5T-04

RoboCup サッカーの Half Field Offense タスクへの深層強化学習の適用

川上 翔平† 田村 啓朗† 相馬 隆郎†
首都大学東京†

1.はじめに

近年，深層学習は機械学習の分野で大きな注目を集めており，Alpha-Go をはじめとするゲーム AI の分野や，ロボットの制御など，様々な分野で研究がされている．本研究では RoboCup Soccer 2D Simulation のサブタスクであり，強化学習に適した Half Field Offense タスクでマルチエージェントシステムにおける Deep Q-Network[1](以下 DQN)とその応用手法の適用を試みた．

2. Half Field Offense タスク

2.1 問題設定

Half Field Offense はサッカーコートの半分にオフェンス及びディフェンスエージェントを配置し，オフェンスがディフェンスを突破しゴールするまでの行動を学習するタスクである．その際，ディフェンスは学習せず，オフェンスはボール保持中のみ行動を学習する．エピソードはゴール時またはディフェンスがボールを獲得した時，ボールがコート外に出たときに終了する．本研究ではオフェンスは 3 台，ディフェンスはゴールキーパーを含め 2 台とした．

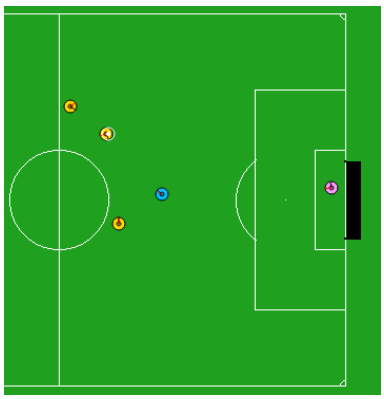

図 1 Half Field Offence タスク

A Deep Reinforcement Learning for Half Field Offense in RoboCup Soccer
†Kawakami Shohei, Tamura Hiroaki and Soma Takao, Tokyo Metropolitan University

2.2 状態変数

状態変数は全エージェントの x,y 座標，ゴールの中心までの距離及び角度，シュート時にゴール可能な角度の大きさ，各ディフェンスまでの距離の最小値の 14 次元とした．

表 1 Half Field Offence の状態変数

状態変数	要素数
各オフェンスエージェントの x,y 座標	6
各ディフェンスエージェントの x,y 座標	4
ゴール中心までの距離	1
ゴール中心までの角度	1
シュート時にゴール可能な角度の大きさ	1
各ディフェンスまでの距離の最小値	1

行動選択肢はボール保持，最適なドリブル，シュート，近いオフェンスへパス，遠いオフェンスへパスの 5 つとした．ここで最適なドリブルとは，Agent2D[2]を利用したプログラムにより速度や角度が決定されるドリブル行動命令である．オフェンスエージェントはこれにより学習し、ディフェンスエージェントは Agent2D を利用した NPC エージェントを使用した．

2.3 実験結果

30000 エピソード学習後の平均ゴール率について，DQN の各ハイパーパラメータを変えて実験した．表 2 に示すハイパーパラメータにおいて，最も高い平均ゴール率 0.8 を達成した．学習曲線を図 2 に示す．

表 2 DQN のハイパーパラメータ

Experience Replay 容量	2^{16}
中間層の構造	128-128-128
活性化関数	ReLU
Target Network 更新周期	1
最適化アルゴリズム	Adam
行動方策	Boltzmann

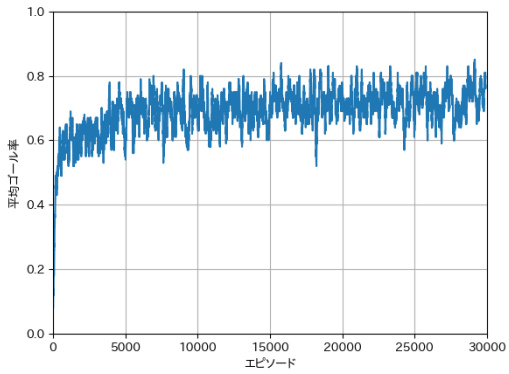

図 2　DQN の学習曲線

また，Double Deep Q-Network[3] (Double DQN)，Dueling Deep Q-Network[4] (Dueling DQN)と，この 2 つを組み合わせた Dueling Double Deep Q-Network(D3QN)の 3 つのアルゴリズムにおいても，中間層以外は表 2 のハイパーパラメータを適用し，中間層は層数を 1 層から 3 層，ニューロン数を 2^4 から 2^{10} の間で変化させながら実験をし，同条件の DQN の結果と比較した．それぞれのアルゴリズムにおいて最も高い平均ゴール率を達成したものを表 3 に示す．また，それぞれの学習曲線を図 3 から図 5 に示す．

