------------------------+
| Do optimized exploration(RL) for number of trials              |
|  (1) Let S denotes current state                               |
|  (2) Let P denotes action                                      |
|  (3) Let R denotes discounted reward value                     |
|  (4) Choose action P <- Policy(S, P) in RL                     |
|  (5) Move available directions by agent                        |
|  (6) Update the learning model in a new value                  |
|  (7) Update the current state S <- Sn                          |
+----------------------------------------------------------------+
```

Figure 3: Use case of Multi-agent Reinforcement Learning

Multi-agent reinforcement learning (RL) in distributed nodes can improve the overall system performance to transfer or share information from one node to another node in following cases; expanded complexity in RL technique with various experimental factors and conditions, analyzing multi-agent sharing information for agent learning speed.

7.2. Use case of Shortest Path-planning via sub-goal selection

Sub-goal selection is a scheme of a distributed multi-agent RL technique based on selected intermediary agent sub-goal(s) with the aim of reducing the initial random trial. The scheme is to improve the multi-agent system performance with asynchronously triggered exploratory phase(s) with selected agent sub-goal(s) for autonomous shortest path-planning.

[TBD]

7.3. Use case of Asynchronous Triggered Multi-agent with Terrain Cluttered-index-based

This is a new proposed technical reward scheme based on the proposed environment-clutter-index for the fast learning speed path-planning.

[TBD]

8. IANA Considerations

There are no IANA considerations related to this document.

9. Security Considerations

[TBD]

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

10.2. Informative References

[I-D.jiang-nmlrg-network-machine-learning]
Jiang, S., "Network Machine Learning", ID draft-jiang-nmlrg-network-machine-learning-02, October 2016.

[Megherbi]
"Megherbi, D. B., Kim, Minsuk, Madera, Manual., "A Study of Collaborative Distributed Multi-Goal and Multi-agent based Systems for Large Critical Key Infrastructures and Resources (CKIR) Dynamic Monitoring and Surveillance", IEEE International Conference on Technologies for Homeland Security", 2013.

[Teiralbar]
"Megherbi, D. B., Teiralbar, A. Boulenouar, J., "A Time-varying Environment Machine Learning Technique for Autonomous Agent Shortest Path Planning.", Proceedings of SPIE International Conference on Signal and Image Processing, Orlando, Florida", 2001.

[Nasim]   "Nasim ArianpooEmail, Victor C.M. Leung, "How network monitoring and reinforcement learning can improve tcp fairness in wireless multi-hop networks", EURASIP Journal on Wireless Communications and Networking", 2016.

[Minsuk]   "Dalila B. Megherbi and Minsuk Kim, "A Hybrid P2P and Master-Slave Cooperative Distributed Multi-Agent Reinforcement Learning System with Asynchronously Triggered Exploratory Trials and Clutter-index-based Selected Sub goals", IEEE CIG Conference", 2016.

[April]   "April Yu, Raphael Palefsky-Smith, Rishi Bedi, "Deep Reinforcement Learning for Simulated Autonomous Vehicle Control", Stanford University", 2016.

[Markus]   "Markus Kuderer, Shilpa Gulati, Wolfram Burgard, "Learning Driving Styles for Autonomous Vehicles from Demonstration", Robotics and Automation (ICRA)", 2015.

[Ann]   "Ann Nowe, Peter Vrancx, Yann De Hauwere, "Game Theory and Multi-agent Reinforcement Learning", In book: Reinforcement Learning: State of the Art, Edition: Adaptation, Learning, and Optimization Volume 12", 2012.

[Kok-Lim] "Kok-Lim Alvin Yau, Hock Guan Goh, David Chieng, Kae Hsiang Kwong, "Application of reinforcement learning to wireless sensor networks: models and algorithms", Published in Journal Computing archive Volume 97 Issue 11, Pages 1045-1075", November 2015.

[Sutton]   "Sutton, R. S., Barto, A. G., "Reinforcement Learning: an Introduction", MIT Press", 1998.

[Madera]   "Madera, M., Megherbi, D. B., "An Interconnected Dynamical System Composed of Dynamics-based Reinforcement Learning Agents in a Distributed Environment: A Case Study", Proceedings IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Italy", 2012.

[Al-Dayaa]
"Al-Dayaa, H. S., Megherbi, D. B., "Towards A Multiple-Lookahead-Levels Reinforcement-Learning Technique and Its Implementation in Integrated Circuits", Journal of Artificial Intelligence, Journal of Supercomputing. Vol. 62, issue 1, pp. 588-61", 2012.

[Chowdappa]
"Chowdappa, Aswini., Skjellum, Anthony., Doss, Nathan, "Thread-Safe Message Passing with P4 and MPI", Technical Report TR-CS-941025, Computer Science Department and NSF Engineering Research Center, Mississippi State University", 1994.

Authors' Addresses

Min-Suk Kim
ETRI
218 Gajeongno, Yuseong
Daejeon 305-700
Korea

Phone: +82 42 860 5930
Email: mskim16@etri.re.kr

Yong-Geun Hong
ETRI
161 Gajeong-Dong Yuseung-Gu
Daejeon 305-700
Korea

Phone: +82 42 860 6557
Email: yghong@etri.re.kr