Q 学習による多数 AGV の自律搬送に関する研究*

渡辺美知子** 古川正志** 木下正博*** 嘉数侑昇***

Acquisition of Efficient Transportation Knowledge by Q-learning for Multiple Autonomous AGVs and Their Transportation Simulation

Michiko WATANABE, Masashi FURUKAWA, Masahiro KINOSHITA and Yukinori KAKAZU

This paper reports on simulation results of multiple AGVs transportation in a virtual factory. AGVs drive on their ways autonomously. To do so, AGV learns paths between machining cells, and automated warehouse and machining cells by use of Q-learning before simulation. The fast learning methods, which take a consideration in failure learning, are proposed and the efficient learning method is found out. After the paths are learnt, transportation simulation for workpieces by autonomous AGV driving is performed. Results show that AGVs can transport the workpieces from the warehouse to machining cells and transport them back to the warehouse without any pre-programming and instructions, if enough number of AGVs are set in the factory. However, it is observed that deadlock phenomenon is observed very few times during transportation.

Key words: autonomous distributed production system, Q-learning, reinforcement learning, AGV, autonomous driving

1. 緒 言

多変種少量短納期の製造を必要とされる時代に入り、新しい生産システムの構築が求められている。こうした状況の中でアジャイル生産システム (agile production system)¹⁾、ホロニック生産システム (Holon production system)²⁾及び生物型生産システム (bionic production system)³⁾などがここ 10 年の間に提案され、積極的な研究が実施されている。こうした生産システムは、従来の効率優先に基づいた中央集権型から自律分散型の生産システムを目指していると考えられる。自律分散型生産システムは一般に冗長という欠点をもつが、前者に比べてシステムの拡張・縮小・変更等が容易に可能である機敏性・柔軟性の利点から注目を浴びている。本研究においても上述のような自律分散型の生産システムを確立するための基礎技術開発を目的とする。

自律分散型生産システムの構築では、無軌道型の無人搬送車 (automatically guided vehicle; AGV) がその自律性に大きな自由度をもち、素材、半製品、製品等の搬送を柔軟性高く行う重要な役割を担っている。しかしながら、こうした自律性を AGV に与えようすると、多くの状況に応じた意志決定を必要とする。そのような意志決定には、搬送物 (仕事) の決定、目的地の選択、現在地から目的地までの経路の決定、走行中の障害物の回避、緊急時の対応、等が含まれる。従来の工場は、こうした搬送装置の行動をコンベア型の搬送装置か有軌道型の AGV によって、予め定めたスケジュール通りに稼働させる中央集権的生産システムであると言える。本研究では、現在地から目的地までの経路の決定に Q 学習を採用する。Q 学習は経路をはずれても自律的に経路に復帰できる性質をもつためである。Q 学習による経路獲得の方法としては Watkins⁴⁾がすでに報告を行っている。また、著者らも簡易

SDM を採用した Q 学習による経路獲得法⁵⁾を提案している。しかし、Watkins による Q 学習を直接採用した場合、学習時間が長大になる。これは Q 学習が環境からのフィードバック情報に成功報酬のみを採用するためである。Q 学習のような強化学習の場合、環境からの成功報酬は初期の場合ほとんど皆無である。そこで、本研究ではこのような経路獲得に失敗から学ぶ強化学習を導入する。すなわち、三種類の逆強化学習法を導入し、数値計算によって最も効率のよい失敗を取り入れた強化学習法を見つけ出す。ついで、この経路学習法を採用した経路学習済みの多数 AGV を用いて工場内の自律搬送シミュレーションを実施する。このシミュレーションにおいては、自動倉庫、工作機械群、AGV をエージェント群とみなしたモデルを採用する。これは AGV に相当するエージェントが工場内を歩き回り、自動倉庫、工作機械群に相当するエージェントが仕事を指示する信号を周囲に発信している時にそれを見つけたし、搬送を実施するモデルである。このようなモデルは一見冗長であるが、数値計算シミュレーション結果では、ある数のエージェントが導入されればそれ以上のエージェントの数が投入されたモデルとほとんど同じ搬送時間となることが判明した。このような搬送を実施する時に、シミュレーションの最中に時折デッドロック現象が観られた。これについても考察する。

2. 失敗に基づく Q 学習経路獲得

ここでは、失敗から学ぶ 3 種類の Q 学習を提案し、これらに Watkins のレーストラック問題⁶⁾を適用した数値計算実験を示す。

2.1 Q 学習

Q 学習は、現在から将来得られるであろう利得を推測した期待値である Q 値を行動決定の評価とした強化学習法である。得られた Q 値は、対象とする問題がマルコフ決定問題に属するとき、Watkins が確率的ダイナミックプログラミングの最適性の原理から Q 値が最適となりかつ収束することを証明し

