

## AGV の強化学習による衝突回避行動の獲得—最適センサ数の決定—

旭川高専 ○池田将晴 渡辺美知子 古川正志

## 要旨

AGV を自律的に運転させるためには、目的地への最適経路の獲得と AGV 同士の衝突回避行動の獲得が必要不可欠となる。これまでセンサを装備した AGV の自律走行を可能とするため、強化学習の一つである Q 学習を 3 種類の実験環境で学習する衝突行動獲得法を提案してきた。本研究では AGV に衝突パターンを何種類学習させれば、あらゆる状況に対して正確な衝突予測が可能であるかを調べ、報告する。

## 1. はじめに

少品種変量生産を要求される現代において、ロバストかつ柔軟な生産システムを構築するための研究が行われている。特に自律分散型生産システムのモデル開発は多くの研究が行われ、同時に生産システムを構築するシステム要素の固有技術を開発することも重要な課題となっている。このような状況で無軌道型自動搬送車(Automatically Guided Vehicle, AGV)は工場内での移動自由度が高いため、自律分散型生産システムには不可欠な搬送装置となっている。本研究では、AGV を用いた搬送システムの構築を目的とし、工場内で起こり得る AGV 同士の衝突の回避行動の獲得を目指す。既に衝突行動の学習から回避行動を得る方法を提案しているが、あらゆる状況に対して回避ができるものではなかった。ここでは従来の研究を基に、様々な状況においても衝突回避を可能とする状況数(学習数)を調べ、報告する。

## 2. 衝突行動獲得問題

AGV が目的地に向かって自律走行する場合、事前に他の AGV や人間の行動知識がプログラムされていなければ衝突する危険性が生じる。従って衝突が予想される場合、AGV は自律的に衝突を回避するような行動が必要となる。本研究では、この知識の獲得を機械学習によって獲得する。一方、衝突回避行動は間接的に衝突行動の逆問題と捉えることができる。衝突を行う知識が得られれば、その逆の行動を衝突回避に際して行わせればよい。また、衝突問題はそれ自体多くの応用が考えられる重要な問題でもある。本研究では AGV の衝突回避問題をその逆問題、即ち衝突問題として取り扱い、知識の獲得に Q 学習を採用したアプローチを実施した。

## 3. 衝突行動獲得法

以下では衝突問題の知識獲得を行うモデルを述べる。

## 3.1 AGV のモデルと前提条件

取り扱う AGV は、1)前方に 5 個のセンサを持ち、2)センサ範囲内では 3 段階の距離を認識できる。よって視野は等角度の小領域に 15 分割されている。このセンサによって近づいた他の AGV の距離・方角を 15 分割の領域の 1 つとして知覚できる(図 1)。

