

積み重ね強化学習を用いた RoboCup Soccer Keepaway タスクの検証

橋本敏成[†] 田原秀平[†] 相馬隆郎[†]

首都大学東京大学院 理工学研究科[†]

1. 研究背景

強化学習とはロボットが環境との相互作用により試行錯誤的に行動則を獲得していく手法である。応用事例としてロボットの歩行動作獲得、倒立振子の振り上げ安定化、サッカーエージェントの学習など様々な分野で研究が進められている。また学習を効率化する手法として積み重ね強化学習が提案されている。

RoboCup Soccer のサブタスクである Keepaway タスクについても強化学習を用いた様々な研究が行われている[1-3]。本研究では積み重ね強化学習を Keepaway タスクに適用させて、その有効性を検証することを目的とする。

2. 積み重ね強化学習

文献[4]で提案された積み重ね強化学習について述べる。人が歩行を行う時、例えば階段を登るときに水平な地面を歩くときの経験を忘れゼロから学習をし直すかといえそうではなく、水平な地面での歩行経験をもとに調整を行う形で環境の変化に対応していると思われる。本研究では、強化学習においてもこの考え方を導入した積み重ね型の強化学習システムを使用する。システムの構造を図1に

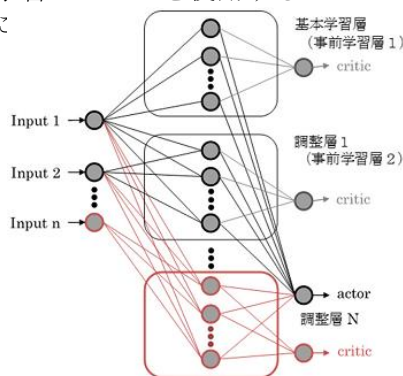


図1 積み重ね強化学習システム

本システムは事前に学習を行ったネットワークを層状に保持しており、各調整層はこれまでに学習したシステムの出力に対する補正値を学習する。

Verification of RoboCup Soccer Keepaway task using multi-layered reinforcement learning

[†]Toshinari Hashimoto, Shuhei Tahara, Takao Soma
Graduate School of Science and Engineering, Tokyo Metropolitan University

最初に学習を行ったネットワークの層を基本学習層、それに対し補正值の学習を行った各層を調整層と呼ぶことにする。また最新の調整層以外の層を事前学習層と呼ぶことにする。学習手順は、もっとも難易度の低い基本的なタスクから学習をはじめ、徐々に難易度の高いタスクを学習させることにより最終的に目標とする複雑なタスクの達成を目指す。具体的には以下の手順で学習を行う。

- (1) 低難易度の基本となるタスクを、基本学習層を用いて学習する。
- (2) i 番目の中間的タスク ($i = 1 \sim N-1$) に対して、調整層 i を用いて学習を行う。ここではそれ以前に学習したネットワーク (基本学習層と調整層 $1 \sim i-1$) の出力に対する補正値を学習する。
- (3) 最終的な目標タスクに対して、(1), (2) で学習したネットワークの出力に対する補正値を調整層 N で学習する。

3. Keepaway タスク

Keepaway タスクとは RoboCup Soccer のサブタスクである。本タスクは、図2のように限られた領域内で相手にボールを奪われないようにボールをキープすることを目的とする。相手からボールを取られるか、ボールがエリアから出るとエピソードが終了し、次のエピソードに進む。単独ではボールをキープすることができないため味方との協調行動を必要とするマルチエージェントタスクである。

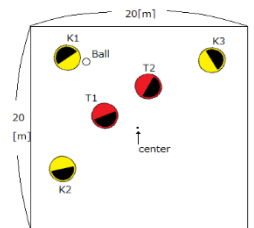


図2 Keepaway タスク

3.1 行動集合

RoboCup soccer シミュレーターでは各プレイヤーが `turn(angle)`, `dash(power)`, `kick(power, angle)` といった基本行動を 100[ms] 毎に行う。Keepaway タスクでは基本行動の組み合わせによる行動集合の爆発の問題を避けるため、ロボットの行動に stone らによって定義された基本行動を組み合わせたマクロ行動を用いる。マクロ行動の定義を以下に示す。

- `HoldBall()`: 可能な限り taker から遠い位置にボールを保持し、その状態を維持する。
- `PassBall(Kj)`: keeper Kj へボールをパスする。
- `GetOpen()`: taker から離れ、味方からパス

を受け取れるような位置に移動する。

- **GoToBall()**: 動いているか、または静止しているボールを取りに行く。
- **BlockPass(Kj)**: ボールと keeper Kj の間に移動する。

