

群ロボットの強化学習による棒押し問題の協調行動獲得

旭川高専

○中澤大輔

渡辺美知子

古川正志

要旨

複数のロボットが棒状の物体を目的地まで搬送する問題を、マルチエージェントの協調問題として取り上げ、強化学習によって協調行動を、獲得できることを既に報告した。本研究では、既に得られた知見を元に、実記モデルを対象としてオフラインで搬送の為の協調行動の知識獲得がQ-学習によって可能なことを報告する。

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の一つとして、マルチエージェントが協調して物体を押し運ぶ問題がある。このような問題の解決に、ここでは、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としてKimuro, Ohtsukaら¹⁾による研究がある。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。また、実機を用いて行った研究として、山田ら²⁾による研究がある。この研究においてロボットはトップダウン式のIF-THEN文によるルールを用いて行動を決定する。

本研究では、小型移動ロボットとして、VACCINEを前提とする。VACCINEは赤外線センサ2つと、フォトリフレクタセンサ4つを持ち、2つのDCモータで動作する。環境には障害物を設定し、荷物は棒状の物体とする。計算機上シミュレーションでは、実際に扱う小型移動ロボットのセンサ数、モータの動作等を再現し、各センサ値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、自律的に搬送問題を解決し、シミュレーション上で得られた状態-行動の関係を実機にプログラミングし、搬送問題を実際に解決することを目的とする。

2. 対象モデル

図1~2に、本実験で使用する実機モデルの概要を示す。

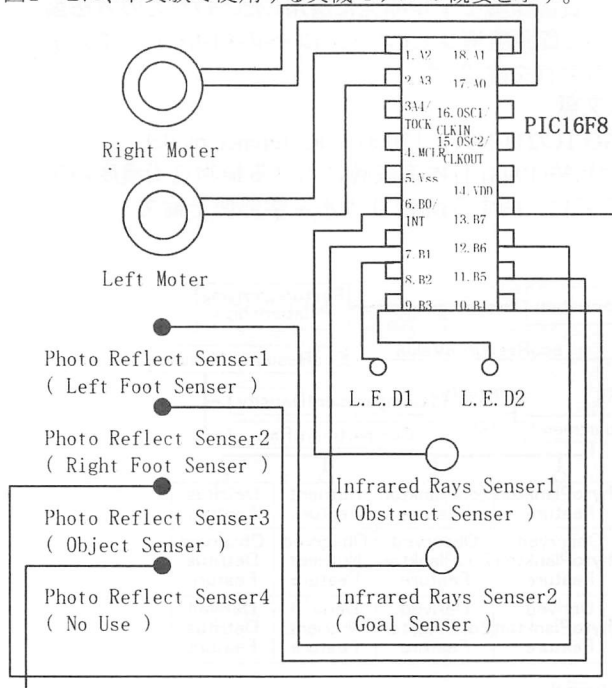


Fig.1 Vaccineのブロック図

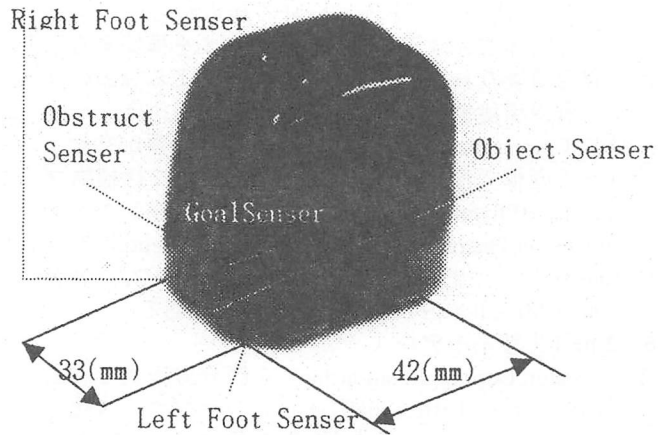


Fig.2 Vaccineの寸法及び、センサ位置

3. エージェントの学習モデル

計算機シミュレーション上の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。

$$A = \{ AGV_i; i = 0, 1, 2, \dots, N_{agv} \} \quad (1)$$

$$AGV_i = \{ m_i, p_i, f_i, a_i, s_i, g_i, Q_{igs} \} \quad (2)$$

ここで、AはAGVの集合を表し、AGV_iは個々の自律エージェントである。また、AGV_iの各項 $m_i, p_i, f_i, a_i, s_i, g_i, Q_{igs}$ はそれぞれAGV_iの質量、位置、力、行動番号、現在の状態を表す状態番号、状態に対し決定した行動番号、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。

$$OBJ = \{ m, p, f, v, a, \omega, \theta, L \} \quad (3)$$

ここで、OBJの各項 $m, p, f, v, a, \omega, \theta, L$ 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。

