複数AGVのQ-学習による衝突回避行動の獲得

旭川高専　○池田将晴　渡辺美知子　古川正志

要旨

複数AGVの自律運転技術を確立するためには，他のAGVとの衝突を避け円滑に走行する必要がある．本研究では，逆に他のAGVとの衝突予想行動ををQ学習によって獲得することを行い，得られた知識により衝突回避を行う方法を提案し，衝突予想のシミュレーション結果を示し有効性を検証した．

1．はじめに

自律分散型生産システムの構築において，各ショップ間を自由に移動できるAGVの自律行動を獲得することは難しい問題である．このような問題には経路獲得問題と衝突回避問題があり，すでにQ学習を採用した経路獲得法を提案してきた．本研究では，AGV同士が様々な方向から衝突する可能性があることを考慮して，自由度の高い衝突回避行動を逆に衝突予想行動をQ学習により獲得することで実現する．

2．衝突行動獲得問題

衝突回避行動を獲得するにあたって，走行中のAGVを避ける走行ルートは多数存在する．衝突回避は衝突が予想される行動以外の経路であればよいから，本研究では，衝突行動を獲得する方法を取り扱う．衝突回避問題ではこの衝突行動を行わなければ目的が達成される．

2.1　AGVのモデルと前提条件

実験に用いるAGVのモデルを図1に示す．

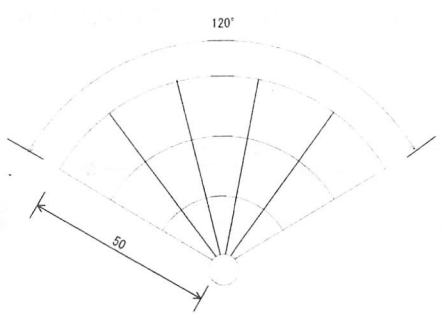

Fig.1 AGVのモデル

AGVには以下のような前提条件を設定する．

1) AGVは前方に5個のセンサを持つ．視野は等角度の小領域に15分割されている．
2) 前提としたセンサによって近づいた他のAGVの距離･方角を知覚できる．

2.2　問題の記述

前提条件をもとにAGVが最短時間で相手AGVに到達するような一連の行動を求めることである．

2.3　オートマトンによる学習モデル

AGVを1個のオートマトンと見なす．自AGVと相手AGVをそれぞれA_i，A_jとすると，各AGV，A_iは，

$$A_i = \{ I_i, O_i, S_i, F_i, G_i \} \quad (1)$$

の5項組で記述できる．ここで各記号は，

I_i ：入力　　S_i ：AGVの状態
O_i ：出力　　F_i ：状態遷移関数
G_i ：出力関数

である．また，A_jも同様である．

1) 入力　I_i

入力は以下のように構成する．

$$I_i = \{ T_i(t) , D_i(t) \}$$

ここで，tは時刻を表し，

$T_i(t)$ ：センサーがとらえたA_jの情報
$D_i(t)$ ：A_iの進行方向

である．

センサから得る情報$T_i(t)$は，

$$T_i(t) = \{ DIS_{Aj}, ANG_{Aj} \}$$

とする．ここでDIS_{Aj}，ANG_{Aj}は相手AGVとの距離，方角である．図1で示すように，これらの情報はAGVのセンサの小領域に相手AGVがいるか否かで定まる．

AGVの進行方向D_iは，図2に示すように

$$D_i(t) = \{ D \mid 0,1,\cdots,11 : \text{時計方向に対応} \}$$

とする．

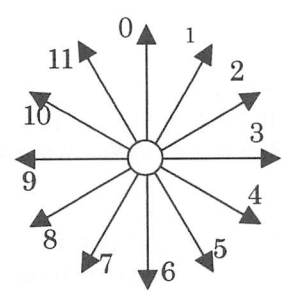

Fig.2 AGVの進行方向

2) 出力　O_i

出力は時刻 t における AGV の行動 $O_i(t)$ とする．

$O_i(t)$ = { h_left, l_left, forward, l_right, h_right

: 左急旋回，左旋回，前進，右旋回，右急旋回 }

AGV は５方向に移動することができる．

3) 状態　S_i

状態は，センサーの情報である．

$S_i(t)$ = { $T_i(t)$, $T_i(t-1)$, $T_i(t-2)$ }

衝突経路は一連の行動のつながりであること考え，時刻 t-2 までのセンサ情報を時刻 t における状態とする．

4) 状態遷移関数　F_i

AGV の状態は相手 AGV がセンサ内に入るたびに遷移する．

状態の遷移関数としてはQ学習で用いられるQ値を採用する．

$$Q_{i,t+1}(s_{i,t},o_{i,t})=(1-\alpha)Q_{i,t}(s_{i,t},o_{i,t})+\alpha\left[f_c+\gamma\max_{b\in O}Q_{i,t}(s_{i,t+1},b)\right] \qquad (2)$$

