

# 連続状態空間の強化学習問題に対する合議アルゴリズムの効果

深沢 竜大<sup>†</sup> 橋 完太<sup>†</sup>  
<sup>†</sup> 工学院大学情報学部情報デザイン学科

## 1. はじめに

機械学習のアプローチとして強化学習が提案された [1]. 強化学習とは、環境から受けた報酬を基に行動の価値を逐次更新する手法である. 強化学習はロボティクスへの応用が盛んにおこなわれ、実空間に近い連続状態空間において、強化学習により行動を獲得する手法も検討されている [2]. 一方、離散空間において、強化学習手法の異なるエージェント同士の合議による行動は、単一の強化学習エージェントの行動よりも最適化されることが明らかにされた [3]. しかし、既存手法は状態空間の離散化手法の違いには着目しておらず、学習手法の違いのみでエージェントを区別していた. 本稿では、離散化方法の違いに着目し、連続状態空間における強化学習エージェントの合議アルゴリズムの効果検証を行う.

## 2. 帆走シミュレータ

連続状態空間として、風の吹く洋上におけるヨットの帆走を想定する [4]. ヨットの帆走は図 1 のように、揚力の前方成分が推進力となる. ただし、絶対風がヨットの船頭から  $-45^\circ \sim +45^\circ$  の方向(デッドゾーンと呼ばれる)から吹いてくる場合、推進力が得られない. ヨットは左右に  $3^\circ$  旋回可能である. ヨットの目標物として、洋上のランダムな位置にゴミを出現させる. ゴミを取得した時刻では  $+1000$ , それ以外は毎時刻  $-1$  の報酬が環境から与えられる.

## 3. 使用手法

### 3.1 離散化パラメータ

連続状態空間の離散化には離散化パラメータ  $\theta$  を設定する. この時、ヨットの正面が  $-\theta^\circ \sim +\theta^\circ$  の範囲となる. 本稿では、状態変数の数を 3 つに設定する.

### 3.2 合議

合議内のエージェントの学習手法は本稿では同一とし、離散化パラメータの値、学習パラメータの値、取得状態数について多様性を持たせる. 合議に含まれているエージェントの行動価値の更新には、仮想ヨットを用いる. 仮想ヨットの駆動はシミュレータ上のヨットと同様となる. 合議を行う前に、シミュレータ上のヨットと全く同じ状態の仮想ヨットをエージェントが操作し報酬を受ける. この時、仮想ヨットの受けた報酬を基に、行動価値の更新を行う. 合議手法は、Boltzmann Multiplication を用いる.

## 4. 実験

### 4.1 実験方法

本稿では、合議エージェントと単一強化学習エージェントの帆走性能を比較する. 2 隻のヨットを帆走シミュレータ上に設置する. スタート位置は同一である.

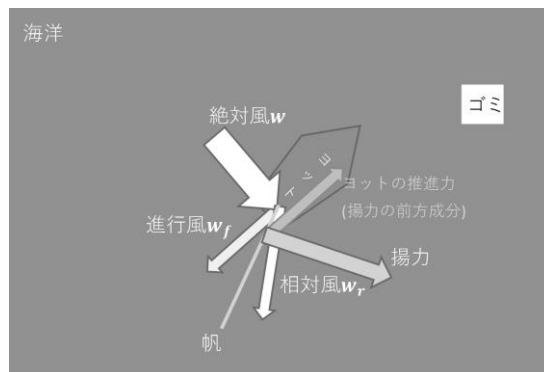


図 1 ヨット帆走の力学的原理

ヨット同士の衝突干渉はしない. 一方のヨットは合議エージェントによって駆動し、もう一方のヨットは、単一強化学習エージェントによって駆動する. 洋上に出現するゴミを早く拾った方に勝利点を与え、これを 1000 試行繰り返す.

### 4.2 実験結果

表 1 より、勝率は合議エージェントの方が高い事が明らかとなった ( $p < 1.0 \times 10^{-14}$ ). 連続状態空間においても、合議エージェントは、単一の強化学習エージェントよりも、帆走性能が高くなる.

表 1 合議エージェントと単一エージェントの勝率

	勝率	引き分け率
合議	.382	.415
単一	.200	

## 5. まとめと今後の課題

本稿では、連続状態空間の離散化方法に着目した多様な強化学習エージェントを用いて合議を行った. 結果、単一の強化学習エージェントよりも、その性能が高くなる事が明らかとなった. 今後、別の強化学習アルゴリズムにおいても同様の結果が得られるのか検証する必要がある.

## 参考文献

- [1] R. S. Sutton, A. G. Barto, “強化学習”, 森北出版, 2000.
- [2] 海津宏, “強化学習による倒立振子の安定化制御について,” 東京都立産業技術高等専門学校研究紀要 5, 2011.
- [3] W. Marco, V. H. Hasselt, “Ensemble algorithm in reinforcement learning,” Systems, Man, and Cybernetics, 2008.
- [4] 橋完太, “セーリング戦略の強化学習シミュレーション(移動体の空力特性),” シンポジウム: スポーツ・アンド・ヒューマン・ダイナミクス講演論文集, 2013.