A4-01　マルチエージェント環境下におけるファジィ状態分割型 Profit Sharing

嶋津 圭介 (指導教官：前田 陽一郎 助教授)

福井大学 工学部 知能システム工学科

1　緒　言

ここ数年、機械学習あるいはロボット工学の分野で研究が進んでいるテーマの１つとして強化学習法がある。人間の手によってロボットの制御を行う事が困難であるような、時々刻々と変化する未知環境において、ロボット自身が試行錯誤して得られた経験から適切な行動を自ら獲得することのできる強化学習は自律エージェントの学習法として注目されている。また、最近では複数のエージェントが協調的に問題解決を行うマルチエージェント系での研究も行われている [1]。

しかしながら、従来の強化学習法では主に状態、行動が離散的であるものが対象とされており、一般に連続的に状態や行動が変化する実世界の対象は扱いにくい。実際に現実の複雑な環境での学習を必要とする際には、連続値を含めた多様な入出力データを取り扱う必要がある。そのため Q-Learning において、連続値の状態、行動を取り扱うことを可能にするための研究も進められている [2]。

また、本研究室では強化学習をより複雑なタスクへ適用できるように、連続値の状態とともに、行動戦略の切り替えのようなケースが多いロボットを対象とした離散行動を取り扱うファジィ状態分割型 Q-Learning[3]、ファジィ状態分割型 Profit Sharing[4] もすでに提案している。しかしながら、これらの手法は単体のロボットの実世界における行動学習には有効であるが、例えばサッカーロボットのようなマルチエージェントの協調行動などの学習には適用できない。

そこで、本研究では、前述のファジィ状態分割型 Profit Sharing をマルチエージェント環境において適用可能な手法に拡張する方法について検討を行う。その際、報酬を各エージェントが分配する必要が生じるが、報酬を得た行動に対し、各エージェントが関与する度合いを寄与率として推定し、その値に従い報酬を比例分配し、集団として協調行動学習が可能な手法を提案する。さらに、本手法の有効性を検証するため、内部エネルギーを持つエージェントがエネルギーを補給しつつ、追跡するマルチエージェントシミュレーションを行ったので、その結果についても報告する。

2　マルチエージェント環境下におけるファジィ状態分割型 Profit Sharing 法の提案

本研究室で提案されたファジィ状態分割型 Profit Sharing (Fuzzy state division-type Profit Sharing:FPS) をマルチエージェント環境において適用可能な手法に拡張する方法について検討を行う。FPS は、状態分割にファジィ推論を導入することで、連続値の状態および一連の行動学習取りが扱える。

学習者がある環境から状態としてｎ次元連続値ベクトル $x = (x_1, x_2, ..., x_n)$ を観測した時、PS の行動決定を行うための行動重み w_{ik} を導出するためのファジィルールは以下のようになる。

$$R_i \ : \ IF\ x_1\ is\ A_{i1}\ and\ ...\ and\ x_n\ is\ A_{in}\ then\ w_{ik} \quad (1)$$

ここで $A_{ij}(i = 1, ..., l; j = 1, ..., n)$ は状態を表すファジィ集合で、状態の数 (=ファジィルールの数) は l 個存在する。また各ファジィルールの後件部シングルトン $w_{ik}(k = 1, ..., m)$ は PS におけるルール系列の行動重みを示しており、本手法においてはある状態における行動の行動重みを表し、行動の数 (m 個) だけ存在する。

式 (2) により、各ファジィルール R_i についての前件部の適合度 C_i を計算し、これらの重み付き平均により以下の式 (3) に従って最終的な推論結果である行動重み w_k^* が導出される。

$$C_i = A_{i1}(x_1) \wedge A_{i2}(x_2) \wedge ... \wedge A_{in}(x_n) \quad (2)$$

$$w_k^* = \frac{\sum_{i=1}^{l} C_i \cdot w_{ki}}{\sum_{i=1}^{l} C_i} \quad (3)$$

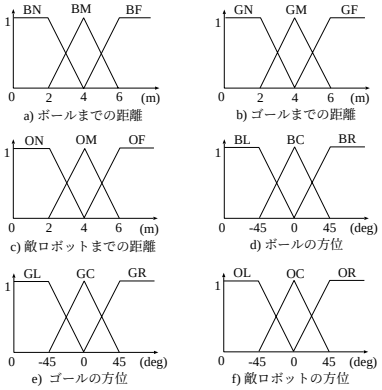

図 1. 前件部メンバーシップ関数

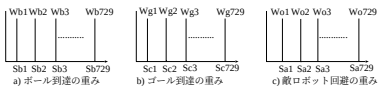

図 2. 後件部シングルトン

シュミレーションでは移動物体までの距離、餌場までの距離、障害物までの距離、エネルギー量を各エージェントの状態量 (前件部) として３つの行動に対する行動重みを後件部としてのように設定した。

