Deep reinforcement learning in multi-agent soccer

Sanket Lokegaonkar, Junkai Zeng, Ali Pasha

Virginia Tech

Introduction

- Reinforcement learning: For an agent in a typical AI system, a sequence of tuple (s,a,r) is observed. The learning task is to find the long term reward Q(s,a), for a state s given action a. From a supervised learning point of view, the problem is equivalently about training a model with dynamically updated data features and targets.
- Deep Q-network based AI has been proved to achieve human level control on certain tasks, e.g., Atari games. In these tasks only one agent is concerned while the opponents, if exist, are treated purely as part of the environment.
- We investigated an expanded version of Deep Q-Network where the reinforcement learning scheme is implemented to perform a multi-agent task. In this approach, a new structure of neural network, the so called Deep Reinforcement Opponent Network (DRON) is made use of in order to model the Q network for the primal agent and the team and opponents' policy simultaneously.

DQN Background

- The problem can be modeled as a Markov Decision Process in the action space of one agent.
  - This can be solved by computing the *Q Value.*
  - Optimal policy always selects the action with the highest *Q Value*

$$Q^*(s,a) = \sum_{s'} \mathcal{T}(s,a,s') \left[r + \gamma \max_{a'} Q^*(s',a') \right]$$

- For complex problems with complex states, continuous approximation is required.
  - In a DQN, we use a neural network to find the optimal *Q* by minimizing the loss function *L*, given a sample consisting of state, action, final state and reward.

$$L_i(\theta^i) = E_{(s,a,s',r)\sim U(M)}[r + \gamma \cdot max_{a'} Q(s',a';\theta^{i-1}) - Q(s,a;\theta^i)]^2$$

DRON Background

- The problem can be modeled as a Markov Decision Process in the joint action space of all agents.
  - A^M defines the joint action space
    - $A^M = A^1 \times A^2 \ldots\ldots\ldots A^n$, where n is the number of agents.
  - *'a'* is the action of the primary agent, *'o'* is the joint action of all the other agents.

$$Q^{\pi|\pi^o}(s_t,a_t) = \sum_{o_t} \pi_t^o(o_t|s_t) \sum_{s_{t+1}} \mathcal{T}(s_t,a_t,o_t,s_{t+1}) \left[\mathcal{R}(s_t,a_t,o_t,s_{t+1}) + \gamma \mathbb{E}_{a_{t+1}} \left[Q^{\pi|\pi^o}(s_{t+1},a_{t+1}) \right] \right].$$

- Used to explicitly model the opponent action as a hidden variable and marginalizing over it.
  - The expected *Q value* is obtained by combining predictions from multiple *expert networks.*

Mixture of Experts

$$Q(s_t,a_t;\theta) = \sum_{i=1}^{K} w_i Q_i(h^s,a_t)$$
$$Q_i(h^s,\cdot) = f(W_i^s h^s + b_i^s).$$

Environment

- Environment used is RoboCup 2D Half Field Offense.
- Feature set can be divided into High Level and Low Level

High Level Features

- Agent Coordinates
- Orientation
- Ball's Coordinates
- Able to kick ball
- Position relative to goal
- Position relative to opponent and teammates
- Teammate's position, opponent's position
- Kit numbers

Low Level Features

- Agent position, velocity, angle
- Stamina, Player frozen
- Collision with ball, player, post
- Ball velocity, angle
- Penalty box location
- Corners, nearest playable field

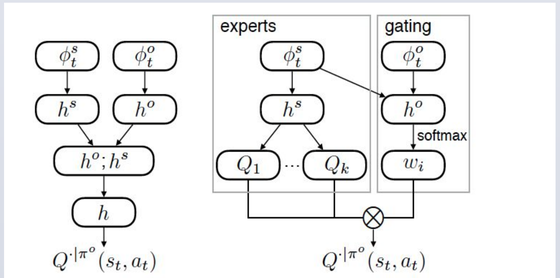
Actions

- High Level
  - Move, Shoot, Pass, Dribble, Catch, Reduce Angle To Goal, Defend Goal, Go To Ball, Mark Player
- Mid Level
  - Kick To, Move To, Dribble To, Intercept
- Low Level
  - Dash, Turn, Tackle, Kick

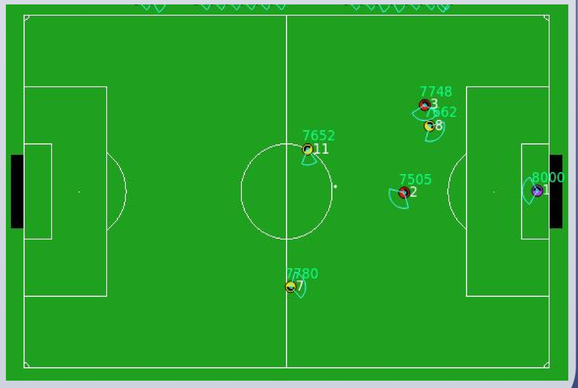
In this work, in order to simplify the training task, we only make use of high level features and high level actions.