また、障害物OBS_jはその位置であらわされる。

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, s_t 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}, s_t は、次式により更新される。

$$Q_{t+1,sg} = (1 - \alpha)Q_{t,sg} + \alpha [f_c + \gamma \max_g(Q_{t,sg})] \quad (4)$$

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。

$$P_{sg} = (\exp(Q_{t,sg})/T) / \sum_g \exp(Q_{t,sg})/T \quad (5)$$

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

本実験のシミュレーションではAGV $_i$ の状態 s_i を以下の4項目で2進的に表す。

$$s_i = \{ \text{Goal, Obj, LFS, RFS, OBS} \} \quad (6)$$

ここで、状態 s の各項Goal, Obj, LFS, RFS, OBSはそれぞれ、ゴールセンサーの値、荷物センサーの値、左前足センサーの値、右前足センサーの値、障害物センサの値で、0または1の値を持つ。

また、AGV $_i$ が決定する行動はモータの動作を表し、前方向回転、停止、後方向回転のいずれかに決定される。AGVは2個のモータからなるので行動の種類は6種類となる。

環境からの直接報酬は、以下の式で表される。

$$f_c = R_1 \cdot W_1 + R_2 \cdot W_2 + R_3 \cdot W_3 + R_4 \cdot W_4 + R_5 \cdot W_5 \quad (7)$$

ここで、

- R_1 { 報酬値 (荷物を発見する)
ペナルティ (荷物を見失う)
- R_2 { 報酬値 (荷物がゴールに近づく)
ペナルティ (荷物を見失う)
- R_3 { 報酬値 (ゴールを発見する)
ペナルティ (ゴールを見失う)
- R_4 { 報酬値 (障害物から離れる)
ペナルティ (障害物にぶつかる)
- R_5 { 報酬値 (障害物に近づく)
ペナルティ (障害物にぶつかる)
- 0 (それ以外の状態)

$W_1 = R_1$ に関する重み
 $W_2 = R_2$ に関する重み
 $W_3 = R_3$ に関する重み
 $W_4 = R_4$ に関する重み
 $W_5 = R_5$ に関する重み
 である。

4. 数値実験シミュレーション

以上のようにモデル化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を検証するために数値実験シミュレーションの結果を実施した。

4.1 実験条件

数値実験の仮想環境を以下のように設定する。

フィールドの大きさ	:	(0.4×0.4) [m ²]
AGVの数	:	3 [台]
荷物の数	:	1 [個]
AGVの重さ	:	0.2 [kg]
荷物の長さ	:	0.1 [m]
荷物の重さ	:	0.2 [kg]

学習率 α	:	0.3	[・]
減衰率 γ	:	0.7	[・]
割引率	:	0.5	[・]
静摩擦係数	:	0.1	[・]
動摩擦係数	:	0.08	[・]
1回の学習にかかる			
学習時間	:	2000	[sec]
学習回数	:	13000	[回]
Reward	:	1	[・]
Penalty	:	-0.01	[・]
W_1	:	5	[・]
W_2	:	1	[・]
W_3	:	1	[・]
W_4	:	1	[・]
温度係数 T	:	0.2	[・]

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3は学習回数とゴールまでの時間の関係を表す。

また、図4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3から学習を学習を重ねるに従い、ゴール達成率が向上するのがわかる。

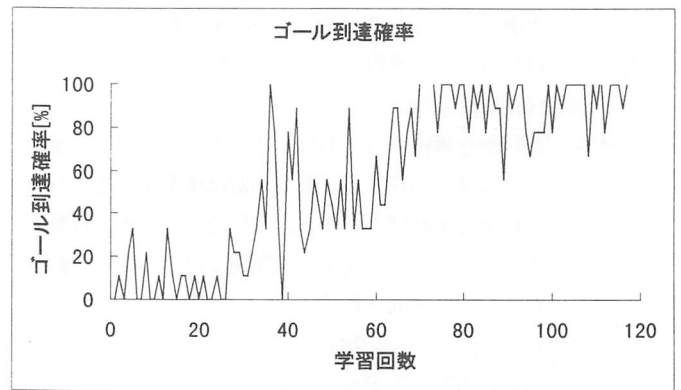


Fig.3 学習回数とゴール到達時間の関係

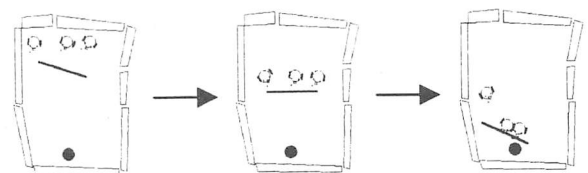


Fig.4 ゴールまでの軌跡

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障害物を有した荷物搬送問題の協調行動の獲得方法を提案した。今後は、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
- (2) Seiji Yamada, Jun'ya Saito, "Adaptive Action Selection without Explicit Communication for Multi-Robot Box-Pushing", Journal of the Robotics Society of Japan Vol. 17 No. 6 (1999) pp. 818-827