

マルチエージェント問題での効率的な協調戦略獲得の検討

松永敬輔[†], 米田慎吾[†], 梶川嘉延^{†,‡}, 野村康雄[†]

関西大学工学部電子工学科[†]

関西大学学術フロンティアセンター[‡]

1. はじめに

従来の強化学習には対象となる問題の規模が大きくなるにつれ必要となる資源や時間が爆発的に増加する問題[1]があり, それに対し様々なアプローチ[2]が行われてきた. 本研究では, 学習環境の要素を制限した簡易学習環境にて学習を行い, その結果を主成分分析し, 獲得されたルールを用いることで学習の効率化をはかる. そして, その有効性を Robocup Rescue Simulation[3]にて検討する.

2. Robocup Rescue Simulation について

後に述べる提案手法の有効性の検討に大規模災害における人命救助を取り扱う Robocup Rescue Simulation を用いる. これは大規模災害直後の環境を再現し, 道路を直す警察官, けが人を運ぶ救急士, 火災を消す消防士の三種類のエージェントが存在し, その協調行動によって人命救助をシミュレーションするマルチエージェント問題である.

従来研究[3]では全エージェント同士が情報をやり取りし, それをもとに各自独自に判断し活動を行うのが効率的な人命救助方法とされていた. しかし, 情報の種類や救助対象, エージェント数といった要素数が膨大で効率的なエージェントの設計は困難であり, また, 一試行の所要時間が長い為強化学習を用いて効率的な救助行動を探索することも難しいという問題がある.

3. 提案手法

上記で述べたような大規模なマルチエージェント問題における効率的な学習方法を提案する.

3-1. 協調戦略の獲得

マルチエージェント問題は個々のエージェントの協調によって目的を達成する問題のことである. しかし, エージェントの数や種類が増加すると, 単純な学習の適用は困難になる. そこで人工生命における群集合アルゴリズム[4]の導入を行い, 複数の異なるエージェントに複雑な協調行動ではなく図1

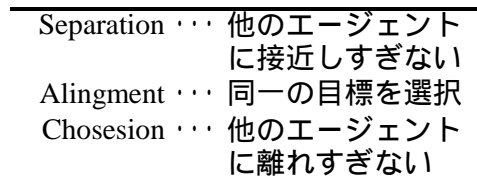
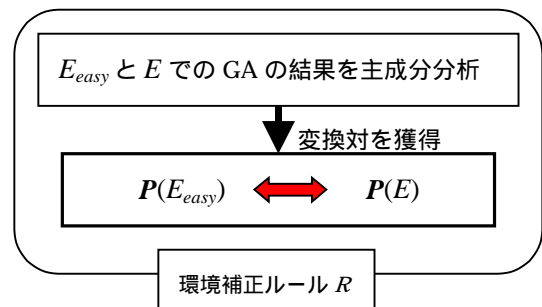


図1 群集合アルゴリズム



$$R\{P(E'_{easy})\} \quad P(E')$$

図2 学習効率化の概念

に示す単純なルールによる集団行動を行わせる. そして, 単純な集団行動を取ることで, 複雑な協調行動を行った場合と同等な結果の獲得を行うことを提案する. なお, 集団化した複数のエージェントを Package と呼ぶことにする.

そして, より精度の高い行動を獲得する為に環境 E に応じて Package の数, 内容を示すパラメータ $P(E)$ を動的に変化させる. その為, 最適な Package を遺伝的アルゴリズム(GA)にて獲得する. 遺伝子は Package に含まれるエージェントの構成とする. しかし, 一試行に時間がかかる場合, そのまま GA を行うのは困難である為, 以下の手法を提案する.

3-2. 主成分分析を用いたルール獲得

試行に要する時間が膨大な場合, 前章で述べた GA を各環境において毎回行うのは困難である. よって, GA による学習の効率化の新たな手法を提案する. まず, 要素を簡略化し, 所要時間を短縮した簡易学習環境 E_{easy} と簡略化していない学習環境 E で GA を実行する. 次に, 各々の学習で得られた Package のパラメータ $P(E_{easy})$ と $P(E)$ の主成分分析を行いパラメータ同士の関連性を獲得する. 最も関連性の高いパラメータ対を環境補正ルール R と呼

「A Study on Efficient Cooperation Strategy Acquisition in Multi-Agents Problem.」

[†]「Keisuke Matunaga, Shingo Yoneda, Yoshinobu Kajikawa, Yasuo Nomura · Department of Electronics, Faculty of Engineering, Kansai University」

