群ＡＧＶの強化学習法による自律的意志決定

旭川高専　○小関秀夫　渡辺美知子　古川正志

要　旨

本研究では一方向のみＡＧＶ（Autoatically Guided Viecle）が移動できる機械工場において、複数ＡＧＶが各ＡＧＶ間の衝突を回避しながら自律行動するときの行動決定問題を取り扱う。この行動決定法として、各ＡＧＶが全体の周回完了時間を最短とするような強化学習法（Reinforcement Learning）に基づくモデルを提案する。また、数値計算実験より本モデルが有効である事を検証する。

１．はじめに

現在、工場での無人搬送手段としてＡＧＶが使用されている。工場内で複数のＡＧＶの行動を決定する場合、大きく分けて二通りの方法がある。一つは事前に全ＡＧＶのスケジュールを決定し与える方法であり、もう一つは各ＡＧＶに自律的意志決定をさせる方法である。本研究は後者の方法を採用し、強化学習法を用いて全ＡＧＶの行動時間を最小化する行動パターンの決定を強化学習を採用して行う方法を提案する。また、本方法の有効性を、数値シミュレーションにより検証する。

２．ＡＧＶモデル

本研究で採用した自動化工場を図１に示す。ＡＧＶに関する制約条件を、以下に記述する。

１）各ＡＧＶの速度は一定とする。
２）移動時間は、図１の１マス分を移動する時間を１単位時間とし、機械工作に要する時間はその５倍とする。
３）レーンはＡＧＶ１台が走行可能とし、並走はできないものとする。
４）レーンは一方通行とする。
５）各ＡＧＶの工程表は乱数により決定する。
６）ＡＧＶは、工程表に記された機械を全て回り出発点に戻った時点で作業を終了する。また、巡回順は自由とする。
７）ＡＧＶ間の衝突は回避する。

ＡＧＶの自律的行動決定方法として強化学習法(ＲＬＡ)を適用し、確率的にＡＧＶの行動を決定する。個々のＡＧＶは、添え字ｉを用いて以下のように記述される。

ＲＬＡｉ＝｛Ｓｉ、Ａｉ、ｒ、Ｐｉ、Ｑｉ｝　　（１）

Ｓｉ：ＡＧＶの走行レーン上の状態
Ａｉ：出力
ｒ：ＡＧＶの移動に体する応答
Ｐｉ：出力関数
Ｑｉ：強化値更新アルゴリズム

（ａ）ＡＧＶ、工場内等の状態（Ｓｉ）

工場内のＡＧＶは、位置及び作業の進行状況等の状態を８ｂｉｔｓのストリングで表し、これをＳｉ＝{S0、S1･･･Sn}とする。各々のｂｉｔの意味は以下のとおりである。

Ｓ０：前方４ｓｔｅｐ内の他のＡＧＶの有無
Ｓ１：前方４ｓｔｅｐ内の作業場の有無
Ｓ２：右方の進路の有無
Ｓ３：進行方向の壁の有無
Ｓ４：右方４ｓｔｅｐ内の他のＡＧＶの有無
Ｓ５：右方の作業場の有無
Ｓ６：全作業の終了判定
Ｓ７：ＡＧＶの状態（作業中または移動中）

