

On-line Profit Sharing による複数 AGV の自律走行 Autonomous Driving of Multiple AGVs by Use of On-line Profit Sharing

村田 賢太[†]
Kenta MURATA

吉井 伸一郎[†]
Shinichiro YOSHII

渡辺 美知子[‡]
Michiko WATANABE

古川 正志[‡]
Masashi FURUKAWA

嘉数 侑昇[†]
Yukinori KAKAZU

Controlling AGVs driving is important to operate FMS efficiently. This research represents a method of autonomous driving for multiple AGVs equipped with FMS. FMS model has a two-way driving lane for AGVs and neither passing nor encounter between AGVs is allowed. On-line Profit Sharing is applied to AGV model to have a capability of making driving decision by itself. Numerical experiments show that AGV can learn proper driving actions and operate its run in the shortest possible distance.

1. はじめに

複数台の自動搬送車 (Automatically Guided Vehicles, AGVs) を用いた多品種少量生産システム (FMS) が加工機械の高い稼働率を上げるためには、AGV がどのように走行するかを定めるスケジューリングが大きなウェイトを占めている。このようなスケジューリング問題は一般に NP 完全問題に属する複雑な組み合わせ最適化問題となり、厳密に解くことはほぼ不可能である。このような問題を厳密に解くかわりに、古川等 [1] は確率的学習オートマトンによる自律走行型 AGV を稼働させる方法を採用し、5~20 台の AGV に関して総 AGV の周回完了時間を最小化するような走行スケジューリングを可能としている。しかし、山本 [2] はこの研究の問題設定は走行通路が一方通行であるため、実問題に適用できないとしている。その上、「出口の方向へ向かわせる」という行動が AGV の行動集合に存在しているため、AGV は工場から出る事を真に学習しているわけではない。

本研究では、両方向に移動可能な走行通路を持つ工場内自律走行 AGV を稼働させる手法を提案する。行動は単純な移動のみを設定し、観測可能空間は古川 [1] よりも小さく設定した。状態は部分観測となることから、学習アルゴリズムとしては On-line Profit Sharing [3] を採用した。ここで提案された方法の有効性を数値計算実験で検証した結果、走行スケジュールの獲得に関して有効性が確認された。また、古川 [1] による結果との比較によって、一方通行という制限には妥当性がある事が分かった。

2. 複数 AGV の自律走行問題

本研究は図 1 に示されるような両方向走行通路を持つ FMS 工場を仮定する。工場内は加工機械が走行通路に沿って並べられているものとする。機械は 1~10 の番号を持っている。そして、同じ番号を持つ 2 つの機械は同じ仕事を行うものとする。各 AGV はそれぞれ指定された番号を持つ加工機械を両方向走行通路に沿って周回しなければならない。

本研究で取り扱う問題で与えられる制約条件を以下のように設定する。

- (1) 走行通路は両方向に進行可能である。
- (2) 各 AGV は走行通路上で擦れ違ふことはできない。
- (3) 加工機械はパッファを持ち、AGV を 1 台だけ収納できる。この AGV は、機械が加工可能な被加工物を搬送し

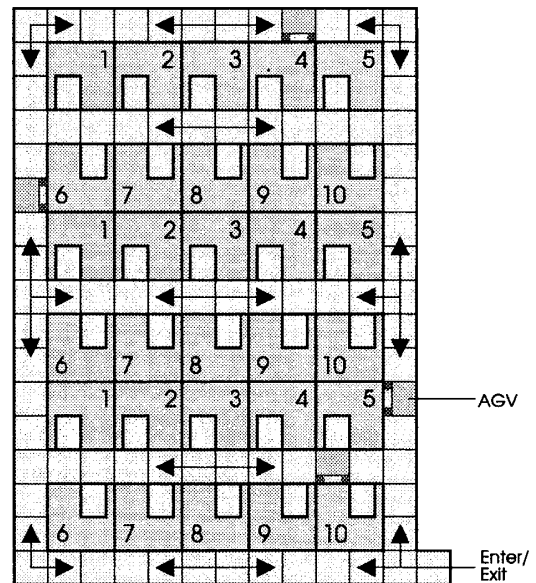


Fig. 1 An FMS model with driving lanes for AGVs

ている必要はない。

- (4) 各 AGV は単位時間 t_s 毎に出発する。

3. 自律走行学習 AGV モデル

3.1 On-line Profit Sharing

On-line Profit Sharing は行動優先度推定型の強化学習アルゴリズムであり、適格度トレース [4] を応用した信用トレースを用いることで Profit Sharing [5] をオンライン型に変形し、有限のメモリ容量で実行できるように改良したアルゴリズムとして Matsui [3] によって提案された。

On-line Profit Sharing では、行動優先度 P と同じ要素数の信用トレース c を持つ。信用トレース c は各状態行動対に対する信用割当ての割合を表わす数値である。これは、Profit Sharing における状態行動対の履歴の代わりに、その信用割当ての割合だけを記憶することに相当する。

