

制御系における状況認識と行動判断に関する研究

Study on state recognition and behavior
decision under control systems

主任研究員：竹口知男

分担研究員：岩佐哲夫 宮井義裕 能勢和夫

概 要

少子高齢化の急速な進展にともない、仕事や家庭生活において人間の作業を代行または支援してくれる自律(移動)ロボットの開発が強く求められている。そのような自律移動ロボットにおいては、ロボット周辺環境をロボット自らが状況認識し、遂行すべきタスクに合わせて最適な行動を判断し、行動実行を行う必要がある。これらロボットによる状況認識、最適行動判断ならびに行動実行に対して、多くの研究がなされてきているが、それぞれ十分な成果を得るまでには至っていない。

そこで、本長期的共同研究は、人とインタラクションを有する移動体(自動車、自律移動の福祉ロボット等を含む)の制御系において、人間特性を考慮した知的処理を図るためのヒューマンインターフェイスについての研究を行うことを研究目的として、平成9年度より研究を開始した。本研究目的を遂行するために、

(1)「制御系の学習機能」(竹口担当)

自律移動ロボットの制御系における学習機能についての研究を行う

(2)「制御系の状況認識と定量的表現」(岩佐・宮井担当)

人および自律移動体の状況判断の定量的表現方法、ならびに移動体間および移動体と人間間の通信手段についての研究を行う

の2つの分担研究課題からスタートし、平成10年度からは

(3)「自律エージェントの知識獲得」(能勢担当)

マルチエージェントによる知識獲得ならびに知識の共有・協調についての研究を行う

という分担研究課題も加えて研究を行ってきた。本長期的共同研究の5年間の各分担研究によって、つぎのような研究成果を得た。

(1) 分担研究「制御系の学習機能」においては、

行動ベース型知能ロボット処理系の概念である包摂アーキテクチャ(Subsumption Architecture: SA)の自律移動ロボットに対する有効性検証実験結果として、SAは環境ならびに行動目的が比較的単純な場合には条件反射的行動(昆虫的自律行動)を実行することから非常に効率良く作動し、自律移動ロボットの行動生成機能として有効である。しかし、装備する行動モジュールならびに行動モジュール間の競合解消(発火順位付け)方法については設計者があらかじめトップダウン的に決定してやらねばならず、より複雑な環境に対応させ、より高度な知的処理を行わせようとした場合、それらの問題点を含めた自律的な学習機能が必要であることが分かった。

そのため、自律的な学習機能として、目標状態を与えれば、そこに至る動作・行動列の試行錯誤を通じて、自律的に行動戦略を獲得する学習システムとして強化学習(Reinforcement Learning : RL)法を採り入れることとし、強化学習の代表的な評価手法である経験強化型の利益共有法(Profit sharing plan : PSP)ならびに環境同定型の Q-learning についての有効性検証を 2 次元平面上の障害物回避最短経路探索問題に対するシミュレーション実験によって行った。そのシミュレーション実験結果によって得られた PSP および Q-learning の得失を踏まえて、それらをハイブリッド化した“Hybrid-learning”と呼ぶ強化学習法を本研究において提案し、同一の障害物回避最短経路探索問題のシミュレーション実験によって、本提案手法が PSP および Q-learning より獲得経路の最適性ならびに学習速度の両方の点において優れていることを示した。

(2) 分担研究「制御系の状況認識と定量的表現」においては、

移動するビーグルの制御に関係する色々な対象について量的表現による情報獲得を目的として実験を行ってみた。その経験から将来の移動体の制御は視覚すなわち画像処理に負うところが大きであると考えられる。また聴覚の情報を得るための基礎的な実験として、騒音計および FFT による色々な試行も検証の得られる方向は残念ながら見出せなかった。しかし指向性のある音を取り出す方法、例えば電波に音声を乗せるように超音波に音声を乗せればアンプも不要で筒の中を通るような音が空間で獲得できる。この技術が実用できれば光と同様に有効な応用が期待できると考える。

(3) 分担研究「自律エージェントの知識獲得」においては、

狭い道路や交差点における車両 2 台の同時協調運転操作の生成問題を考察し、従来の進化計算法に改良を加えたアルゴリズムを提案した。数値計算の結果ではあるが、改良点であるゼロ時間幅挿入処理の導入が同時協調運転操作の生成に有効であることが検証できた。ただ、現在の手法では、車両の台数が増加すると染色体長さも比例的に増加してしまうため、計算時間も増大してしまう。そのため、車両の台数に比例しない解の染色体表現方法を検討し、より多くの車両間での協調運転操作の生成について考慮する必要がある。