[‡]「Yoshinobu Kajikawa · Frontier Science Center, Kansai University」

ぶことにする．そして，他の学習環境 E' で GA を行う場合，事前にその環境の簡易学習環境 E'_{easy} で，GA を実行し $P(E'_{easy})$ を得る．そして， $P(E'_{easy})$ と同じパラメータを持つ変換対を環境補正ルール R から式(1)より獲得し，それを学習環境 E' で使用する．

$$R\{P(E'_{easy})\} \quad P(E') \quad (1)$$

これによって学習の効率化を行うことができる．

次に，分析に用いる主成分分析について述べる．主成分分析とはいくつかの要素を合成・圧縮して1つの主成分にし，それによって要素の特性を解析する分析方法である．式(2)に示す GA のパラメータ $P(E)$ に対し共分散行列から全パラメータを圧縮した主成分を求め，その内，第一主成分による評価のより近いもの同士を環境補正ルール R とする．

$$P(E) = \{N_{\text{警察官}}, N_{\text{消防士}}, N_{\text{救急隊}}, X, Y\} \quad (2)$$

$$\begin{cases} X = \frac{N_{\text{警察官}} + N_{\text{消防士}} + N_{\text{救急隊}}}{N_{\text{all}}} \\ Y = \text{同じ構成のPackageの個数} \end{cases}$$

ここで， $N_{\text{警察官}}, N_{\text{消防士}}, N_{\text{救急隊}}$ は Package に含まれる各エージェントの数， N_{all} はシミュレーションに使用される全エージェント数である．

4. 結果と考察

Robocup Rescue Simulation にて本手法の有効性を示す．評価は Robocup Rescue Simulation で用いられる人命救助の成功率を表す評価値 F を用いることにする． F は大きくなるほど人命救助の成功率が高いことを表す．シミュレーション条件を表 1 に示すシミュレーションでは従来手法による結果，提案手法による結果，学習環境で長時間かけて実際に GA を実行した結果を，それぞれ結果 A，結果 B，結果 C として表 2 に示す．また，結果獲得まで要した時間 $T[\text{min}]$ に対するシミュレーションの評価値 F の対時間改善率 C_T を以下の式(3)より求め，表 2 に併せて示す．

$$C_T = \frac{|T_{\text{MAX}} - T|}{T_{\text{MAX}}} \times \frac{F}{F_{\text{MAX}}} \quad (3)$$

但し， $T_{\text{MAX}}, F_{\text{MAX}}$ は T, F の最大値である．表 2 から評価 F を比較すると従来手法より提案手法が良い結果を得ているが，実際に GA による学習を行った結果には及ばない．しかし，対時間改善率 C_T では提案手法が良い結果を得ており，提案手法によって学習の効率化を行えたといえる．

表 1 シミュレーション条件

(a)使用マシン		
CPU:Pentium4 2.8GHz		
メモリ:512MB		
(b)シミュレーション設定		
	簡易学習環境	学習環境
$N_{\text{警察官}}$	2	10
$N_{\text{消防士}}$	2	10
$N_{\text{救急隊}}$	1	5
N_{all}	5	25
救助対象	219	1573
一試行[ターン]	50	300
一試行[min]	1	8.3
(c)GA の設定		
個体数：5，エリート保存戦略		
交差確率：1，突然変異確率：1		

表 2 シミュレーション結果

	結果A	結果B	結果C
F	81.7047	83.5843	86.302056
F_{MAX}	98		
T		2694	22534
T_{MAX}	22535		
C_T		0.750939	3.91×10^{-5}

5. まとめ

本稿では大規模なマルチエージェント問題における効率的な学習方法を提案し，その有効性を示した．今後は，より詳細な学習環境補正ルールの獲得を目指す．

参考文献

- [1]R. E. Bellman, Adaptive Control Processes, Princeton Univ. Press, New Jersey, USA, 1961.
- [2]伊藤昭，金淵満，“知覚情報の粗視化によるマルチエージェント強化学習の高速化- ハンターゲームを例に-”，信学論(D-I)，Vol. J84-D-I, No. 3, pp. 285-293, Mar. 2000.
- [3]RoboCup Rescue Official Web page, URL:<http://robomec.cs.kobe-u.ac.jp/robocup-rescue/index.html>.
- [4]Craig. W. Reynolds, “A Distributed Behavioral Model”, Computer Graphics, Vol. 21, No. 4, pp. 25-34, July. 1987.