## 3.2 問題の記述

上記の前提条件の下に、AGV が最短時間でセンサ領域に観測された AGV に衝突する一連の行動知識を求めることである。

## 3.3 オートマトンによる学習モデル

1 台の AGV をオートマトン  $A_i$  と表現すると、全 AGV の集合は集団オートマトン  $A$  を、

$$A = \{A_i : i=1, 2, \dots, n\} \quad (1)$$

とする。ここで  $n$  は AGV の総数である。

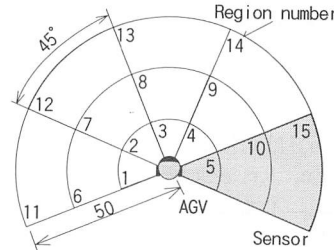


Fig. 1 AGV sensed areas

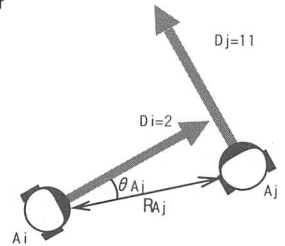


Fig. 2 Input information

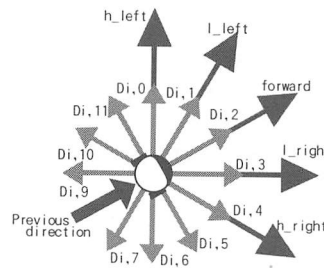


Fig. 3 Driving direction

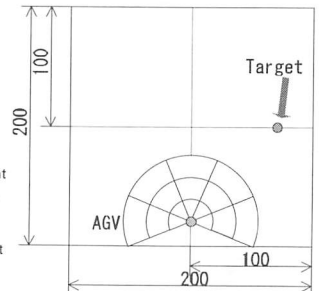


Fig. 4 A learning field

個々のオートマトン  $A_i$  を、

$$A_i = (I_i, O_i, S_i, F_i, G_i) \quad (2)$$

の 5 項組で記述する。ここで、 $I_i, O_i, S_i, F_i, G_i$  はそれぞれ、入力、出力、状態、状態遷移関数、出力関数である。

## 3.3.1 入力 I

ここからは衝突行動を行う AGV とそのターゲットとなる AGV をそれぞれ  $A_i, A_j$  とする。

入力は以下のように構成する。

$$I_i = \{T_i(t), D_i(t)\} \quad (3)$$

ここで、 $t$  は時刻を表し、

$T_i(t)$ : センサがとらえた  $A_j$  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、

$$T_i(t) = \{R_{A_j}, \theta_{A_j}\} \quad (4)$$

とする。ここで  $R_{A_j}, \theta_{A_j}$  は  $A_j$  との距離、方角である。図 2 で示すように、これらの情報は AGV のセンサ小領域に  $A_j$  がいるか否かで定まる。

AGV の進行方向  $D_i(t)$  は、図 3 で示すように、

$$D_i(t) = \{D_{ij} : j=0, 1, \dots, 11\} \quad (5)$$

とする。

## 3.3.2 出力 O

出力は時刻  $t$  における AGV の行動  $O_i(t)$  とする。

$$O_i(t) = \{h\_left, L\_left, forward, L\_right, h\_right\} \quad (6)$$

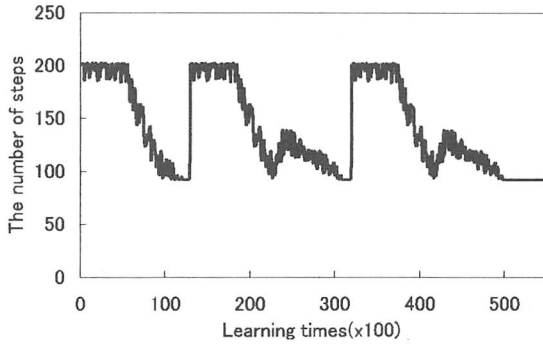


Fig. 5 The number of steps for collision vs. Learning times(r=3)

$h\_left, l\_left, forward, l\_right, h\_right$  はそれぞれ、左急旋回、左旋回、前進、右旋回、右急旋回である (図3)。

### 3.3.3 状態 S

状態  $S_i(t)$  は、一連のセンサ情報  $T_i(t)$  のグループである。

$$S_i(t) = \{ T_i(t), T_i(t-1), \dots, T_i(t-k) \} \quad (7)$$

獲得する経路は一連の行動であると考え、時刻  $t-k-1$  までのセンサ情報を時刻  $t$  における状態とする。ここで  $k$  は過去にさかのぼる数とする。

### 3.3.4 状態遷移関数 F

$A_i$  の状態は  $A_j$  がセンサ内で移動する度に遷移する。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。

$$Q_{i,t+1}(s_i(t), o_i(t)) = (1 - \alpha)Q_{i,t}(s_i(t), o_i(t)) + \alpha \left[ f_c + \gamma \max_{b \in O} Q_{i,t}(s_i(t+1), b) \right] \quad (8)$$

ここで、 $Q_{i,t}(s_i(t), o_i(t))$  は、状態-行動間にも与えられる Q 学習値、 $\alpha, \gamma, f_c, O$  はそれぞれ学習率、減衰率、環境報酬値、行動集合である。

### 3.3.5 環境報酬値 $f_c$

(8)式における環境報酬値からの情報は、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、パケットブリゲードアルゴリズムにより  $k$  ステップ前の報酬値  $f_c$  を  $f_c^{ik}$  として割引係数  $\beta (0 < \beta < 1)$  を用いて以下のように与える。

$$f_c^{ik} = \beta^k f_c, \quad f_c = \begin{cases} R & : \text{接近した場合} \\ P & : \text{その他の場合} \end{cases} \quad (9)$$