以上の研究成果の一部については、つぎのように外部発表を行い、評価を得た。

- 1) 喜多, 能勢, 平松: 適応的進化計算法を用いた車の運転操作の最適化, 第 43 回自動制御連合講演会前刷, pp. 225-226, 2000
- 2) 竹口, 小尻, 大橋, 今井, 能勢: 強化学習によるロボット移動経路の探索, 日本機械学会関西支部第77回定期総会講演会講演論文集, No. 014-1, pp. 5-47 - 5-48, 2001
- 3) 喜多, 能勢, 平松: 適応的進化計算手法を用いた複数台の車の運転操作最適化, 第 44 回自動制御連合講演会前刷, pp. 548-551, 2001
- 4) 喜多, 能勢, 平松: 適応的進化計算手法のトレーラ・トラック運転操作最適化への応用, 第 45 回システム制御情報学会研究発表講演会講演論文集, pp. 459-460, 2001
- 5) M. Kita, K. Nose, A. Hiramatsu, H. Imai, T. Takeguchi: Cooperated Drive Operation of Two Cars Using Adaptive Evolutionary Computation, Proceedings of the Seventh International Symposium on Artificial Life and Robotics, pp. 401-406, 2002
- 6) 能勢, 平松, 天目, 下浦: 高速道路の最適流入制御への進化計算手法の適用, システム制御情報学会論文誌, Vol. 15, No. 9, pp. 486-494, 2002

目 次

1. まえがき
2. 制御系の学習機能「自律移動ロボットの学習機能に関する研究」
3. 制御系の状況認識と定量的表現「制御系における状況認識と定量的表現について」
4. 自律エージェントの知識獲得「複数車両の協調運転操作の生成に関する研究」
5. あとがき

1. まえがき

少子高齢化の急速な進展にともない、仕事や家庭生活において人間の作業を代行または支援してくれる自律(移動)ロボットの開発が強く求められている。そのような自律移動ロボットにおいては、ロボット周辺環境をロボット自らが状況認識し、遂行すべきタスクに合わせて最適な行動を判断し、行動実行を行う必要がある。これらロボットによる状況認識、最適行動判断ならびに行動実行に対して、多くの研究がなされてきているが、それぞれ十分な成果を得るまでには至っていない。

そこで、本長期的共同研究は、人とインターアクションを有する移動体(自動車、自律移動の福祉ロボット等を含む)の制御系において、人間特性を考慮した知的処理を図るためのヒューマンインターフェイスについての研究を行うことを研究目的として、平成9年度より研究を開始した。本研究目的を遂行するために、

(1)「制御系の学習機能」(竹口担当)

自律移動ロボットの制御系における学習機能についての研究を行う

(2)「制御系の状況認識と定量的表現」(岩佐・宮井担当)

人および自律移動体の状況判断の定量的表現方法、ならびに移動体間および移動体と人間間の通信手段についての研究を行う

の2つの分担研究課題からスタートし、平成10年度からは

(3)「自律エージェントの知識獲得」(能勢担当)

マルチエージェントによる知識獲得ならびに知識の共有・協調についての研究を行う

という分担研究課題も加えて研究を行ってきた。平成14年度をもって本長期的共同研究を終了するにあたり、5年間の研究成果を報告する。

なお、本研究成果報告書においては、2章で分担研究「制御系の学習機能」における研究結果を、3章で分担研究「制御系の状況認識と定量的表現」についての研究結果を、4章では分担研究「自律エージェントの知識獲得」の研究結果について述べる。

2. 制御系の学習機能「自律移動ロボットの学習機能に関する研究」

2.1 はじめに

急速な少子高齢化社会の到来にともない、人間の代わりに家庭やオフィスの作業を代行したり、高齢者や障害者の介護を支援したりするロボットの開発が強く求められている。そのようなロボットには人間との協調・協働機能ならびに未知行動環境下での適切な行動を創発する機能を備えている必要がある。しかし、これまでの多くのロボット研究においては、設計者によってあらかじめ与えられた行動環境設定の下で、与えられた処理方法に沿って動作・行動するロボットの研究が主であった。このような設計方法では、設計者がすべての環境を掌握・検討し、それぞれの環境に適した処理方法を組み込まなければならず、少しでもそれらの環境からずれた環境が生じた場合にはロボットは対応不能となってしまう。このことは、設計者に大きな負担を課することとなり、環境変動の大きい実環境に対する処理方法の確立を困難なものとしている。

そこで本分担研究においては、設計者はある基本的な機能のみをロボットに与えることによって、後はロボット自身が環境に適応する行動を経験的に学習することにより自律的に行動戦略を獲得する機械学習法についての研究を行うこととした。そのためまず、系の中で並列に存在する要素行動群を協調させて、相互に表層に現れる環境に合わせた行動を生成する概念に基づく包摂アーキテクチャ (Subsumption Architecture : SA) と呼ばれる行動ベース型知能ロボット処理系の有効性検証実験を行うとともに、生物行動学における動物の躰などの用いる強化学習を工学的に模倣した機械学習法である強化学習 (Reinforcement Learning : RL) 法についての研究を行うこととした。

2.2 包摂アーキテクチャ(SA)による自律移動ロボットの行動生成について

有限状態マシンの考え方を応用し、1980年代後半にMIT人工知能研究所のRodney A. Brooksが提唱した包摂アーキテクチャ (Subsumption Architecture : 以下SA) と呼ばれる行動ベース型知能ロボット処理系は、従来の知能ロボットの処理系である計画ベース型ロボットでは実現できなかった、未知環境下での行動を可能にした¹⁾。

