413

強化学習における状態圧縮の研究
－ＡＧＶ自律走行問題の事例－

旭川高専　○平井幸基　渡辺美知子　古川正志

要　旨

一方向を走行するＡＧＶの自律走行を可能とする強化学習法を報告した．この結果を検討すると，多くの状態が同じ強化学習結果をもつことが判明した．本研究では，こうした状態を一つの状態空間とできるような状態圧縮法を報告する．

１．はじめに

これまでに一方向走行レーンをもつFMS工場におけるAGV（自動搬送車）の自律走行法として強化学習に基づく方法を報告してきた[1),2)]．これらの学習結果を検討すると，いくつかの状態が同じ行動の学習を行っていることが観察された．実際の人間の学習行動では，状態空間すべてを学習しているとは考えられず，いくつかの状態に関する学習で大体の行動が行えることも推測できる．

これらを考慮に，本研究ではAGVの自律走行を限られた状態数に基づく学習によって行可能であるかを，状態空間にGAのスキーマ理論である Don't careとHamming距離を導入して調べることを行った．

２．一方向走行レーンのAGV問題

本問題では一方向走行レーンをもつFMS工場を取り扱う[1)]．工場の内部は加工機械が走行レーンに沿って並べられる．各工作機械を

$$M=\{M_{x,y};x=1,2,\ldots,u\ \ y=1,2,\ldots,v\}$$

と置く．走行レーンを走るAGVを

$$A=\{A_i;i=1,2,\ldots,n\}$$

とする．いま，AGV，Ai，が行き先を指定された加工機械の集合を

$$M=\{A_{x,y};x\in X(A_i),y\in X(A_i)\}$$

と置く．ここでX(Ai)，Y(Ai)はAGV，Ai，の行先の加工機械の添え字集合である．ここで取り扱う問題は，「全AGVが操車場から出発し，すべての行先をまわり，操車場に戻る周回完了時間を最小にするようにAGVが行動を学習する」ことである．AGVには以下の様な制約を設定した．

１）走行レーンは一方向にのみ進行可能である．また，周回を可能とする．

２）AGVは走行レーン場での追い越しを禁止する．

３）各加工機械はAGVの待避バッファーをもち，一台のみ収納可能である．被加工物はAGVからこの待避場で加工機械に渡され，指定された加工が終了するまで待機する．

４）AGVは待避バッファーのAGVを追い越し可能である．

５）各AGVはAGV間の衝突を回避する．

６）AGVの移動単位時間を用いて被加工物の加工時間を設定する．

７）AGVは指定された加工機械をすべて巡回すると，速やかに操車場に戻り，移動を中止する

３．AGVのQ-学習モデル

各AGVは(I,O,S,f,g)の５項組から定義される．ここで，I,O,S,f,gはそれぞれ入力，出力，状態，状態遷移関数，出力関数を示す．

３．１　入力　I=r

時刻nと時刻(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を計算して応答rを以下の式で決定する．但し，作業中については応答はないものとする．

$$r=DIS(n-1)-DIS(n)$$

３．２　出力　$O=\{A_k;\ k=0,1,\ldots,4\}$

出力A_kは，AGVの行動パターンを記述する．行動は５種類あり，それらは以下の行動をとる．A_0：停止，A_1：直進，A_2：右折，A_3：作業中，A_4：出口に向かう．

但し，通路が一方通行のため，左折および後退は行わない．

３．３　状態　$S=\{\wedge S_j;\ S_j=0,1\}$

行場内のＡＧＶは，位置および作業の進行状況等の状態を８ビットのストリングで表わす．各ビットS_iの意味は以下の通りである．S_0：前方４step内の他のAGVの有無，S_1：前方４step内の指定加工機械の有無，S_2：右方の通路の有無，S_3：進行方向の壁の有無，S_4：右方４step内の他のAGVの有無，S_5：右方の指定加工機械の有無，S_6：全作業の終了判定，S_7：ＡＧＶの状態（作業中または移動中）．

3．4　状態遷移関数f(I,S,O)

AGVの学習を実行するために，状態行動遷移行列を状態遷移関数f(I,S,O)として採用する．すなわち，

$$f(I,S,O)=[Q_{j,k}(I)]$$

とする．ここで$Q_{j,k}(I)$は状態jにおける行動A_kのQ値を示し，3．6に述べる学習によって更新される．

3．5　出力関数g(O)

出力Oは，Q値を確率に変換して行動を決定する．これは，j番目の状態に対して

$$g(O)=\mathrm{Prob}\left(\frac{e^{Q_{jk}/T}}{\sum_{b=0}^{4} e^{Q_{jk}/T}}\right)$$