* 原稿受付 平成 13 年 4 月 2 日

** 正 会 員 旭川工業高等専門学校 (旭川市春光台 2-2-1-6)

*** 正 会 員 北海道大学大学院工学研究科 (札幌市北区北 13 西 8)

ている。自律搬送を行う経路獲得に Q 学習を採用する理由は以下である。

Q 学習は、未知環境においてモデルを作成し、行動をプランニングし行動を決定する能力をもつ。経路生成の代表的な手法であるダイナミックプログラミングは環境が完全に知られている時に有効な手法である。しかしながら、単純な経路生成として離散的ダイナミックプログラミングを採用した場合には、最適経路が複数えられても搬送装置が衝突回避や何らかの事情で獲得されている経路からはずれた場合、その後の行動をえることができずその場からの再計算となり、自律搬送には不向きである。一方、確率的ダイナミックプログラミングを適用した場合には、現在地で搬送装置がとろうとする行動すべてに対して経路評価に対する確率遷移行列が必要となる。この点においてダイナミックプログラミングは分布モデルといわれるが、一般にこのような経路獲得に関する確率遷移行列をえるのは不可能である。Q 学習はサンプリングによる学習から経路生成の確率遷移行列すなわち環境モデルを同定し、かつ行動を決定できる特徴をもつ。従って、環境モデルとして同定された地点では同時に経路の評価に対する経路生成の確率遷移行列を作り出している。これは、衝突回避等において獲得経路を離れた場合に有効な行動選択規範となり自律行動を可能とする。

また、Q 学習は環境が変化した場合に追加学習を可能とする適応性をもつ。これらの利点から経路獲得に Q 学習を採用する。

今、時刻 t におけるエージェントの可能な状態を $s(t) = \{s_i(t); i = 1, 2, \dots, L\}$ 、可能な行動を $a(t) = \{a_m(t); m = 1, 2, \dots, M\}$ とする。その時、時刻 t で状態が $s_i(t)$ のときに選択した行動 $a_m(t)$ の利得期待値 $Q(s_i(t), a_m(t))$ は、

$$Q(s_i(t), a_m(t)) = (1 - \alpha) Q(s_i(t), a_m(t)) + \alpha [f + \gamma \max_{a_k} Q(s_i(t+1), a_k)] \quad (1)$$

と表される。 f はエージェントの環境からの成功報酬値、 α は学習率、 γ は利得の将来の減衰率である。エージェントの行動は以下の確率的選択関数を用いて決定される。

$$a_m = \arg \max_k \frac{\exp(Q(s_i(t), a_k(t))/T)}{\sum_{a_k(t) \in a} \exp(Q(s_i(t), a_k(t))/T)} \quad (2)$$

ここで、式 (2) に示される計算式はボルツマン確率分布、 T は計算温度、関数 $\arg \max$ はボルツマン分布確率に従って行動 a_k を選択する関数、関数 $\arg$ は選択した行動 a_k の引き数を a_m に引き渡す関数である。

2.2 経路獲得問題

図 1 は Watkins に採用された Q 学習のための経路獲得問題である。左端のスタート地点から右上のゴール地点までの経路を各格子点で Q 値を学習し獲得する。状態は、AGV の現在の位置ベクトル x と現在の速度ベクトル v によって定められ、行動は加速度ベクトル g から選択する。行動 g は、

$$a = \{g : (-1, -1), (-1, 0), (-1, 1), (0, -1), (0, 0), (0, 1), (1, -1), (1, 0), (1, 1)\}$$

から採用される。時刻 t から $t+1$ への状態遷移は

$$x(t+1) = x(t) + v(t) + g(t) \quad (3)$$

で定める。壁に衝突した場合は、 $v(t) = [0, 0]$ および、 $x(t+1) = x(t)$ とする。AGV の初期状態は $v(0) = [0, 0]$ とする。スタート地点はスタートの 4 点からランダムに選ばれる。 $t=100$ でゴールに達成しなかった場合は AGV を出発点に戻し、これを 1 回の学習回数とする。 t が 100 以内でゴールに達成した場合も同様とする。