従来の知能ロボットの処理系においては、Fig. 2.1に示すようにセンサ群の入力から外界認識、モデル化、行動計画の作成、タスクの実行を経てアクチュエータ群の駆動までの直列的な信号の流れの処理を基本としており、この処理系では行動可能なパターンがあらかじめ設計者の想定した範囲内に限られてしまうとともに、センサ入力からアクチュエータ出力までの処理に時間がかかり環境変化に高速に追従することが困難である。それに対しSAでは、Fig. 2.2に示すように、障害物回避や探索などの目標達成に必要な機能を行動モジュールごとに分割し、並列的かつ階層的に配置することで、各モジュールは、センサおよびアクチュエータと直接結合され、常にセンサ入力を監視し、センサ入力に合わせた処

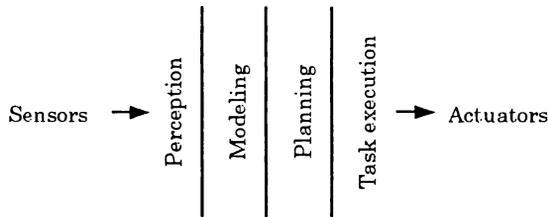


Fig. 2.1 Procedure of ordinary AI

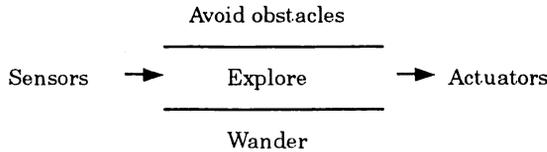


Fig. 2.2 Procedure of subsumption architecture

理方法をすばやくアクチュエータに出力・動作させることができるため、ロボットは予測不可能な環境の変化に対しても、反射的に行動することが可能となる。

車輪駆動用 DC モータならびに光センサなどの各種センサを有し、それらのセンサ情報を基準として SA の概念に基づくナビゲーション・アルゴリズムによって動作する車輪駆動型自律移動ロボットを製作し、実機実験を通して、SA の概念がどのように有効であるかを考察した。

有効性検証実験としては、事前に障害物などの環境に関する情報をまったくロボットに与えず、点光源をゴールとした二次元平面の障害物回避経路探索問題と先導ロボットと追尾ロボットの 2 台のロボットによる協調物体搬送問題に対して実験を行った。

2.2.1 実験用ロボット

知能ロボット実験研究用に MIT 人工知能研究所の監修で開発された 2 輪駆動型ロボット “Rug Warrior” (Fig. 2.3) を 2 台組み立てて実験に用いた。Rug Warrior は、マイクロプロセッサ (Motorola 製 MC68HC11A0) で制御され、左右両車輪駆動用 DC モータとエン

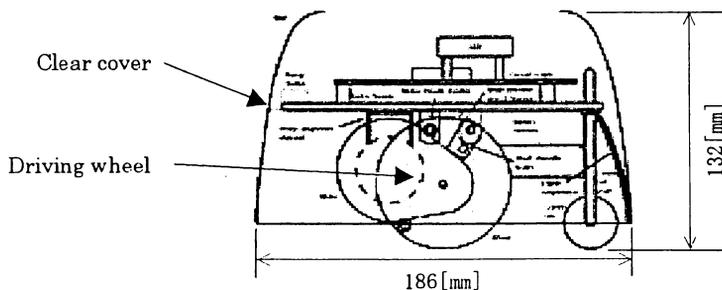


Fig. 2.3 Side view of Rug Warrior

コーダおよび光センサ、赤外線センサ、接触センサ、音響センサの各種センサを備えている。

制御用言語としては IC (Interactive C) ²⁾ を用い、パソコンよりシリアルポートを通じて、コンパイルしたプログラムを Rug Warrior 本体上の RAM に転送し、実行することにより、Rug Warrior を自律動作させることが可能である。

また、協調搬送作業を行わせるために、追尾ロボットには超音波ソナーセンサ (A. K. Peters. Ltd. 製 Rug Bat) およびマイクロサーボ (Futaba S3102) を取り付け (Fig. 2.5 参照)、ソナーセンサを Fig. 2.4 に示すように左右に首振をしながらスキャンさせるようにして、先導ロボットとの距離ならびに障害物までの距離を計測できるようにした。

先導ロボットの最大移動速度が、追尾ロボットの最大移動速度を上回らないようにするために、追尾ロボットのソナーセンサの回転角を $\pi/6$ [rad]、追尾ロボットの最大移動速度を 0.28 [m/sec] として、先導ロボットの最大移動速度を $0.28 \cos \pi/6 = 0.24$ [m/sec] に設定した。