ただし、 $k=i, i-1, \dots, i-m$ ,  $k$  は成功報酬値  $R > 0$ , 失敗報酬値  $P < 0$  とする。

ここで、 $m$  は過去に状態をさかのぼる数である。

### 3.3.6 出力関数 G

$G_i$  は時刻  $t$  における AGV の状態  $S_i$  から行動  $O_i$  を確率的に決定する関数である。本研究では  $G_i$  をボルツマン分布式によって以下の式に基づいて決定する。

$$G(s, o) = \arg \operatorname{prob}_{O_i} \frac{\exp\{Q(s, o)/H\}}{\sum_{b \in O} \exp\{Q(s, b)/H\}} \quad (10)$$

ここで、 $H$  は温度係数である。

## 4. 数値計算実験

あらかじめ衝突点を設定し、そこへの衝突 AGV の進入可能範囲  $\theta$  とする。これを  $r$  分割したセクタを作成する。ついで、

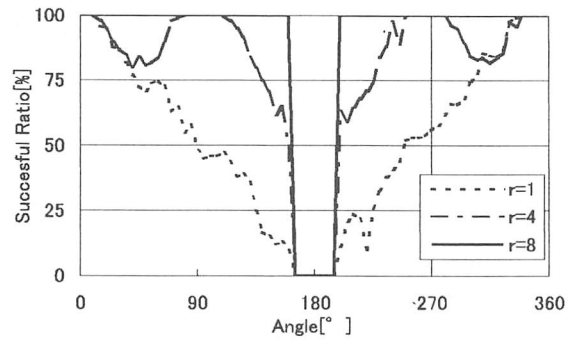


Fig. 6 Ratio of performance(r=1,4,8)

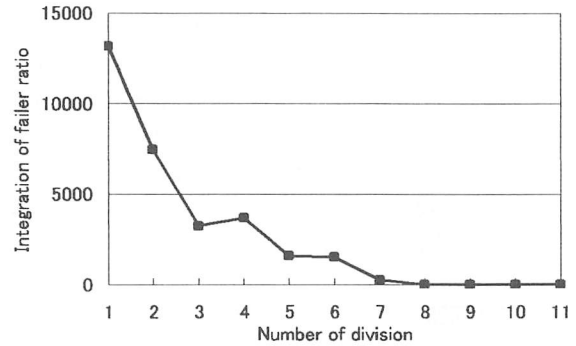


Fig. 7 Integration of failer ratio

各セクタの中心に衝突点に向かって AGV を走行させ衝突行動獲得実験を実行する。最適学習数は  $r$  を変化させて求める。このときの学習に、設定した学習パラメータをそれぞれ、Q 初期値=2.0,  $R=0.08$ ,  $P=-0.04$ ,  $\alpha=0.3$ ,  $\gamma=0.8$ ,  $\beta=0.5$ ,  $H=1.0$ ,  $m=8$ ,  $k=3$  とする。図5は  $r=3$  の時の追加学習による収束曲線を示している。図6は、 $r=1, 4, 8$  のときの学習に対して  $360^\circ$  からの進入衝突を行わせた時の 200 回での成功率を示した。これから分割数を増やす毎に全体の衝突成功率が上昇していくことが分かる。図7は、 $r$  に対する  $(1 - \text{成功率})$  の角度による積分値を示したものである。これから  $r=8$  以上の時、全衝突を可能とすることが判明した。

## 5. おわりに

本研究では、AGV が走行する際に生じる衝突回避問題について逆問題としての衝突問題を Q 学習による獲得法を提案し取り扱った。その結果、単純なセンサを持たせた AGV でもターゲットとの衝突を予測し回避行動を行わせることができた。またセンサ領域を一定にして最適学習数を調べたところ、ある程度の環境数を設定して学習を行わせることで、全方向に対して十分な衝突回避行動が取れることが実証された。これにより全方向をくまなく学習する必要がなく、学習時間の短縮につながる。と考える。

### 参考文献

- (1) A.G.Barto, S.J.Brake and S.P.Singh: Learning to Act Using Real-time Dynamic Programming, Artificial Intelligence, 72(1995)81.
- (2) 川上敬, 嘉数侑昇: クラシファイヤーシステムによる自律ロボットナビゲーションに関する研究, 日本機械学会論文集(C編)59,564(1993)2339.