Fig. 1 Driving course for Watkins' racetrack problem

2.3 失敗の行動に基づく Q 学習

多くの試行錯誤による知識獲得は、初期段階において失敗に終わることが多い。Watkins の Q 学習ではゴールにたどりついてダイナミックプログラミングの原理がスタート点に伝搬される。従って、初期段階ではある状態に対して選択した失敗の行動を採用しない逆学習が、学習にとって効果的と考えられる。今、式 (1) で与えられる Q 値を、

$$Q(s, a) = Q_s(s, a), Q_f(s, a) \quad (4)$$

とし、成功を学習する Q 値を $Q_s(s, a)$ 、失敗を学習する Q 値を $Q_f(s, a)$ とおく。ここで、 $Q_s(s, a)$ は式 (1) において $f=f_s(>0)$ 、 $Q_f(s, a)$ は式 (1) において $f=f_f(>0)$ と設定する。以下にこれらを用いたアルゴリズムを 3 種類提案する。

[アルゴリズム 1] 単純加減法

成功の Q 値と失敗の行動の Q 値を負にしたものを重みを用いて単純に加算する。従って、行動を決定する Q 値は、以下の式で計算される。

$$Q(s, a) = w_1 Q_s(s, a) - w_2 Q_f(s, a) \quad (5)$$

$w_1=w_2=1$ のとき、これは単純に式 (1) において報酬に成功の時 f_s 、失敗の時 f_f を与えるのに相当する。

[アルゴリズム 2] 確率分布加減法

Q 値を単純加減するのではなく、それぞれの Q 値をボルツマン分布式による確率に計算し、この確率の加算によって行動選択関数を作成し、行動を決定する。すなわち、

$$p_s(a_k) = \frac{\exp(Q_s(s_i(t), a_k(t))/T)}{\sum_{a_k(t) \in a} \exp(Q_s(s_i(t), a_k(t))/T)}, \quad (6)$$

$$p_f(a_k) = \frac{\exp(Q_f(s_i(t), a_k(t))/T)}{\sum_{a_k(t) \in a} \exp(Q_f(s_i(t), a_k(t))/T)}$$

を計算し、

$$p(a_k) = (1 - \eta) p_s(a_k) + \eta p_f(a_k) \quad (7)$$

から

$$a_m = \arg \max_k p(a_k) \quad (8)$$

に基づいて行動を決定する。ここで、 $0 < \eta < 1$ とする。

[アルゴリズム 3] 比較法

2 つの Q 値は学習の進み方が異なるので、その時点で学習の進んでいる方を次式により選択する。

$$Q(s, a) = \max(Q_s(s, a), Q_f(s, a)) \quad (9)$$

2.4 数値計算シミュレーション

図 1 に示される Watkins のレーストラック問題に対して、3 種類のアルゴリズムを適用する。学習パラメータは、 α を 0.3、 γ を 0.8 とした。アルゴリズム 1 ではウェイトを $(w_1, w_2) = (2, 1)$ 及び $(w_1, w_2) = (1, 1)$ とした 2 通りの実験を行った。その時の学習収束曲線を図 2~3 に示す。このアルゴリズムではどちらの実験でも失敗から学ぶ学習が早い、両者のウェイトが同じ場合は高速で精度のよい学習が実施されること

Fig. 2 Learning curves obtained by algorithm 1 when (w_1, w_2) is set to (2,1)

Fig. 5 Learning curves obtained by algorithm 2 when η is set to 0.2

Fig. 3 Learning curve obtained by algorithm 1 when (w_1, w_2) is set to (1,1)

Fig. 6 Learning curves obtained by algorithm 3

Fig. 4 Learning curves obtained by algorithm 2 when η is set to 0.8

がわかる。アルゴリズム2では η を0.8と0.2の二通りについて実験を行った。その時の学習収束曲線を図4～5に示す。やはり失敗を取り入れた方が学習が早い。失敗のウェイトを大きくすると学習精度が落ちることが判明した。アルゴリズム3の学習曲線結果を図6に示す。この結果は、失敗を取り入れた学習が従来法より高速で、よりよい学習を行っていることを示した。図7～8に学習回数が5000及び10000回の時のアルゴリズム3による獲得経路を示す。これらの結果から重み及びパラメータ調整の不要なアルゴリズム3が失敗を考慮した学習法として優れていることが分かった。

Fig. 7 Trace of AGV driving when learning time is five hundreds

Fig. 8 Trace of AGV driving when learning time is ten hundreds
Optimum trace is almost accomplished

3. 仮想工場モデルによるAGVの自律搬送

経路獲得に従来よりも高速なQ学習を用い、図9に示すような小規模の仮想工場モデルにおいてAGVの自律搬送と工場の稼働シミュレーションを実施する。工場は4台のマシニングセル、