いずれも DQN の結果を上回ることはなかったが，Double DQN では中間層の層数が少ない時に DQN よりも良い結果が出る傾向にあった．また，Dueling DQN では各層のニューロン数が少ない時に DQN よりも良い結果がでる傾向にあった．これより、応用手法の適用は性能向上よりもパラメータ数の削減という点で有用であった。

表 3　応用手法のハイパーパラメータ

	Double DQN	Dueling DQN	D3QN
中間層	512	64-64-64	256-256
ゴール率	0.772	0.772	0.767

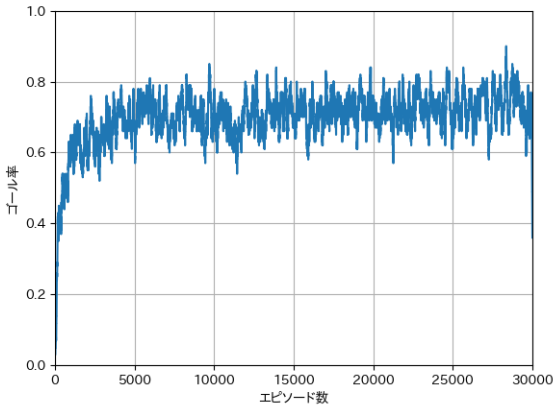

図 3　Double DQN の学習曲線

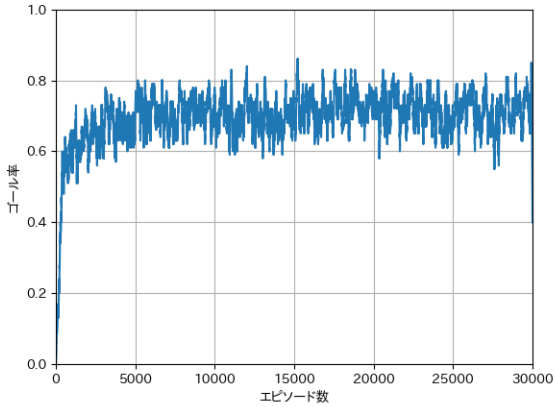

図 4　Dueling DQN の学習曲線

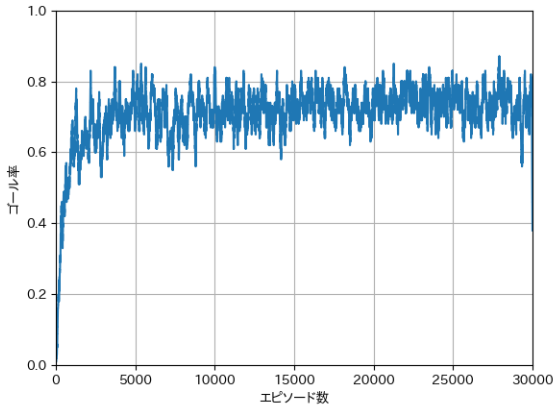

図 5　D3QN の学習曲線

3. おわりに

本研究では Half Field Offense タスクに Deep Q-Network とその応用手法を適用する実験を行った．Deep Q-Network を適用した場合，中間層 3 層，各ニューロン数 128 個としたときに平均ゴール率 0.8 を達成した．また，3 つの応用手法についても確認し，層数やパラメータ数の削減という点でその有効性を確認した．

参考文献

[1] V. Mnih, et al. “Human-level control through deep reinforcement learning,” Nature, 518 (7540):529–533, 2015.
[2] Akiyama Hidehisa. “Agent2d base code,” 2010.
[3] H.Hasselt, et al. “Deep Reinforcement Learning with Double Q-learning,” CoRR, abs/1509.06461, 2015.
[4] Z.Wang, et al. “Dueling Network Architectures for Deep Reinforcement Learning,” CoRR, abs/1511.06581, 2015.