式(2)における環境情報からの報酬は，AGV のスタートから衝突するまでの行動で評価する．さらに衝突するまでの経路は一連の行動であることを考えて，バケットブリゲードにより報酬値 f_c を割引係数Γ，割引率 d_c を用いて以下のように与える．

$$\Gamma_{i+1} = \Gamma_i \cdot d_c \qquad (3)$$

直接評価 f_c = R(成功)・Γ_i

f_c = P(失敗)・Γ_i

ただし，i = 1,2,…T

R>0, P<0

ここで，T は過去に状態を遡る数である．

5) 出力関数　G_i

G_i は時刻 t における AGV の状態 S_i から行動 O_i 確率的に決める関数である．本実験では，G_i をボルツマン分布式(4)によって確率的に決定する．

$$O(s,o)=\frac{\exp\{Q(s,o)/T\}}{\sum_{b\in O}\exp\{Q(s,o)/T\}} \qquad (4)$$

3．　衝突回避問題

衝突問題より理想的な衝突行動が獲得されれば，衝突回避ではそれ以外の行動をとればよい．

しかし，衝突回避後は本来の目的地を目指す行動に移行する必要がある．

4．　実験結果

AGV が走行する環境を図 3 に示す．ターゲットの経路は一定方向に直線移動し，外周を障害物とする．Q 学習パラメータは，学習率α=0.8，減衰率γ=0.8，温度係数=1.0 とした．オフラインのQ学習では 17000 回の学習で最適解に収束した．図 4 に学習終了間際の AGV の行動を示す．また図 5 にQ学習の学習曲線を示す．

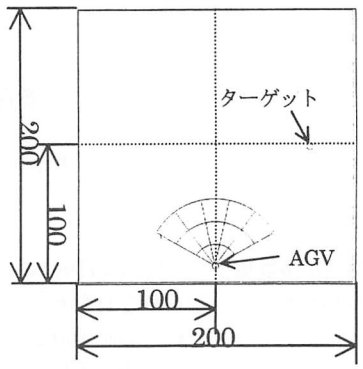

Fig.3 実験モデル

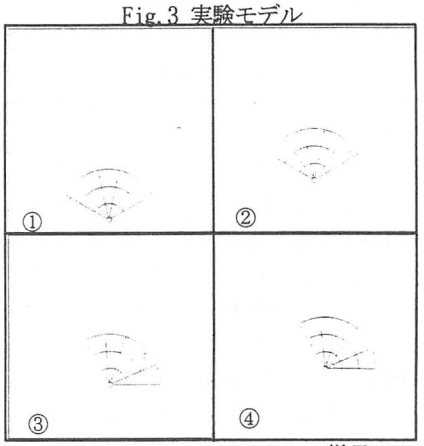

Fig.4 シミュレーションの様子

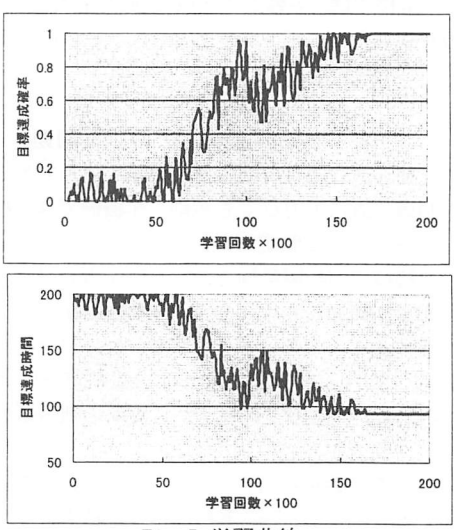

Fig.5 学習曲線

(上:達成時間収束グラフ，下:達成確率収束グラフ)

5．　おわりに

本研究では，AGV に単純なセンサーを持たせているのにもかかわらず，ターゲットの行動を予測し適確に行動させることができた．ただし，この強化学習に対するパラメータに関しては最適な値であるとは限らず，何らかの方法で最適値を求める必要がある．

参考文献

1) 池田将晴，渡辺美知子，古川正志 ： 仮想工場における多数 AGV の自律運転に関する研究，1999 年度精密工学会北海道支部講演会論文集(1999)44