Fig. 9 Virtual factory for AGV driving simulation

自動倉庫、自動倉庫出庫口からマシニングセル、マシニングセル間、マシニングセルから自動倉庫の入庫口への搬送を行う複数の AGV、からなる。工場の被加工物は自動倉庫に格納されており、自動倉庫の出庫信号によりこれを AGV が出庫口から定められマシニングセルに搬送する。マシニングセルは被加工物半製品を終了すると周囲に信号を発信し、これを見つけた AGV は次の指定されたマシニングセルへと被加工物を搬送する。すべての加工が行われた被加工物は最後のマシニングセルの終了信号を見つけた AGV により完成品として自動倉庫へと搬送される。このような仮想工場は現在存在しないが、蜂や蟻等にみられる自律的な社会行動を工場としてモデル化したとも考えられる。

3.1 AGV モデル

3.1.1 AGV の経路獲得モデル AGV は操車場から出発し、工場内をランダムに移動する。自動倉庫、マシニングセルそれぞれが発する後述する信号に出会うとその行動を決定する。例えば、AGV は信号を発しているマシニングセルあるいは自動倉庫に被加工物あるいは半製品受け取りに移動し、被加工物または半製品に指定された次の目的地、すなわち自動倉庫の入庫口あるいは次の加工を行うマシニングセルに被加工物または半製品を搬送する。AGV にはあらかじめ自動倉庫出庫口から各マシニングセル、マシニングセル間、各マシニングセルから自動倉庫入庫口への経路獲得を前章で述べた失敗を考慮したアルゴリズム 3 で前もって学習させる。AGV の速度は 1 セル/1 単位時間とする。AGV の状態は現在地 x とし、AGV の行動 a には移動方向 v を与える。従って、状態遷移は

$$x(t+1) = x(t) + v(t) \quad (10)$$

で計算する。AGV の行動は a は

$$a = \{v : (1,0), (1,1), (0,1), (0,0), (1,-1), (-1,0), (-1,-1), (0,-1), (1,-1)\}$$

の 9 種類から選択する。

3.1.2 衝突回避 AGV 間の衝突回避には、単純に先に AGV が他の AGV をそのセンサー領域内に発見したものが進行方向に対して右に避ける方法を適用する。文献³⁾は多くの場合に衝突回避はこのような単純な方法で実施できることを示している。経路への復帰は Q 学習の性質を利用する。前節で示した経路獲得法は、フィールド内の各地点に対して Q 値によって進行方向の道しるべをつくる学習と考えられる。従って、経路をはずれてもその場からの目的地への経路へ復帰することが可能である。図 10 は最大 Q 値をもつ方向を矢印で示した時の理想とする経路と、それをはずれた時の復元のコースの原理を示している。なお、AGV の障害物感知領域は AGV が位置する周囲の 8 セルとする。

(a) Ideal course

(b) Corrected course when AGV avoids collision

Fig. 10 Ideal course and corrected course. Black painted square shows the current AGV and white square shows the obstacle AGV

Fig. 11 Information flow between AGVs, machining cells and workpieces

3.2 被加工物

被加工物は自動倉庫に格納されており、ランダムに順次出庫される。被加工物はマシニングセルの加工順序と加工法を情報としてもち、AGV に積載されると行き先（マシニングセルと自動倉庫）を加工順序に従って AGV に指定する。

3.3 マシニングセル

仮想工場内のマシニングセルには、図 9 に示すように工場内の 16 セルをその大きさとして与える。また、マシニングセルの 1 セルを被加工物の取り入れ口、他の 1 セルを加工後に出庫口とし、これらの地点で AGV は被加工物及び加工済み半製品の積みおろしと受け渡しを実行する。これらの取り入れ口と出庫口はそれぞれ 10 個までの被加工物及び加工済み半製品をおくことが可能なバッファを与える。マシニングセルは、取り入れ口のバッファに余裕がある時イエローの信号、取り入れ口のバッファに余裕がない時はレッドの信号、出庫口のバッファに半製品がある時はグリーンの信号を発する。信号の発信範囲はマシニングセルの外周 20 セルとする。AGV はあるマシニングセルが搬送先であり、それがイエローの信号を発している時、そのマシニングセルの取り入れ口に被加工物を搬送し受け渡す。もしレッドの信号を発している時はその周辺で待機する。なにも搬送していない AGV がマシニングセルがグリーンの信号を発しているのを受信した場合は、その AGV は受信したマシニングセルの出庫口に向かい、半製品を受け取り、次の目的地へと搬送を行う。AGV、マシニングセル及び被加工物の信号の流れを図 11 に示す。