先導ロボットおよび追尾ロボットのプログラムは、Fig. 2.6 に示すようなサブサンプション構成とした。

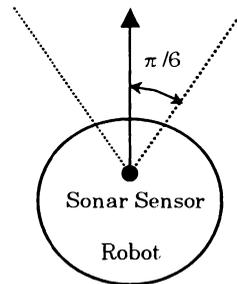


Fig. 2.4 Detected range of sonar sensor

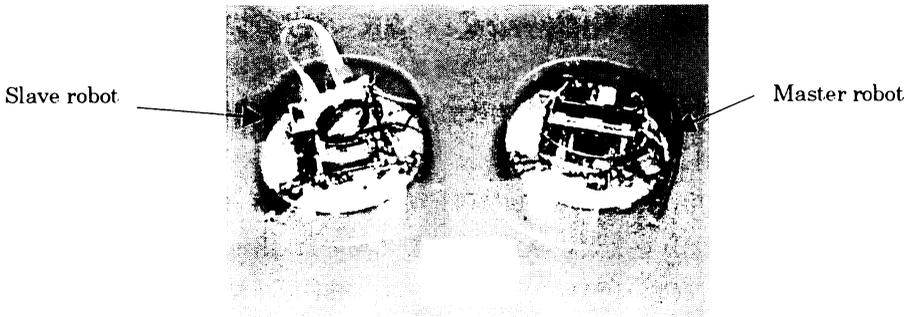
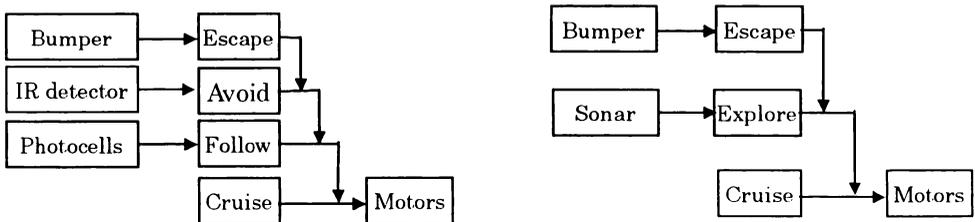


Fig. 2.5 Master and slave robots



(a) Master robot

(b) Slave robot

Fig. 2.6 Subsumption architecture for Rug Warriors

2.2.2 実験結果

ロボットの移動空間はいくつかの障害物を含む二次元平面とし、ロボットの移動目標位置は、障害物回避経路探索問題および協調物体搬送問題ともに、ある高さに設置された点光源（60[W]の電気スタンド）直下の床面とした（Fig. 2.7 参照）。

(1) 障害物回避経路探索問題に対して

先導ロボット用のSAプログラム（Fig. 2.6(a) 参照）を用いて実験を行った。実験の結果として、事前に障害物などの環境に関する情報をまったくロボットに与えなくても、障害物を含む二次元平面において、すばやく障害物を回避しながら移動目標点である点光源に向かって進むという、昆虫的自律(条件反射的)行動を行わせることが可能であった。その実験の様子を Fig. 2.7 に示す。

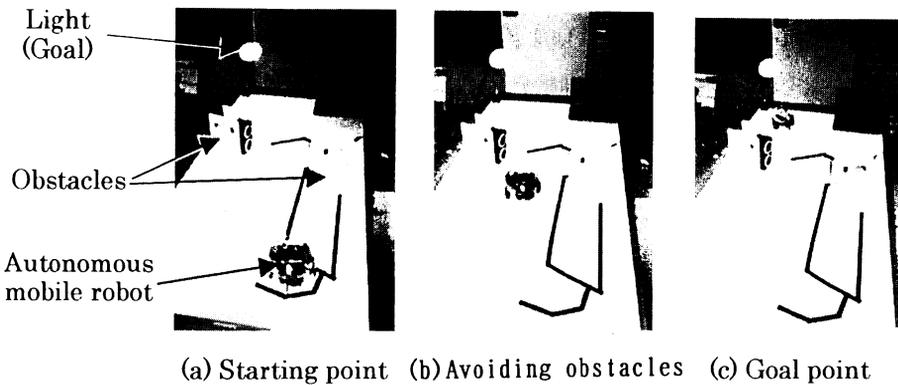


Fig. 2.7 Experiment for avoiding obstacles

(2) 協調物体搬送問題に対して

Fig. 2.8 に示すような搬送イメージで、先導ロボットが障害物を避けて移動目標点まで移動し、追尾ロボットは先導ロボットと一定の距離間隔を保ちながらその先導ロボットに追従する。ただし、本実験では、実際に搬送物体をロボットに載せずに、2台のロボット間の距離を一定間隔に保ちながらの移動実行可能性についての検証を行った。なお2台の設定距離間隔としては0.40[m]~2.00[m]まで変化させた設定とした。