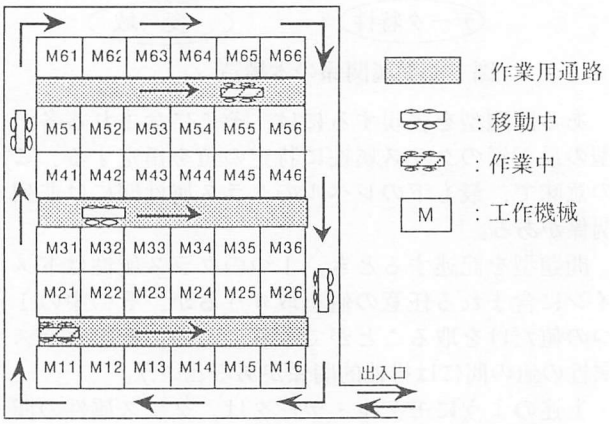

図１　機械工場モデル

（ｂ）出力（Ａｉ）

出力Ａｉは、ＡＧＶの行動パターンを記述する。行動は５種類あり、それを以下に示す。

ＡＣＴ１：停止　　ＡＣＴ２：直進
ＡＣＴ３：右折　　ＡＣＴ４：作業中
ＡＣＴ５：出口に向かう

ただし、いくつかの状態では行動を制限するので、あらかじめ行動を設定し上記の行動を禁止する。また、一方通行のため左折及び後退は行われない。

（ｃ）ＡＧＶの行動に対する応答（ｒ）

時間ｎと時間（ｎ－１）における最短の目的地（作業場または出口）までの距離ＤＩＳを用いて応答ｒを以下の式で決定する。ただし、作業中は応答はないものとする。

ｒ＝ＤＩＳ（ｎ－１）－ＤＩＳ（ｎ）　　（２）

このとき、ｒがプラスであれば成功、０またはマイナスであれば失敗とする。

（ｄ）出力関数（Ｐｉ）

出力Ａｉ（ＡＧＶの行動パターン）は、状態ｘｉにおける各行動の強化値のマトリックスＱｉから

$$P(S|a)=Prob\left(e^{Q_{S,a(t)}/T}/\sum_{b=0}^{m} e^{Q_{S,b(t)}/T}\right) \qquad (3)$$

から、確率的に決定する。ここでｂは行動を示している。ただし、作業場付近、衝突回避状態では各々作業中、停止が絶対的に選択される。

（ｅ）行動重み更新アルゴリズム（Ｑｉ）

強化学習の重みのマトリックスの学習、更新する。ただし、応答がないならば、学習はしないものとする。

$$Vx(t)=\max_{b=0}^{n}[Q_{x,b(t)}]$$

$$Q_{a,i(t+1)}=(1-\alpha)Q_{a,i(t)}+\alpha r(t)+\alpha\gamma Vx \qquad (4)$$

このときのｂは行動でＶｘはその状態のなかで一番大きい

Qの値である。また、αは学習率、γは割引率とする。

3．計算機実験

図1のような工場に対しＡＧＶの台数を5台、一つのＡＧＶが受け持つ作業数を5ケ所、$\alpha=0.0003$、$\gamma=0.8$、$T=0.05$として数値シミュレーションを行った。全体の作業数は25ケ所である。学習回数1000回とした結果を図2に示す。学習を重ねると全てのＡＧＶが作業を終える時間は160前後に収束するが、140から200の間で変動する。また、各ＡＧＶの学習状況を図3に、ある状態での強化値の変化を図4に示す。尚、1000世代での最短の作業所要時間は111であった。

4．考察

本実験での学習が収束している状態での作業時間が変化する理由として、以下のことが考えられる。

1）行動が確率的に決定されるので、必ずしも良い行動が選択されるとは限らない。このことは、逆に新しい環境に対して学習適応性を持つ。

2）ＡＧＶに与えられる作業場が複数ケ所に接している場合、ＡＧＶは移動時間にロスを殆ど被ることなく、複数の作業を実行できる。

3）ＡＧＶに与えられる作業場は、ランダムに選んでいるために作業場が近くなったり遠くなったり毎回変化する。

また、状態数は256通り用意したが、実際に使用されたのは62通りである。これは、62通りの規則生成を行えたことを意味する。

5．おわりに

本実験ではＡＧＶの自律意志決定問題に強化学習法を適用し、良好な結果を得ることができた。しかし、パラメータの調整が収束に影響することが分かった。今後の課題は、このパラメータの調整と、自動化工場のモデルの大きい場合、ＡＧＶや工作機械の性能の違いを持たせた場合についての検討をする必要がある。

