4J－1　重みの上限値及び負の強化を導入したProfit Sharingのマルチエージェントシステムにおける諸特性

西山 敦　堀井 亨　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，分散人工知能の分野においてマルチエージェントシステムの構築を目指す研究が盛んに行われている．このような研究におけるエージェントの自律的な学習[1]として経験強化型の学習法であるProfit Sharingが注目されている．しかし，環境が非マルコフ的なマルチエージェント問題では，エージェントの視覚範囲を制限すると部分観測問題が生じるため，無限ループに陥ってしまう．そこで，Profit Sharingに重みの上限値及び負の強化[2]を導入することで，部分観測問題を含むマルチエージェントシステムにおいても学習可能であることをシミュレーションにより示す．

2. Profit Sharing

Profit Sharingとは，報酬が与えられた時にそれまでの行動系列を一括に強化する経験強化型の学習法である．しかし，強化する行動系列には報酬に関係のない無効ルールが含まれる場合がある．従って，エージェントが適切な行動を学習するためには，この無効ルールの強化を抑制する必要がある．そこでProfit Sharingでは，以下の式(1)で示すような等比減少関数を強化関数に用いることにより，無効ルールの強化を抑制している，ここで，Sは強化減少値，Lは同一感覚入力下に存在する有効ルールの個数である．

$$f_n = \frac{1}{S} f_{n-1} \qquad \text{但し，} S \geq L+1 \quad \cdots\cdots(1)$$

Properties of Profit Sharing with Upper Limit of Weights and Negative Reinforcement in Multi-Agent System.
Atsushi Nishiyama, Toru Horii,
Yoshinobu Kajikawa, Yasuo Nomura
Department of Electronics, Faculty of Engineering,
Kansai University
3-3-35, Yamate-cho, Suita-shi, Osaka, 564-8680, Japan

3. 追跡問題

追跡問題とは，Fig.1に示すようなn×nの格子状トーラス環境上で複数の追跡者エージェントが1体の獲物エージェントを捕獲することを目的とした協調問題解決型マルチエージェントシステムのベンチマークである．

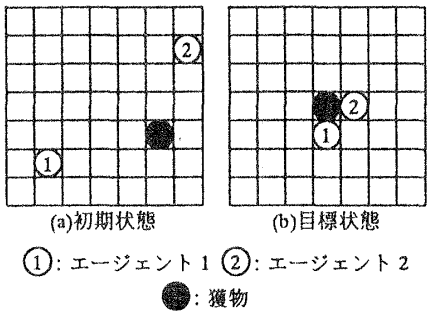

Fig.1　The hunting problem.

Fig.1の環境において，(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，(b)のように2体の追跡者エージェントが獲物エージェントに隣接すれば獲物を捕獲したとし，追跡者エージェントに報酬を与える．また，全てのエージェントにはFig.2で示す5×5の視覚範囲を設定し，獲物は視覚範囲内の全ての追跡者エージェントから遠ざかる方向へ移動する逃避的な行動を取るものとする．

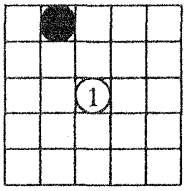

Fig.2　Observable range.

4. 収束特性

まず，通常のProfit Sharingを用いた時のシミュレーション結果をFig.3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.

強化関数	等比減少関数
公比	1/4
ルール重みの初期値	全て 10
目標状態の報酬	各エージェントに 100
環境サイズ	7×7

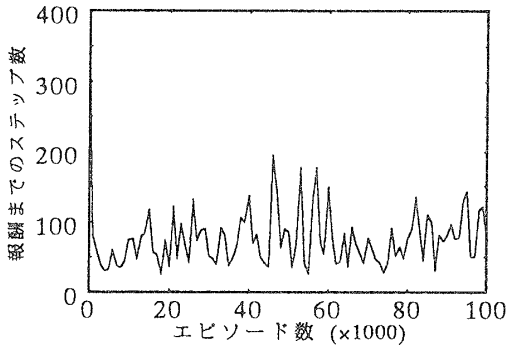

Fig.3 Convergence properties.

Fig.3 より，ある程度のステップ数で収束しているが，収束特性が乱れていることがわかる．これは，マルチエージェント問題では環境が非マルコフ的であるために部分観測問題が生じてしまい，獲物が見えない状態で無限ループに陥っているからである．そこで，これを解決するために，Profit Sharing に重みの上限値及び負の強化を導入する．

5. 重みの上限値及び負の強化を用いた場合の収束特性

次に，Profit Sharing に重みの上限値及び負の強化を導入し，その収束特性を検討する．但し，重みの上限値は

①各状態における 90%

②各状態における 95%

とし，負の強化は学習時に 1000 回の平均ステップ数の最小値を逐次的に記憶しておき，この最小値より学習ステップ数が大きくなった場合に，エピソードに余分な無効ルールが含まれると判断できるため，マイナス報酬を Profit Sharing と同様の方法で分配する．また，その他の条件は Table 1 と同様とする．この場合の収束特性を Fig.4 に示す．

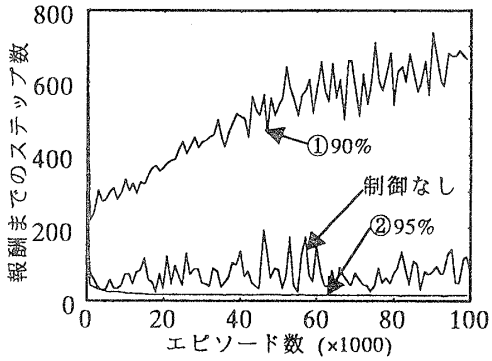

Fig.4 Convergence properties.

Fig.4 より，重みの上限値を 95%と設定した場合には平均ステップ数 15 程度で収束した．この結果から，重みを制御することにより，無限ループに陥ることを回避できることがわかる．しかし，90%と設定した場合にはステップ数が乱れ，しかも徐々に大きくなっていくことがわかる．これは，重みの上限値を 90%と設定すると，最適な行動を学習している状態においてもその行動を選択する確率が 90%となってしまうため，捕獲するまでに余計なステップ数がかかってしまうことになるからである． また，負の強化によってエージェントの過学習による無限ループを回避していることがわかる．

以上の結果から，重みの上限値及び負の強化を用いる場合，上限値を適切に設定することが必要で，これを誤ると収束特性がより悪化することがわかった．しかし，適切な設定ができれば，確率的に無限ループを脱することができ，収束特性を向上させることが可能となる．

6. まとめと今後の方針

本稿では，Profit Sharing に重みの上限値及び負の強化を導入することにより，確率的に無限ループを脱し，収束特性の向上が可能であることが分った．今後は，より効率的な問題解決のため通信を導入しその収束特性について検討を行う．

【参考文献】

[1] 宮崎他：人工知能学会誌，Vol.9,No.4，pp.580-587 1994.

[2] 山口他：人工知能学会誌，Vol.12, No.4，pp.570-581 1997.