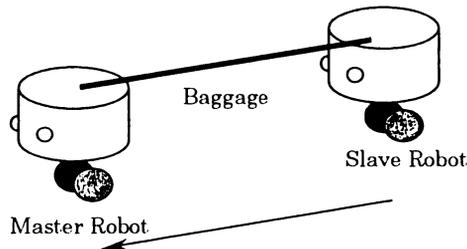


Fig. 2.8 Illustration of cooperative conveying

実験の結果としては、搬送物の長さ（ロボット間設定距離間隔）が 0.9[m]までの短いときには、先導ロボットへの追従は障害物を回避しながら間隔誤差が比較的小さくスムーズに並走移動したが、搬送物が長くなれば長くなるほど間隔誤差が大きくなり、2.0[m]以上では並走動作が破綻し、各ロボットが独自走行に近い形態で走行し協調搬送が不可能であった。この原因としては、制御処理系の問題というよりは、ロボット間距離のセンシング方法に問題があるものと思われる。すなわち、Fig. 2.4 のように超音波ソナーセンサを首振りさせながら距離検出をさせた場合、ロボット間距離が長くなればなるほど間隔距離検出可能な首振り角度が狭くなり検出時間が短くなること、および障害物など先導ロボット以外の周辺物体からの外乱信号が大きく作用してくることなどによる問題であると考えられる。

2.2.3 SA 有効性検証実験に対する考察

SA の概念を用いた行動ベース型ロボットによる実機実験の結果より、SA の概念は、環境ならびに行動目的が比較的単純な場合には条件反射的行動(昆虫の自律行動)を実行することから非常に効率良く動作し、有効である。しかし、SA そのものの概念では、装備する行動モジュールならびに行動モジュール間の競合解消(発火順位付け)方法については設計者があらかじめトップダウン的に決定してやらねばならず、より複雑な環境の下で、より高度な知的処理を行わせようとした場合、効率的に動作させることができないことが分かった。

このことは、ロボット自身に行動モジュール生成機能ならびに各行動モジュール競合解消機能を有していないことに起因している。したがって、より高度な目的やタスクをロボットに達成させるためには、ロボット自身に学習機能を持たせ、行動モジュールの各機能を協調させるように各行動モジュール間の優先順位を試行により学習獲得させ、さらには、その行動モジュール間の優先順位に基づき新たなる行動モジュール群を生成する機能を備えることによって、環境に応じた適切な行動戦略を生成させるようにする必要があることが分かった。

2.3 強化学習(RL)による自律移動ロボットの行動生成について

環境に対する動物の適応過程を工学的にモデル化した強化学習(Reinforcement Learning: RL)は、エージェント(自律移動ロボット)がとった行動に対して、評価としての報酬または罰を与えることによってエージェントに望ましい行動を自律的に獲得させる機械学習法であり、設計者は報酬や罰を設定するのみで、エージェントが自律的学習により行動獲得する点や不確実・不定形な環境に対しても対処可能であるという点に特徴がある⁸⁾。

強化学習において、ロボットは報酬による収益を最大化する行動を選択し、試行(エピソード)を繰り返すことにより適切な行動についての行動戦略を学習する。その行動は観測した状態から期待収益が最大化されるように次の行動を確率的に決め、その決定がどれだけ有効であったかは、試行終了時に与えられる報酬に基づいて定義される価値関数に反映さ

れる。その価値関数の付与法は強化学習法により異なり、強化学習の研究においてはこの価値関数の付与方法が研究対象となっている。

代表的な付与方法としては、Fig. 2.9 に示すように報酬を得た際に報酬を得るに至った過程(行動)に対し一括的に強化する学習法である利益共有法(Profit sharing plan : PSP)，および Fig. 2.10 で示すように状態と行動に依存する Q 値と呼ぶ利得の期待値を評価の見積もりとし、行動毎に Q 値の時間的差分誤差(Temporal difference (TD) error)分に基づき強化改善を図る学習法である Q-learning とがある。PSP は経験強化型の強化学習法であり、学習速度は速いが、最短経路を獲得する保証がなく、Q-learning は環境同定型の強化学習法であり、最適化手法である動的計画法に基づいているため学習速度は遅いが、適正経路の獲得が保証されるということが知られている⁴⁾⁻⁶⁾。

本報告では環境同定型の強化学習である Q-learning を基本に経験強化型の強化学習である PSP を併用した強化学習法である Hybrid-learning を提案し、Hybrid-learning が自律移動ロボットの行動生成問題に対して最適性および学習速度に関し適していることを PSP および Q-learning との比較により明らかにする。

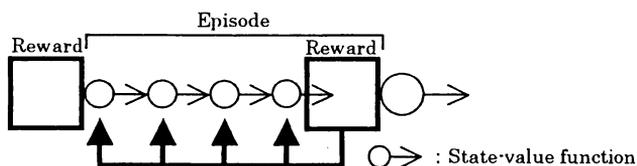


Fig. 2.9 Learning process by PSP

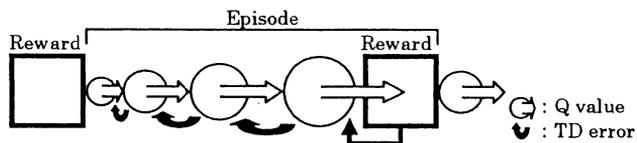


Fig. 2.10 Learning process by Q-learning

2.3.1 強化学習シミュレーション実験における問題設定

本研究で用いたシミュレーション環境を Fig. 2.11 に示す。周囲を壁に囲まれた 14×14 マス目の 2 次元格子状空間を行動領域とし、その格子状空間内に 15 の障害物が存在する。自律移動ロボット(エージェント)は、障害物や壁面との衝突を回避しながら、スタート地点からゴール地点までの最短移動経路を探索する障害物回避最短経路探索問題に対して試行(学習)実験を行った。エージェントは 1 ステップで周辺 8 方向に 1 マス移動することができ、その 8 方向への行動選択にはルーレット選択方式を用いた。

