マルチエージェントのQ－学習による協調搬送

旭川高専　　○中澤大輔　渡辺美知子　古川正志

要旨

複数のロボットが協調して物体を搬送する問題は、マルチエージェントの協調問題として考えることができる。本研究では、搬送物体を棒状の物体とし、隣のエージェントと自身のエージェントの棒までの距離および物体からの反力のみで搬送する協調行動を、Q－学習を採用して獲得する方法を提案し、その実験結果を報告する。

1. はじめに

複数の自律ロボットを用いた荷物の搬送問題は、マルチエージェントが協調して物体を押し運ぶ問題と考えることができる。このような問題の解決に、ここでは、信頼度と拡張性の有利さから、事前に各エージェントに行動を与えるような方法ではなく、それぞれのエージェントが分散し、自律的に各々の行動を決定できるような学習法を提案し、全体として一つの目標を達成する自律協調方法の有効性を検討する。

この協調問題をシミュレーションしたものとして Kimuro、Ohtsukaら[1]による研究があるが、この研究では荷物とする物体は円形であり、また物体を衝突の繰り返しによって移動させている。本研究では荷物を棒状の物体とし、自身と隣のエージェントの位置関係、およびロボットにひ垂直になる物体からの反力で現在の状態を表現し、現在置かれている状態に対する行動の評価を徐々に行動獲得の指標とするQ-学習を採用し、搬送問題を解決する。

2. 物体を多点で押す物理原理

本研究を学習のシミュレーションモデルとするためには、物体の動きに対するエージェントは物理現象の計算を満たすように設計する。すなわち、エージェントの加速度、速度、位置、力、反力はベクトルで表現される。よって以下の式が成り立つ。

（本論文中の太字はベクトルを表すものとする。）

$$\mathbf{F}_{agv} = m_{agv} \cdot \mathbf{a}_{agv} \quad (1)$$

$$\mathbf{F}_{obj} = \mu \cdot \Sigma \mathbf{F}_{agv} \quad (2)$$

$$\mathbf{a}_{obj} = \mathbf{F}_{agv} / m_{obj} \quad (3)$$

$$\mathbf{v}_{agv} = \mathbf{a}_{agv} \cdot t \text{ (物体を押していないとき)} \quad (4)$$

$$\mathbf{v}_{obj} = \mathbf{a}_{obj} \cdot t \text{ (物体を押していないとき)} \quad (5)$$

$$\mathbf{v}_{agv} = m_{agv} \cdot \mathbf{v}_{agv} / (m_{obj} + m_{agv}) \quad (6)$$
（物体を押しているとき）

$$\mathbf{v}_{obj} = m_{agv} \cdot \mathbf{v}_{agv} / (m_{obj} + m_{agv}) \quad (7)$$
（物体を押しているとき）

$$\omega_{agv} = M \cdot 12 / (m_{obj} \cdot L_{obj}^{2}) \quad (8)$$

$$\theta = \omega \cdot t \quad (9)$$

ここで、

m_{agv} ： AGV の質量 [g]
$\mathbf{F}_{agv}$ ： AGV の力 [N]
$\mathbf{V}_{agv}$ ： AGV の速度 [m/s]
$\mathbf{a}_{agv}$ ： AGV の加速度 [m/s^2]
m_{obj} ： 荷物の質量 [g]
$\mathbf{F}_{obj}$ ： 荷物が AGV から受ける力 [N]
$\mathbf{v}_{obj}$ ： 荷物の速度 [m/s]
$\mathbf{a}_{obj}$ ： 荷物の加速度 [m/s^2]
μ ： 摩擦係数
ω ： 荷物の角速度 [rad/s]
θ ： 荷物の角度 [rad]
M ： 各 AGV による棒にかかるモーメント [Nm]
L ： 荷物の長さ [m]
t ： 経過時間 [s]

とする。

3. エージェントの学習モデル

マルチエージェントの定式化と、ここで採用した学習方法であるQ-学習は、以下のように述べられる。

3.1 マルチエージェントモデル

AGV を以下のように定式化する。

$$A = \{ AGV_i ; i = 0,1,2,\ldots,Nagv \} \quad (10)$$

$$AGV_i = \{ m_i, \mathbf{p}_i, \mathbf{f}_i, \mathbf{v}_i, \mathbf{a}_i, s_i, g_i, Q_{igs} \} \quad (11)$$

ここで、A は AGV の集合を表し、AGV_i は個々のエージェントで表された自律ロボットである。また、AGV_i の各項 $m_i, \mathbf{p}_i, \mathbf{f}_i, \mathbf{v}_i, \mathbf{a}_i, s_i, g_i, Q_{igs}$ はそれぞれ AGV_i の質量、位置、力、速度、加速度、現在の状態を表す状態番号、状態に対し決定した加速度、状態とその Q 値を表す。

同様に荷物を以下のように定式化する。

$$OBJ = \{ m, \mathbf{p}, \mathbf{f}, \mathbf{v}, \mathbf{a}, \omega, \theta, L \} \quad (12)$$

ここで、OBJ の各項 $m, \mathbf{p}, \mathbf{f}, \mathbf{v}, \mathbf{a}, \omega, \theta, L$ はそれぞれ荷物の質量、位置、力、速度、加速度、角速度、角度、長さを表す。

3.2 Q-学習

Q-学習は C.J.C.H.Watkins によって提案された、ダイナミックプログラミングを基にニューラルネットワークから発展した強化学習法である。これは TD 法の発展型として考えることができ、状態と行動の組に対する評価を見積もることができる。