各時間ステップ t において、すべての状態行動対の信用トレースは、割引率 γ だけ減り、そのステップで訪問された状態

[†] 北海道大学大学院情報科学研究科

[‡] 旭川工業高等専門学校

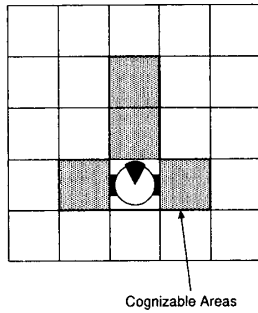


Fig. 2 Cognizable areas for AGVs

s_t と行動 a_t に対応する信用トレース $c(s_t, a_t)$ は 1 増える。

$$c_t(s, a) = \begin{cases} \gamma c_{t-1}(s, a) + 1 & (s = s_t \wedge a = a_t), \\ \gamma c_{t-1}(s, a) & (\text{otherwise}) \end{cases} \quad (1)$$

そして、各ステップにおける優先度の増分を

$$\Delta P_t(s, a) = r_{t+1} c_t(s, a) \quad (\forall s, \forall a) \quad (2)$$

で計算する。ここで r_{t+1} は $t+1$ ステップにおける報酬値である。

On-line Profit Sharing は、エピソードの中間に与えられる報酬が全てゼロの場合、通常の Profit Sharing に一致する事が証明されている。

3.2 AGV の状態

AGV の走行単位を 1 セルとする。走行通路上の AGV はセンサを持ち、前方 2 セルおよび左右 1 セルの領域を知覚可能とする (図 2)。これらセンサにより得られる工場の可視部分の状態に加え、

- AGV が向いている方角 (東西南北)
- タスク終了フラグ
- 過去 4 回分の行動履歴

を合わせて AGV の状態とする。

AGV のセンサは、セルの状態として次の状態を知覚する。

- AGV の走行経路が存在
- 他の AGV が存在
- 搬送中の被加工物を加工可能な加工機械が存在
- 壁が存在
- 出口が存在

加工機械として判定されるセルは加工機械のバッファの部分だけであり、その他の部分は壁として判定される。

3.3 AGV の行動

AGV が取る行動は

- 前への移動
- 右への移動 (右旋回 → 前進)
- 左への移動 (左旋回 → 前進)
- 後ろへ旋回

の 4 種類とする。

3.4 AGV の行動選択

AGV は次のルールによって行動選択を行う。

- (1) 前、右、左の各センサの全てが壁もしくは他の AGV を知覚している場合「後ろへ旋回」を選択 (袋小路からの脱出)。

- (2) 前方のセンサが工場の出口を知覚している場合「前へ移動」を選択。
- (3) 右のセンサが工場の出口を知覚している場合「右へ移動」を選択。
- (4) 左のセンサが工場の出口を知覚している場合「左へ移動」を選択。
- (5) 優先度 P に基づき、重み付きルーレット行動選択を行う。

このルールは、上から順に適用される。ここで重み付きルーレット行動選択は、次式より確率的に行われる行動選択である。

$$\Pr(s, a) = \frac{P(s, a)}{\sum_{a' \in \mathcal{A}(s)} P(s, a')} \quad (3)$$

ここで、 $\Pr(s, a)$ は状態 s で行動 a を選択する確率、 $\mathcal{A}(s)$ は状態 s で選択可能な行動の集合であり次の条件を満たすこととする。

$$\forall s, \{ \text{後ろへ旋回} \} \notin \mathcal{A}(s) \quad (4)$$

つまり、「後ろへ旋回」行動は重み付きルーレット行動選択では決して選択されない。

3.5 報酬値

本研究では最初の AGV が工場に入ってから最後の AGV が工場から出るまでの 1 連のエピソードを最適化することを目的としている。しかし、本研究で扱う問題は、ひとつのエピソード中に

- (A) 与えられた順序で加工機械を探索するタスク
- (B) 工場の出口へ向かうタスク

の 2 種類のタスクを含んでいる。ここで、前述した AGV の状態から、この 2 つのタスクは「タスク終了フラグ」によって区別できる事が分かる。本研究ではこの特性を利用し、報酬を与えるタイミングを

- (a) 目的の加工機械へ到達した時
- (b) 工場の出口へ到達した時

の 2 つに設定し、どちらの場合においても報酬値として +100 を与えることとする。その他の場合については常に報酬を 0 とする。

4. 数値計算実験

4.1 実験条件

図 1 に示されるような両方向走行通路を持つ工場モデルを設定し、学習実験を行なった。本実験では各 AGV が周回する加工機械数を 5 台とした。これら 5 台の加工機械は一様乱数によって無作為に選択した。また、AGV の台数は 5 台とし、加工機械の加工時間は $5t_s$ とし、工場内のセル間の移動時間は t_s と設定した。このとき、すべての AGV が指定された加工機械を周回し、工場の出口へ到達するまでを 1 回のエピソードとし、ついで先に決定した 5 台の加工機械を周回するエピソードを繰り返し行う実験を行った。On-line Profit Sharing における信用トレース c の更新式 (1) で用いる割引率は $\gamma = 0.8$ とした。