また、学習終了条件としては、10 エピソード続けて同一経路を探索したとき、学習が収束したものを見なして学習を終了することとし、 10^4 エピソードを越えても学習が収束しないときには、学習を失敗したものとした。1 エピソード中のステップ数が 10^4 ステップを越えた場合には、デッドロックが生じたものとして学習を終了させることとした。

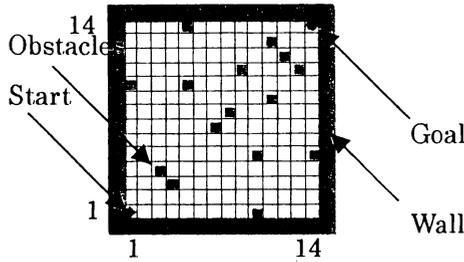


Fig. 2.11 Simulation environment

2.3.2 PSP による移動経路学習シミュレーション

エージェントは状態から収益を計算し、報酬を得るとその収益に基づき目標値(状態価値関数)を更新する。状態価値関数 $V(s_t)$ の更新式は 4),

$$V(s_t) \leftarrow V(s_t) + \alpha [R_t - V(s_t)] \quad (2.1)$$

で与えられる。時刻 t で観測した非終端状態を s_t として、その後起きる事象によって状態価値関数 $V(s_t)$ の推定値を更新する。ここで R_t は時刻 t に対応する実際の収益値で、 α ($0 < \alpha < 1$) は学習率である。PSP は経験強化型の強化学習であり、試行(エピソード)毎に学習結果が大きく異なる。そこで、学習率 α を 0.1 刻みで $\alpha=0.1 \sim 0.9$ の各学習率で 10 回試行実験を行った結果を Fig. 2.12 に示す。縦軸は試行終了時のステップ数とエピソード数であり、横軸は学習率 α である。学習率の変化に対して獲得経路および学習速度に傾向は見られず、デッドロックは全試行で起きなかった。シミュレーション環境から求められる計算上の最短経路ステップ数 14 を獲得することはできなかった。学習率 $\alpha=0.4$ および $\alpha=0.8$ の場合のエージェントが獲得した移動経路を Fig. 2.13(a) および (b) にそれぞれ示す。ここで、学習率 $\alpha=0.4$ の場合のエピソード数、ステップ数および衝突回数はそれぞれ 130, 41 および 13 であり、学習率 $\alpha=0.8$ の場合のエピソード数、ステップ数および衝突回数はそれぞれ 34, 41 および 5 であった。両獲得経路共に適正経路とはいえないが、エピソード数 34 で学習を終了しており、学習速度が比較的速いことが分かる。

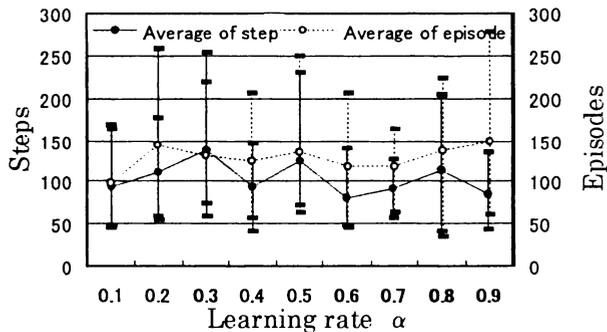


Fig. 2.12 Steps and episodes by PSP

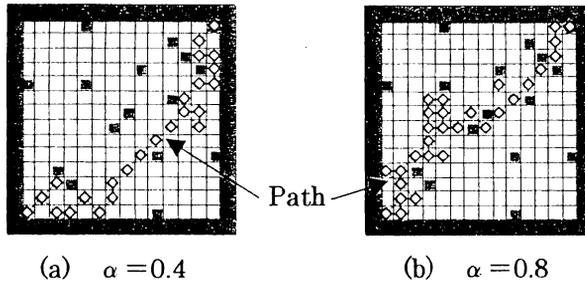


Fig. 2.13 Learned paths by PSP

2.3.3 Q-learningによる移動経路学習シミュレーション

エージェントは現状態で Q 値(行動価値関数)に基づき行動し, 観測により報酬と次状態から Q 値を更新する. 時刻 t において非終端状態 s_t , 行動 a_t とすると, Q 値 $Q(s_t, a_t)$ の更新式は,