時刻 t の状態をs_t、その時の行動を g_t、その時点での Q 値を $Q_{t\,sg}$とし、行動g_tを実行した結果、時刻t＋1の時に状態s_{t+1}へ移行したとすると $Q_{t+1,sg}$は、次式により更新される。

$$Q_{t-1,sg} = (1-\alpha)Q_{t,sg} + \alpha[f_c + \gamma \max_g(Q_{t,sg})] \quad (13)$$

ここで、α は学習率、γ は減衰率、f_cは環境からの直接報酬、関数 max は行動集合 G の中から、その状態に対する最大の Q 値を得る関数である。

行動の選択は、Q 値をボルツマン分布とし、確率的に行動を選択する方法を用いる。行動確率 P_{sg}は以下の式で計算される。

$$P_{sg} = (\exp(Q_{tsg})/T) / \Sigma \exp(Q_{tsG})/T \quad (14)$$

ここで T は計算温度係数を表し、T が小さくなると Q 値の差が反映されにくくなる。

3．3　本実験モデル

本実験では AGV_i の状態 s_i を以下の 4 項目で2進的に表す。

$$s_i = \{ pow, bck, sid, frd \} \quad (15)$$

ここで、状態 s の各項 pow,bck,sid,frd はそれぞれ、荷物からの反力よりも AGV_i の力が大きいか、隣のエージェントよりも棒から遠ざかっているか、隣のエージェントと棒までの距離が等しいか、隣のエージェントよりも棒に近づいているかを表し、0または1の値を持つ。また、AGV_i が決定する加速度は 4 段階あるものとする。環境からの直接報酬は、以下の式で表される。

$$f_c = R_1 \cdot W_1 + R_2 \cdot W_2 \quad (16)$$

ここで、

$$R_1 = \begin{cases} \text{Reward} & \text{：前段階よりもゴールに近づいた場合} \\ \text{Penalty} & \text{：前段階よりもゴールから遠ざかった場合} \end{cases}$$

$$R_2 = \begin{cases} \text{Reward} & \text{：前段階よりもゴールに近づいた場合} \\ \text{Penalty} & \text{：前段階よりもゴールから遠ざかった場合} \end{cases}$$

$W_1 = R_1$に関する重み

$W_2 = R_2$に関する重み

である。

4．　数値実験シミュレーション

以上のように設計されたエージェントによって、Q-学習を用いた行動決定法の獲得について、その有用性を検証するために行った数値実験シミュレーションの結果を報告する。

4．1　実験条件

フィールドの大きさ	:	(640・480セル)(m^2)
AGV の数	:	3 台
荷物の数	:	1 個
AGV の重さ	:	628g
荷物の長さ	:	5m
荷物の重さ	:	500g
学習率 α	:	0.2
減衰率 γ	:	0.8
摩擦係数	:	0.5
一回の学習にかける 学習時間	:	2000 秒
学習回数	:	30000 回
Reward	:	0.4
Penalty	:	0.1
W_1,W_2	:	1

4．2　実験結果

上記の実験条件を用いた数値実験シミュレーションの結果を以下に示す。 図1は学習回数とゴールする確率の関係であり、図2は隣のロボットと並んでいて、荷物からの反力よりもロボットの力が小さい状態における、学習回数とQ値の変動である。

また、図3はロボットがゴールまでたどり着く軌跡を描いたものである。

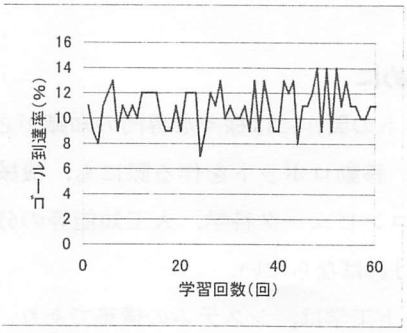

図1　学習回数とゴール到達確率の関係

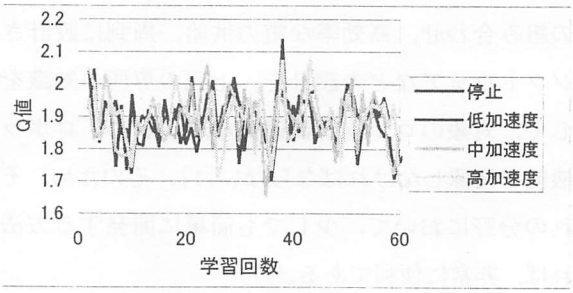

図2　隣のロボットと並んでいて荷物からの反力よりもロボットの力が小さい状態のQ値の変動

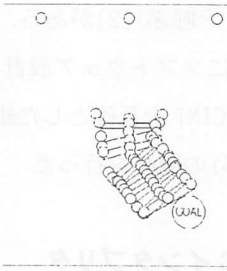

図3　ロボットと荷物のゴールまでの軌跡

5．　おわりに

本研究では Q-学習を用いた複数の自律ロボットによる荷物搬送問題の協調行動の獲得方法を提案した。今後は障害物を考慮した問題や、ロボット同士の衝突回避の問題、および、複数の荷物を取り入れた問題を検討したい。

参考文献

(1)Y.Kimuro,Y.Ohtsuka,H.Zha,T.Hasegawa,"Distributed Planning for Pushing Operation by Multiple Autonomous Robots" ,Intelligent Autonomous Systems Y.Kakazu et al.(Eds.) IOS Press(1998)454