3.4 自動倉庫

自動倉庫は指定された被加工物を順次出荷する。出庫口には被加工物が 10 個まで格納できるバッファを設定する。バ

バッファに被加工物がある時は緑色の信号を出庫口で発し、バッファが空になると赤色の信号を発する。緑色の信号を見つけた AGV は被加工物を積載にくる。加工済の製品は AGV によって入庫口へ搬送される。入庫口も 10 個まで格納するバッファをもち、バッファに製品が 10 個未満の時イエローの信号を発し、そうでない時は赤色の信号を発するとする。AGV はイエローの信号であれば入庫口で製品を積みおろし、赤色の信号の時は付近で待機する。信号の到達範囲は出庫口及び入庫口ともそれぞれ図 9 に示されるそれらの上の 7 セルとする。

4. 搬送シミュレーション実験

2 章で実験を行った経路獲得法のアルゴリズム 3 により自動倉庫から各マシニングセル、各マシニングセル間、各マシニングセルから自動倉庫への経路を Q 学習で獲得させた後、その結果を用いて自動倉庫から各マシニングセルを経て自動倉庫に戻る搬送シミュレーションを実施した。以下に実験条件及び実験結果とその考察を述べる。

4.1 実験条件

Q 学習のパラメータは、2 章と同じ値を用いて AGV に経路獲得の学習を実施した。AGV の台数は 1~25 台とし操車場からランダムに出発させる。自動倉庫から出庫される被加工物は表 1 に示される 4 種類 (A~D) とし、それらは 4 台のマシニングセル (M_1 ~ M_4) で表 1 に示される時間で加工される。単位時間は AGV の 1 セル移動時間に等しくおく。加工順序は設定せずに最も近いマシニングセルを AGV が選択するとした。

Table 1 Transported work pieces and their machining time

Workpiece	Machining time by each machining cell				Number of workpiece
	M_1	M_2	M_3	M_4	
A	10	20	5	15	10
B	20	10	15	5	10
C	5	15	20	10	10
D	15	5	10	20	10

4.2 シミュレーション結果

図 12 は搬送シミュレーションの初期の状況を示す。17 台の AGV の内 2 台が自動倉庫の出庫信号を見つけその搬出口に入り、11 台が工場内にランダムに操車場の出庫口から出発している。操車場の出庫口には 2 台が存在し、2 台が操車場に隠れている。図 13 は 9 台が被加工物の搬送を行っており、AGV の 6 台が搬送せずにランダムに移動している状態を示す。また、2 台はマシニングセルの搬出口で加工済みの半製品を受け取っている。図の下辺に搬送を行っている AGV が多いのは倉庫から被加工物を受け取った AGV が多いことを示している。図 14 は 12 台が搬送を実施している。これらの内図の左下 3 台は自動倉庫の入庫口に全て加工済の製品を運んでいる AGV である。下辺中央の 3 台は自動倉庫出庫口から被加工物を受け取ったものであり、残りは半製品を搬送中である。AGV の 2 台はマシニングセルの搬出口で加工済の製品を受け取っているところである。残る 3 台の内 1 台は左上のマシニングセルに被加工物を受け渡したところであり、他 2 台はランダムに行動を行っている。このような実験を AGV の台数が 1~25 台それぞれの時に、搬送シミュレーションを実施した。そのときの各台数における搬送終了時間 (滞留時間, make span time) をグラフとして図 15 に表した。搬送終了は 4 種の被加工物がすべて自動倉庫に戻ったときの経過時間とした。グラフから AGV の投入台数が増加するに従って滞留時間が短縮されるが、10 台以上の時はほとんど同じ滞留時間となることからわかる。これは各マシニングセル及び自動倉庫