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a_t) - Q(s_t, a_t)] \quad (2.2)$$

で与えられる⁴⁾. ここで α ($0 < \alpha < 1$) は学習率で, γ ($0 < \gamma < 1$) は割引率である. 割引率 $\gamma = 0.1 \sim 0.9$ (0.1 ごと) および学習率 $\alpha = 0.1 \sim 0.9$ (0.1 ごと) の全組み合わせ 81 通りのシミュレーション実験結果を Fig. 2.14 に示す. ただし, 縦軸は試行終了時のステップ数とエピソード数を, 横軸は 1 野線間隔内を同一割引率 γ とし, 各野線間隔内を学習率 $\alpha = 0.1 \sim 0.9$ に内分した座標としている. 割引率および学習率の変化に対して, 獲得経路および学習速度に傾向は見られず, デッドロックは全試行で生じなかった. シミュレーション環境から求められる計算上の最短経路ステップ数 14 を獲得することができなかった. 割引率 $\gamma = 0.4$, 学習率 $\alpha = 0.2$ および $\gamma = 0.5$, $\alpha = 0.8$ の場合のエージェントが獲得した移動経路を Fig. 2.15(a) および (b) にそれぞれ示す. ここで, 割引率 $\gamma = 0.4$, 学習率 $\alpha = 0.2$ の場合のエピソード数, ステップ数および衝突回数はそれぞれ 43, 15 および 0 であり, 割引率 $\gamma = 0.5$, 学習率 $\alpha = 0.8$ の場合のエピソード数, ステップ数および衝突回数はそれぞれ 61, 15 および 0 であった. 両獲得経路ともエージェントがスタートからゴールまで直線的に移動していることが分かる.

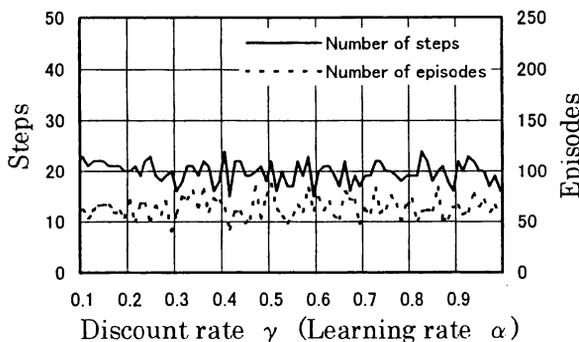


Fig. 2.14 Steps and episodes by Q-learning

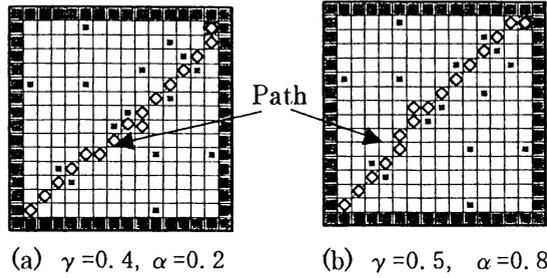


Fig. 2.15 Learned paths by Q-learning

Q-learning による獲得経路は 15 ステップであり、計算上の最短経路の 14 ステップに対して準最適であるといえるが、学習速度の点では学習終了時のエピソード数が PSP の 34 エピソードと比べると 43 エピソードと大きく、学習速度が遅い。しかし PSP と比べると、すべての学習率 α ならびに割引率 γ に対して学習終了時のステップ数がほぼ一定であることから、自律移動ロボットの移動経路探索問題に対して有効であることが分かる。

2.3.4 Hybrid-learning による移動経路学習シミュレーション

本報告では Q-learning での最適保証性ならびに PSP の学習の高速性を有する強化学習法として Hybrid-learning を提案する。また、本提案手法 Hybrid-learning の最適性および学習速度における有効性を、PSP および Q-learning の場合と同一の障害物回避最短経路探索問題を用いて示す。

2.3.4.1 行動決定則と価値関数の更新式

Hybrid-learning は PSP の状態価値関数 $V(s_t)$ と Q-learning の Q 値 $Q(s_t, a_t)$ を組み合わせた価値関数を用いる強化学習法である。それぞれの価値関数の更新は

$$V(s_t) \leftarrow \begin{cases} V(s_t) + \alpha_v [R_t - V(s_t)] & \text{if } Q(s_t, a_t) = \text{Initial Value} \\ V(s_t) + \alpha_v [R_t - \max_a V^*(s_t)] & \text{otherwise} \end{cases} \quad (2.3)$$

$$Q(s_t, a_t) \leftarrow \begin{cases} Q(s_t, a_t) + \alpha_Q [r_{t+1} + \gamma Q(s', a') - Q(s_t, a_t)] & \text{if } Q(s_t, a_t) = \text{Initial Value} \\ Q(s_t, a_t) + \alpha_Q [r_{t+1} + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)] & \text{otherwise} \end{cases} \quad (2.4)$$