4.2 実験結果

乱数の種を変更して 20 回行った実験結果の平均を、横軸をエピソード数、縦軸をステップ数としたグラフとして図 3 に示す。グラフより、5 台の AGV は 5 台の加工機械を周回し工場を出るまでの必要ステップ数を減少させる方向へ学習していることが分かる。

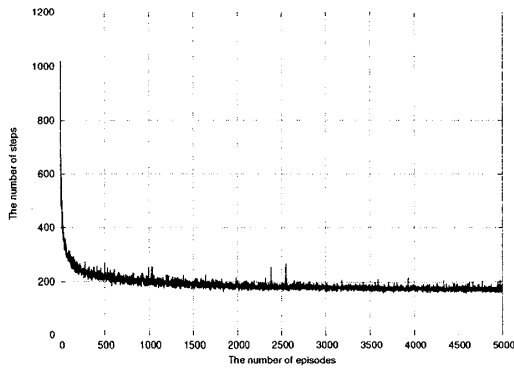


Fig. 3 A learning transition curve for 5 AGVs by OnPS

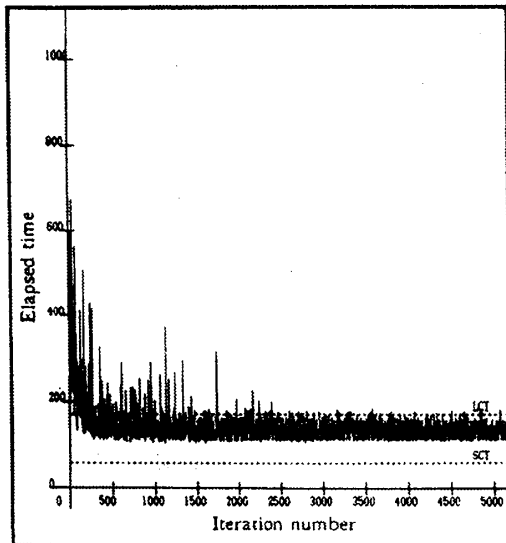


Fig. 4 A learning transition curve for 5 AGVs by SLA[1]

また、図4に文献[1]において行なわれた確率的学習オートマトンによる一方通行通路を持つ工場を仮定した実験結果を示す。両結果を比較すると、文献[1]の結果の方が約50ステップ程度良い解を得ている事が分かる。

5. 結論

両方向走行通路を持つ工場モデルにおける自律走行学習AGVモデルを作成し、各AGVが自律走行を行えるような方法を示した。これより、以下のことがまとめられる。

- (1) 自律走行学習AGVモデルをOn-line Profit Sharingでモデル化する方法を示した。
- (2) On-line Profit Sharingによるモデルに関しては、数値計算実験から総AGVの周回完了時間を最小化するような走行スケジューリングを、学習に基づいた確率的な意思決定で可能である事を示した。
- (3) 一方通行通路下における同様の実験結果[1]と本研究における実験結果を比較した。その結果、本研究の問題設定において移動方向の自由度が学習に及ぼす影響は50ステップ程度であることが分かった。

最後の結論から以下のことが言える。すなわち、エージェントの行動に関する自由度を高くすると状態数が増加するため、学習は進行しにくくなる。そうして得られた結果が、行動に制限がある場合と比較して多少なりとも悪いのであれば、無理に自由度を高くする必要はない。本研究で扱った問題にあてはめれば、工場の通路が一方通行であることは、AGVの学習に有利に働くような制限になっていると考える。さらに、現実世界に一方通行道路が存在している事を考慮すると、工場内の通路が一方通行であることは、大した制限にはならないとも考えられる。

参考文献

- [1] 古川正志, 渡辺美知子, 嘉数侑昇. 確率的学習オートマトンによる複数AGVの自律的走行. 精密工学会誌, Vol. 62, No. 2, pp. 260-264, 1996.
- [2] 山本秀彦. 「21世紀のリーディング生産技術」特集にあたって. 計測と制御, Vol. 42, No. 7, pp. 533-537, 2003.
- [3] Tohgoroh Matsui, Nobuhiro Inuzuka, and Hirohisa Seki. On-line profit sharing works efficiently. In Vasile Palate, Robert J. Howlett, and Lakhmi Jain, editors, *Proceedings of the 7th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES 2003)*, Vol. 2773 of *Lecture Notes in Artificial Intelligence*, pp. 317-324, Oxford, UK, September 2003. Springer-Verlag.
- [4] Richard S. Sutton and Andrew G. Bart. *Reinforcement Learning: An Introduction*. The MIT Press, 1998.
- [5] J. J. Grefenstette. Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms. *Machine Learning*, Vol. 3, pp. 225-245, 1988.