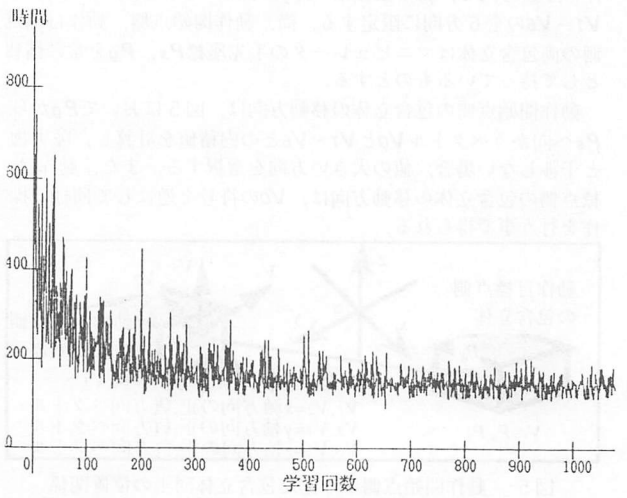

図2　ＡＧＶ全体の周回完了時間の収束

参考文献

1）Jeffery A. Clouse and Paul E. Utgoff;A Teaching Method for Reinforcement Learning,Machine Leaning (ML92)

2）畝見達夫；実例に基づく強化学習法、人工知能学会誌 Vol.7 No.4 July 1992

3）渡辺他；群ＡＧＶのＳＬＡによる自律的意志決定、1994年度精密工学会北海道支部学術講演会論文集、1994

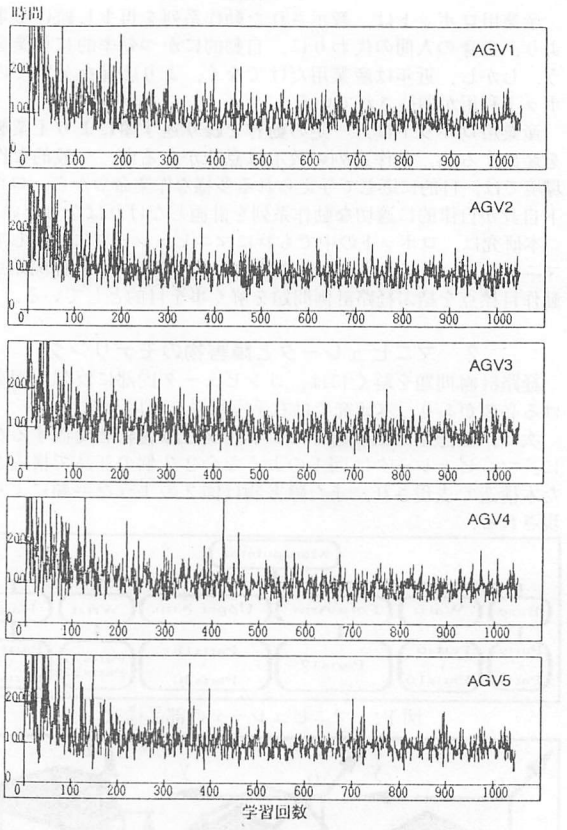

図3　各ＡＧＶの周回完了時間の収束

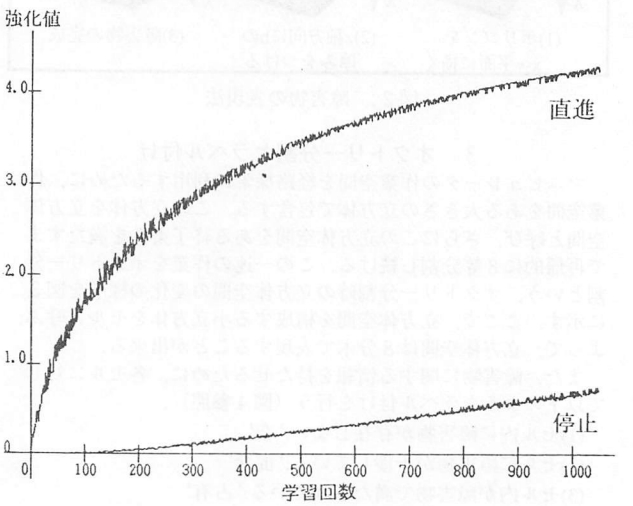

図4　01100110の状態におけるＡＣＴ1とＡＣＴ2の行動の強化値の様子