Architecture

- **DRON-CONCAT**
- **DRON-MIXTURE OF EXPERTS**



Example Environment Snapshot



Results

Soccer Environment with 1 agent and 1 opponent.
The opponent can be learners, random agents , rule-based agent.

DQN vs DRON

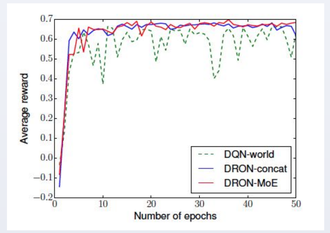


Figure 1: Learning curves over 50 epochs. DRON models are more stable than DQN.

Evaluations against Random Opponent:

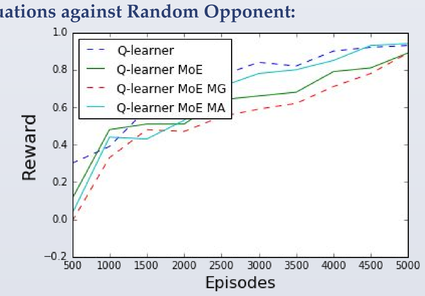


Figure 2: DQN model performs best against a random agent

Evaluations against Competitive Q-learning agent:

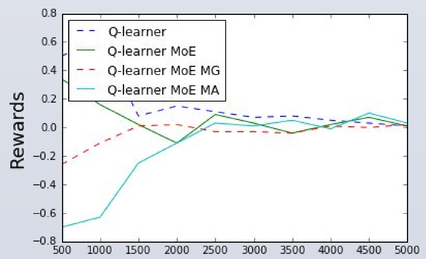


Figure 3: Rewards converge to 0 against a competitive Q-learning agent

Evaluations against Rule-based Agent:

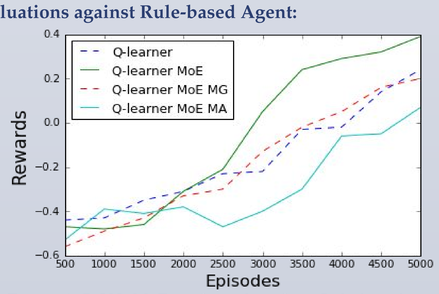


Figure 4: DQN- MoE performs best for a Rule-based agent.

Team and Opponent modeled jointly in concat network configuration with and without LSTMs(*figure 5*) using 3 teammates and 2 opponents. Peaks represent the instances when goal was scored:
After 1000 games, tally is 19 goals , 700 captured by defense.
Opponent is agent 2D base model using BFS based strategy selection configured by the HFO environment. On evaluation, LSTMs did not provide significant improvement.

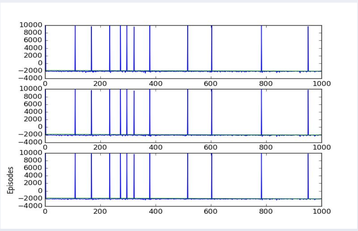


Figure 5

Conclusions/Analysis

Results (Single agent,Single Opponent, 6x9 grid):

- Mixture of Experts Network modeling opponent consistently performs better than the DQN agent, as seen in the rule based agent and random agent evaluations
- Against competitive DQN learning agent, the agent eventually learns to tie with the competitive opponent.

Results (3 Team agents,2 defending opponent, HFO Environment):

- Individually training agents together does not provide significant learning increase
- Some of the issues that we hypothesize could be the issue:
  - In contrast to the previous environment, the opponent being evaluated on is efficient.
  - Individual learning of the agents with either Boltzmann or Linear Annealed Policy leads to a changing joint model with no equilibrium condition.

Potential improvements:

- Adoption of curriculum based training of the individual agents
- Jointly learning single framework simulating multiagent communication

References

- **Opponent Modeling in Deep Reinforcement Learning, He He et.al**
- **Deep Reinforcement Learning in Parameterized Action Space, Matthew Hausknecht, Peter Stone**
- **Coordinating multi-agent reinforcement learning with limited communication, Chongjie Zhang, Victor Lesser**
- **Learning Multiagent Communication with Backpropagation ,Sainbayar Sukhbaatar et.al**
- **Miikkulainen, Risto, and Kristen Grauman. "Making Friends on the Fly: Advances in Ad Hoc Teamwork."**

RESEARCH POSTER PRESENTATION DESIGN © 2015
www.PosterPresentations.com