確率的に選択される．

3．6　Q-学習

Q-学習では，Q値の更新によって学習をおこなう．この更新は，

$$Q_{jk}(t+1)=(1-\alpha)Q_{jk}(t)+\alpha[r+\gamma\{\max Q_{jk}\}]$$

で行う．α，γは学習係数で$1>\alpha>0$である．

4．状態圧縮探索アルゴリズム

4．1　Don't careによる圧縮表現

図1(A)のように，モデルとなる状態ビット列にDon't careをふくませ，一つのモデルで複数の状態に一致させる．

4．2　ハミング距離による圧縮表現

図1(B)のように，比較する状態とモデルとの，異なるビットの数がある許容範囲内ならば同一視する．完全に一致した場合も含む．

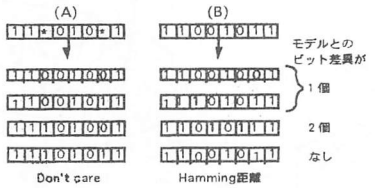


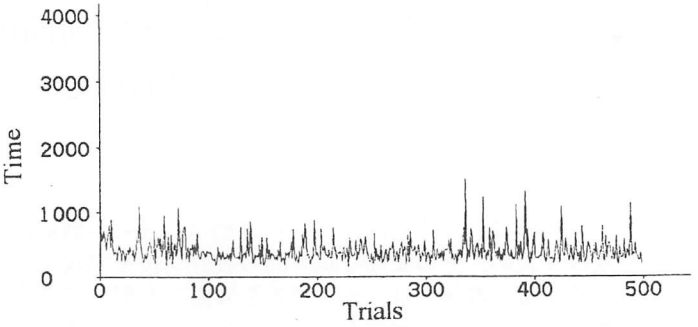
Fig.1 Representation on state compression

4．3　状態探索アルゴリズム

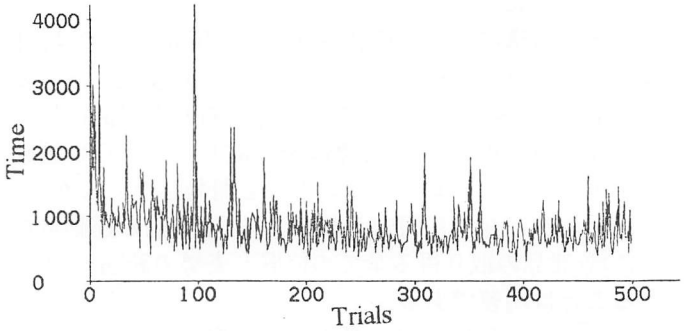
状態探索するのに，まずAGVごとに20個のランダムな状態モデル（これをクラスタと呼ぶ）をつくる．そして通常のQ-学習によるAGV走行を一定数周回させ，クラスタのヒット回数と周回時間を検出する．ここで，最も周回時間の少ないAGVの全クラスタを，ほかのAGVにコピーし，さらにヒット回数順にソートする．ヒット回数の少なかったものは新規クラスタに置き換える．ここまでの過程を繰り返すことにより，クラスタを進化させる事ができる．

5．数値計算実験

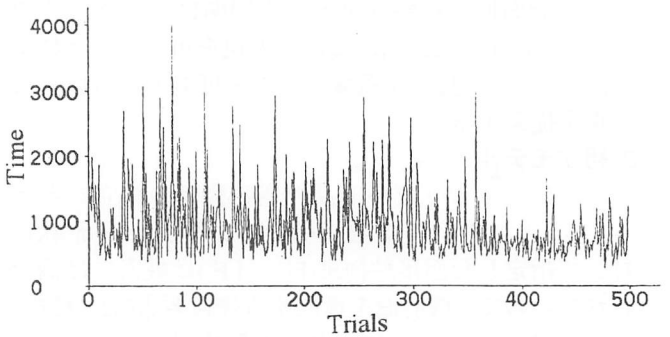
今回実験ではAGV5台，各AGVの作業5回とし，500回ループさせた．クラスタについては10回に1回更新を行っている．結果を図2に示す．



(A) Conventional Q-Learning



(B) Don't care method



(C) Hamming distance method

Fig.2 Learning curve

6．おわりに

本研究では，Don't care法とハミング距離法の2方法で状態圧縮を試みた．通常のQ-学習による結果と比較すると，状態圧縮を行った場合収束に時間がかかり，さらに若干結果が悪化しているが，それでも十分効果があると考えられる．今後，状態探索の方法や圧縮理論に関してさらに改善する必要があると思われる．

参考文献

(1) 古川，渡辺；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Vol.62No.2(1996)260.

(2) 古川，渡辺；一方向走行レーンをもつFMS工場における複数AGVの運行スケジューリング，日本機械学会論文集62巻595号(1996-3)407.

(3) 古川，渡辺；一方向走行レーン上のAGV走行スケジューリング，第8回自律分散システム・シンポジウム資料(1996)95.