本手法で用いたメンバーシップ関数及びファジィルールを図 1 から図 3 に示す。

2.1　寄与率の算出

本研究では、FPS をマルチエージェント化するために、ある行動に対し報酬が与えられた場合、その報酬

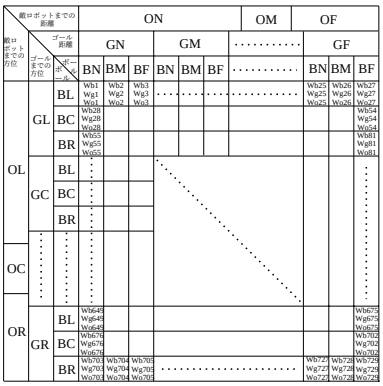

図 3. ファジィールール

行動に関与した度合い (以後、寄与率と呼ぶ) に応じて各エージェントに報酬を分配する強化学習法を提案する。ここでは報酬を獲得した時点でのエージェントの報酬行動への関与の度合いを表したものを寄与率 c_i と考え、式 (4) で表す。この式は報酬獲得時のエージェントとの距離や報酬獲得に関わる時間を用いて表したものである。ここで、直接貢献エージェントを報酬発生の直前に行動を実行したエージェントとする。

$$c_i = \alpha \int_{t=0}^{T} \frac{dt}{d_i(t)} + \beta \frac{T_{distin}}{T_{total}} \qquad (4)$$

d_i は直接貢献エージェントと i 番目のエージェントの距離、T はエージェントが報酬を獲得した時間、T_{distin} は T 中で、直接貢献エージェントのある一定の距離にいた時間、T_{total} は、 報酬獲得時から次の報酬獲得時までの時間、α、β は正規化定数。

2.2 報酬分配

前述の寄与率 c_i を用いて、関与したと考えられるエージェントに式 (5) を用いて報酬を分配する。また、式 (5) は i 番目のエージェントが獲得する報酬を表したものである。この式は全エージェントの寄与率の総和に対するあるエージェントの寄与率の度合いを表したものであり、その値に従い報酬を比例分配するものとする。

$$ri = \frac{c_i}{\sum_{i=1}^{n} c_i} \times R \qquad (5)$$

ri ：i 番目のエージェントの報酬
c_i ：各エージェントの寄与率
R ：報酬値
n ：エージェント数

3 物体追跡シミュレーション

本研究で用いるシミュレータは Linux 上で Xlib のグラフィックスライブラリおよび $Motif$ ウィジェットを利用し、C 言語を使用して開発したものである。シミュレータの画面を図 4 に示す。

この物体追跡シミュレータで用いた前提条件は次の通りである。

- 各エージェント、餌場、捕獲物体の初期位置はランダムに配置される。

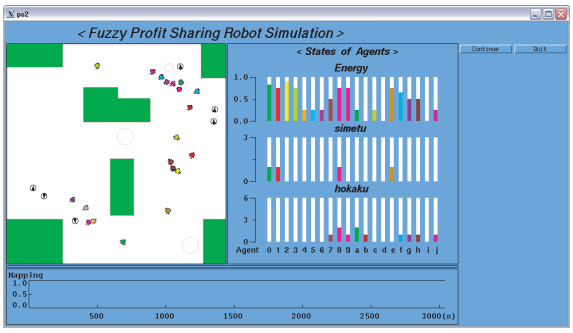

図 4. シミュレータの画面構成

- エージェントは餌場にたどり着くことにより自らのエネルギーを補給しつつ、物体を追跡し捕獲する。
- 餌場の保有量は有限であり、エネルギーがなくなった餌場は消滅し、また新たな餌場がランダムに出現する。
- 自己以外のエージェントは障害物とみなす。
- 移動物体は追跡に来たエージェントが近づいてきた場合のみ遠ざかる方向へ逃げるものとする。

ここで、各エージェントは移動物体の捕獲、餌場でのエネルギー補給、障害物の回避を学習目標として学習を行う。各学習エージェントのとりうる行動としては、1) 移動物体の追跡に向かう、2) 餌場に向かう、3) 障害物を回避するの 3 行動を取るものとする。

4 結 言

本研究では、ファジィ状態分割型 Profit Sharing をマルチエージェント環境において適用可能な手法に拡張する方法について検討を行い、報酬を得た行動に対し各エージェントが関与する度合いを寄与率として推定し、その値に従い報酬を比例分配し、集団として協調行動学習が可能な手法を提案した。

参考文献

[1] 荒井 幸代, 宮崎 和光, 小林 重信: “マルチエージェント強化学習の方法論－ Q-Learning と Profit Sharing による接近”, 人工知能学会誌, Vol.13, No.4, pp.609-618 (1998)

[2] 堀内 匡, 藤野 昭典, 片井 修, 椹木 哲夫: “連続値入出力を扱うファジィ内挿型 Q-Learning の提案,” 計測自動制御学会論文集, Vol.35, No.2, pp.271-279 (1999)

[3] 前田 陽一郎: “ファジィ状態分割型 Q-Learning を用いたマルチエージェントシミュレーション” 第 *17*回ファジィシステムシンポジウム, pp.419-422 (2001)

[4] Y.Maeda and K.Makita, ”Modified Profit Sharing Method with Continuous States Divied by Fuzzy Rules”, The Second International conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2003), CD-ROM (2003)