Fig. 12 Simulation snapshot. AGVs leave for floor

Fig. 13 Simulation snapshot. AGVs autonomously move on the floor

Fig. 14 Another simulation snapshot

に設定したバッファ数に一致する。シミュレーションを観察してわかるのは AGV が被加工物 (半製品) をバッファ数だけ行き先に運び込み、バッファが一杯になると遊休していることが観察された。最小の滞留時間 F は、被加工物がマシニングセンターを順次 M_1, M_2, M_3, M_4 へと巡回すると仮定すれば、次式で目安として近似できる。

$$F = F_1 + F_2 + F_3 + F_4 \quad (10)$$

ここで、 F_1, F_2, F_3 及び F_4 は、それぞれ 4 種類の製品すべての加工時間、最後の搬送被加工物の全マシニングセルでの積みおろし時間、最後の搬送被加工物の自動倉庫での積みおろし時間、及び被加工物が自動倉庫から 4 種類のマシニングセルを経て自動倉庫へ戻る滞留時間である。表 1 から $F_1=2000$ が得られる。マシニングセルの積みおろし時間を 1 とすると、 $F_2=8$ を得る。自動倉庫での積みおろし時間も同様とすれば $F_3=2$ である。AGV が搬送を 4 台のマシニングセルの外周を移動するとすれば、この移動時間は 60 であり、自動倉庫から最短の 1 台のマシニングへの往復時間は 10 とみなすと $F_4=70$ と近似できる。これらから、 $F=2080$ を得る。シ

Fig. 15 The number of AGVs vs. make span time

Fig. 16 Deadlock phenomenon

シミュレーション結果による搬送時間はAGVを10台以上投入した時、約8800で収束している。近似した滞留時間からこの値を検討すると、約3倍の時間はAGVが遊休あるいはランダムな移動を行っていると考えられる。

図16はAGVが15台の時に観察されたデッドロック現象である。多くのAGVが左下のマシニングセルに集中したため、このような現象が生じたと考えられる。今回はデッドロックに関する抜本的な解決策は見出せなかった。Q学習による移動方向決定は確率的なため、時間が経過すればこのようなデッドロックは解決される。しかし、これには長時間を有するため図14に示した実験では、一定時間同じ場所に留まる時にAGVをQ学習で示す方向以外をQ値の逆数に対して

ボルツマン確率分布関数から確率的に行動を選択させ、移動方向を決定するようにした。このようなデッドロックに対しては木下⁹⁾等が目的地と反対に動く学習を時系列の中で取り入れることをクラシファイヤ・システムを用いて提案している。しかしながら、自律分散型の移動問題では常にデッドロックは四人のジレンマ問題として記述される大きな問題となるため、抜本的な解決策が必要であると考えられる。

5. 結 言

自律分散型生産システムにおけるAGVの自律搬送を実現するためにAGVの工場内における経路獲得法をQ学習を用いて提案し、その学習結果を用いた仮想工場でのシミュレーションを実施した。これらから以下のことが明らかとなった。

- (1) Q学習による経路獲得に失敗を取り入れた学習法を3種類提案し、効果的な学習法を調べた。
- (2) AGVが何らの指示を与えられることなく行動する仮想工場モデルを作成し、そのシミュレーションが可能であることを示した。
- (3) AGVの衝突回避を行った場合、学習されたQ値によって自律的に正しい経路に復帰できることを示した。
- (4) シミュレーション結果は、被加工物を搬送するのに一定のAGV台数が必要であるが、それ以上は搬送効率が向上しないことが示された。
- (5) シミュレーション中にデッドロック現象が観察された。これに対しては簡易な解消法を採用したが、本質的な解決策が必要である。

なお、本研究は一部平成10～12年度科学研究費補助金基盤研究(C)2による「マルチエージェントの適応・学習に基づいた自律分散型生産システムモデルの研究」に基づいて行われたことを付記し、関係者に謝意を表する。

参 考 文 献

- 1) Dean and A. Carrie : FUTURE ENTERPRISE TYPES AND STRATEGIES FOR AGILE MANUFACTURING, *Advances in Manufacturing Systems*, Edited by N. Okino, H. Tamura, and S. Fujii, IFIP (1996) 299.
- 2) L. Wang : COLLABORATIVE APPROACH FOR HOLONIC MANUFACTURING SYSTEM", *Proc of FAIM'00*, Vol. II, Econo Printing & Graphic, inc., Maryland (2000) 1146.
- 3) 沖野教郎 : 生物型生産システム, 朝倉書店 (1993)
- 4) C. J. C. H. Watkins : Technical Note Q-learning, *Machine Learning*, 8, (1992) 279.
- 5) 渡辺美知子, 加藤龍, 古川正志, 木下正博, 嘉数侑昇 : Q学習によるAGV衝突回避通信プロトコルの獲得, *精密工学会誌* 66巻1号, (2000) 101.
- 6) M. Kinoshita, T. Kawakami and Y. Kakazu, Taking Block Out Problem by Adaptive Packing Method, *Intelligent Engineering Systems Through Artificial Neural Networks*, 7 (1997) 389.