上記のマクロ行動の内、keeper が行うマクロ行動は **HoldBall()**, **PassBall(Kj)**, **GetOpen()** である。3 対 2 の環境で keeper がボールを保持している状態のとき keeper の行うことのできる行動集合は { **HoldBall()**, **PassBall(K1)**, **PassBall(K2)** } の 3 種類であり、この中から行動を選択する。ボールを保持していないときはボールに一番近い keeper が **GoToBall()** を選択しそれ以外の keeper は **GetOpen()** を行う。taker は次の 2 つの戦略を用いる。戦略 1 は keeper 2 体とも **GoToBall()** を行う戦略、戦略 2 は taker の片方が **GoToBall()** を行い、もう片方が **BlockPass(Kj)** を行う戦略である。

またロボットが環境から与えられる状態入力として、図 3 の状態変数を用いる。状態変数は 11 個の距離変数 ($d1 \sim d11$) と 2 個の角度変数 ($\theta 1$, $\theta 2$) の計 13 個である。

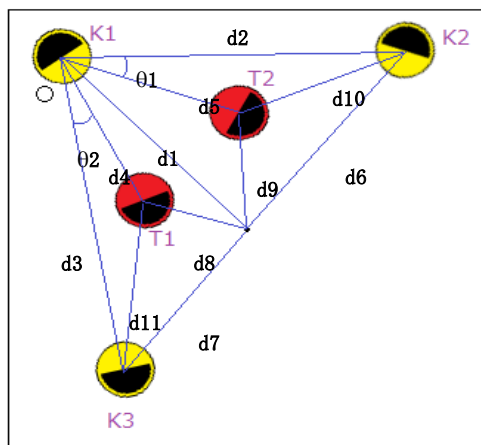


図 3 状態変数

4 シミュレーション

Keepaway タスクに対し、強化学習の適用を行った結果を示す。Keepaway の環境は従来研究と同じく $20[m] \times 20[m]$ の領域内で keeper3 体, taker2 体の 3 対 2 とした。taker は前節で述べた戦略 1 または戦略 2 に基づいて行動する。強化学習は方策オン型強化学習の sarsa を用い、各 keeper がそれぞれ独立して学習を行う。学習パラメータは試行錯誤的に決定した。

本シミュレーションでは戦略 2 に対して、積み重ね強化学習を適用した。なお中間タスクは設定せず、基本タスクと目標タスクの 2 層のみの積み重ね学習を行った。まず戦略 1 に対する Keepaway

を基本タスクとして基本学習層を用いて学習し、その後戦略 2 に対する Keepaway の学習を調整層 2 で行った。積み重ねを用いずに初めから戦略 2 で学習を行った場合との比較を図 4, 表 1 に示す。なお報酬は荒井ら[6]の報酬関数を用いた。積み重ね強化学習を用いた方がエピソード持続時間が長くなる結果を得られた。

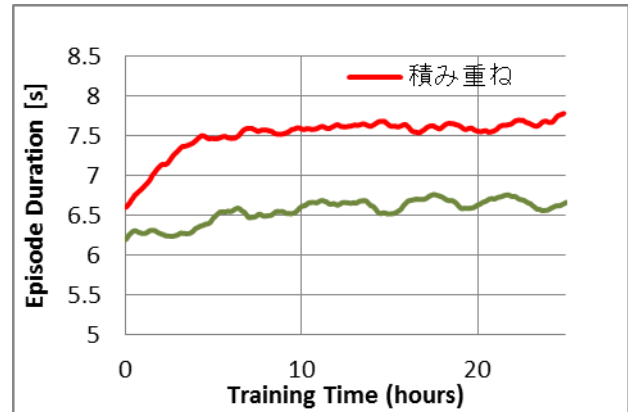


図 4 戦略 2 学習比較

表 1 各戦略における基底ユニット数

戦略	Ave. Num. of Units
1	163.1
2	216.8
2(積み重ね)	99.8

5 まとめ

積み重ね強化学習を keepaway タスクに適用し、その有用性の検証を行った。持続時間に関して従来の手法より劣る結果ではあるが、積み重ね強化学習が Keepaway タスクに対しても有効に機能することが確認できた。

6 参考文献

- [1] 坂本聖也、尾関基行、岡夏樹：RoboCup サッカーの Keepaway サブタスクにおけるパスの受け手の強化学習, 第 29 回人工知能学会 AI チャレンジ研究会資料
- [2] 伊佐野 勝人, 片上 大輔, 新田 克己：ロボカップサッカーシミュレーションの Keepaway における協調行動の学習, 第 22 回人工知能学会全国大会論文集, 2I3-3
- [3] 澤徹, 斉藤佑一, 渡邊俊彦, ファジィ Q 学習に基づくロボカップサッカーエージェントの Keepaway タスクの学習, 第 27 回ファジィシステムシンポジウム, セッション ID: P4-04
- [4] 安達良太, 田原秀平, 相馬隆郎 “積み重ね強化学習を用いた倒立振り子モデルに基づく 2 足歩行制御” 第 32 回日本ロボット学会学術講演会, 2B1-08
- [6] 荒井幸代, 田中信行, マルチエージェント連続タスクにおける報酬設計の実験的考察, 人工知能学会論文誌, 2006.