Q-Learningを用いたマルチエージェントシステムに関する検討
4J－5　－視覚情報の動的切替による学習速度向上の試み－

真鍋 彰宏　梶川 嘉延　野村 康雄
関西大学工学部電子工学科

1. はじめに

近年，自律型のエージェントの学習方法として，強化学習によるエージェントの行動獲得が研究されている．更にそのような強化学習を用いて，複数のエージェントがそれぞれに学習し，なんらかの強調動作を獲得するマルチエージェントシステムが注目されている．このようなマルチエージェントシステムでは，代表的な強化学習である Q-learning[1]を用いてエージェントの学習が可能であることがわかっている[2][3]．しかしマルチエージェントの学習に Q-learning を用いた場合，どうしても状態数の増加により収束速度が低下してしまう．そこで本稿では，エージェントの視覚情報を動的に変化させることにより，収束速度の向上を図る．

2. Q-learning[1]

強化学習の代表的な手法の一つに Q-learning がある．Q-learning は状態と行動の対を Q-値と呼ばれる重みを用いて評価し，それを元に行動の学習を行う手法である．この Q-値の更新は以下の式(1)を用いる．

$$Q(i_t,u_t) = (1-\alpha)Q(i_t,u_t) + \alpha(r_t + \gamma \max Q(i_{t+1},u_t)) \quad \cdots\cdots\cdots (1)$$

式(1)において，αは学習率であり$0<\alpha\leq 1$である．

3. 追跡問題

追跡問題とは，Fig.1 に示すような 2 次元格子状のトーラス構造である環境で複数のハンターエージェントが 1 体の獲物エージェントを捕獲することを目的としたマルチエージェント系のモデルである．

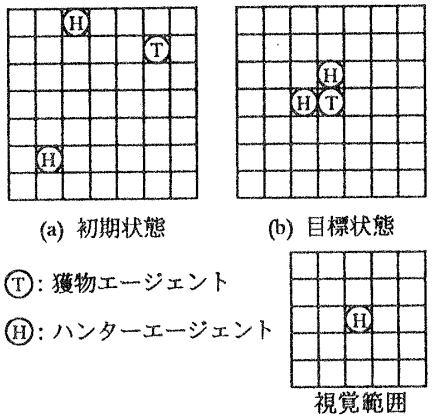

Fig.1　The hunting problem.

Fig.1(a)のように各エージェントをランダムに配置した状態を初期状態として行動を開始し，Fig.1(b)のように獲物エージェントに 2 体のハンターエージェントが隣接すれば獲物を捕獲したとする．本稿では獲物エージェントを捕獲した場合にハンターエージェントに対して報酬を与える．また，エージェントの能力を制限するために視覚範囲を設定する．この設定によりエージェントは環境を完全に認識できないという不完全知覚状態に陥ることとなる．またエージェントは“自分の視覚範囲内のどこに他のエージェントが存在するか”で状態を認識する．

4. 視覚情報の変化

エージェントの視覚情報は，学習における状態数を決定するものである．視覚情報をできるだけ詳しく得た場合，学習結果は良くなるが，収束速度は遅くなる．また視覚情報を削減すれば，収束速度は向上するが，学習結果は悪くなってしまう．そこで本稿では，視覚情報を動的に変化させることによりこの問題を解決する．具体的には Fig.2 に示すように，

A Study on Multi-Agent System by Q-learning.
-The trial of the learning speed improvement by the switching that sight information is dynamic.-
Akihiro Manabe, Yoshinobu Kajikawa,
Yasuo Nomura
Department of Electronics, Faculty of Engineering,
Kansai University
3-3-35, Yamate-cho, Suita-shi, Osaka, 564-8680, Japan

まず学習の初期段階では Fig.2(a)のように大雑把な視覚情報を用いて学習を行い，ある程度学習が進んだ時点で Fig.2(b)のような細かい視覚情報に変化させる．このように視覚情報を変化させることによって，学習の初期段階である程度の学習精度まですばやく学習を行い，その後更に学習精度を上げることが可能であると考えられる．

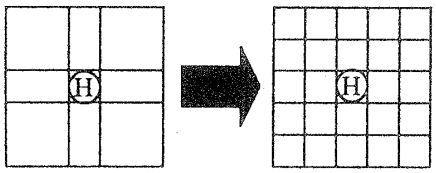

Fig.4 The sight information of agents.

ここで視覚情報を変化させる方法について述べる．具体的には，学習途中である区間のステップ数を用いて単回帰分析を行い傾きを求める．この傾きがある一定値以下の状態が何度か続けば視覚情報を変更する．このように，視覚情報を切り替えるタイミングを判断するようにする．

5. シミュレーション

ここでは実際にシミュレーションを行う．今回のシミュレーションでは単回帰分析に 5 つの点を使用し，しきい値は傾きが 0.1 以下の状態が 5 回続いた場合とする．このときの結果を Fig.5 に示す．但し，シミュレーション条件は Table 1 を用いる．

Table 1　Condition of simulation.

学習率α	0.02
割引率γ	0.001
Q-値の初期値	1000
報酬	1000
環境サイズ	7×7

Fig.5 より確実に収束速度が向上していることがわかる．次に，視覚情報の切り替えのしきい値を別の値に設定しシミュレーションを行った．この結果を Fig.6 に示す．

Fig.6 より，しきい値を変化させても確実に収束速度が向上しているので，本手法は有効である．

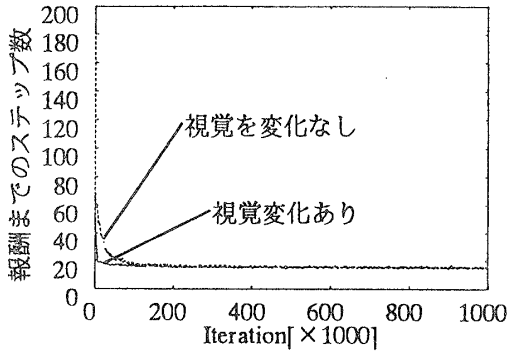

Fig.5　Convergence properties when recurrence analysis is used.

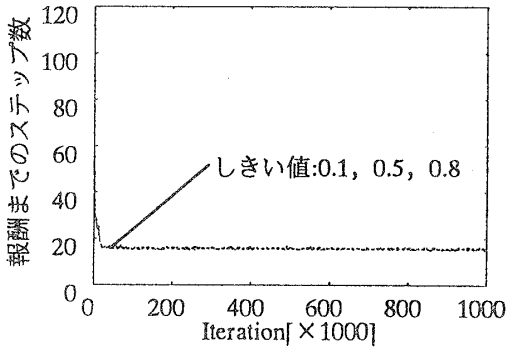

Fig.6　Convergence properties when threshold is changed.

6. まとめと今後の方針

本稿では Q-learning をマルチエージェントシステムに適用した場合に起こる問題である収束速度の低下をエージェントの視覚情報を動的に変化させることにより改善する手法を提案し，その有効性を示した。今後は更に情報量が増加した場合についての検討を行っていく。

【参考文献】

[1] Watkins 他：Machine Learning 8，pp.279-292 1992.

[2] 荒井他：人工知能学会誌，Vol.13，No.4，pp.609-618 1998.

[3] 村田他：コンピュータソフトウェア，Vol.14，No.1，pp.25-29 1997.