によってなされ、各強化学習法が独立して状態の観測と行動の選択を行い、個々に選択された行動は行動決定則に従って実行行動として決定される。

2.3.3.2 Hybrid-learning による移動経路の獲得

状態価値関数 $V(s_t)$ の学習率 $\alpha_v = 0.1 \sim 0.9$ (0.1 ごと)、Q 値の割引率 $\gamma = 0.1 \sim 0.9$ (0.1 ごと)、学習率 $\alpha_Q = 0.1 \sim 0.9$ (0.1 ごと) の全組み合わせ 729 通りのシミュレーション実験を行った。そのシミュレーション実験結果を Fig. 2.16 に示す。ただし、縦軸は試行終了時のステップ数およびエピソード数とし、横軸は Fig. 2.14 の場合と同様に学習率 α_v および α_Q を内点とした割引率 γ 系列として表わしている。Fig. 2.16(a) から割引率 $\gamma = 0.3 \sim 0.9$ で、ステップ数に学習の効果が見られ、Fig. 2.16(b) から割引率 $\gamma = 0.4 \sim 0.9$ で、エピソード数に学習の

効果が見られる。しかし、Q-learning と同様にシミュレーション環境から計算上求められる最短経路ステップ数 14 を獲得することができなかった。学習率 $\alpha_v=0.5$ および $\alpha_q=0.7$, 割引率 $\gamma=0.7$ の場合のエージェントが獲得した移動経路を Fig. 2.17 に示す。エージェントがスタートからゴールまで直線的に移動していることが分かる。このときのエピソード数, ステップ数および衝突回数はそれぞれ 34, 15 および 0 であった。

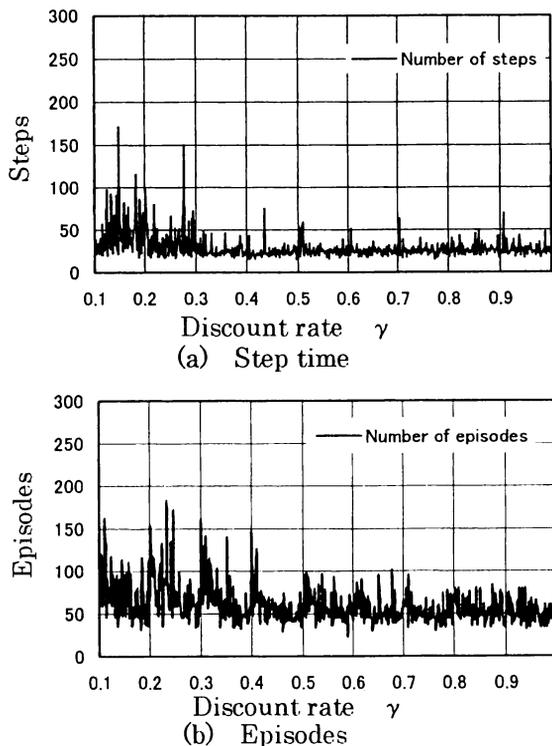


Fig. 2.16 Steps and episodes by Hybrid-learning

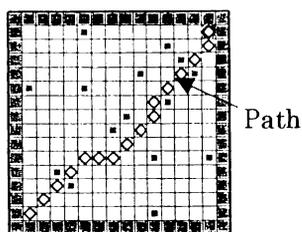


Fig. 2.17 Learned path by Hybrid-learning
 $(\alpha_v=0.5, \alpha_q=0.7, \gamma=0.7)$

Hybrid-learning は、Table 2.1 に示すように PSP と比べると学習終了時のステップ数およびエピソード数共に少なく優れており、また Q-learning と比べると学習終了時のステップ数は同程度であるが、エピソード数は少なく優れている。したがって、Hybrid-learning は学習終了時のステップ数が少なく、エピソード数も少し差はあるものの全体として少な

く、獲得経路も準最適であるといえ、Q-learning 同様に自律移動ロボットの行動生成に適した強化学習法であるといえる。

Table 2.1 Steps and episodes by each methods

	PSP	Q-learning	Hybrid-learning
Steps	41	15	15
Episodes	34	43	34

2.3.4 強化学習(RL)に対する考察

強化学習の代表的な手法である、経験強化型の強化学習である利益共有法(Profit sharing plan : PSP)および環境同定型の強化学習である Q-learning に対して、自律移動ロボットの障害物回避最短経路探索問題によって、機械学習法としての有効性についてのシミュレーション実験検証を行った結果、設計者は報酬を設定するのみで、ロボットが自律的に適切な行動について学習を行い、自律移動ロボットの学習機能の枠組みとして適していることが分かった。また、PSP は学習速度が速いが獲得経路についての最適性の保証がなく、Q-learning は学習速度が遅いが獲得経路についての最適性が保証されているという得失についても確認することができた。

そこで、PSP と Q-learning のそれぞれの利点を生かす手法として Hybrid-learning を提案し、自律移動ロボットの障害物回避最短経路探索問題に対するシミュレーション実験によって本提案手法の有効性を示した。すなわち、Hybrid-learning が最適性および学習速度の両方の点から他の強化学習法と比べて有効であることを示し、Hybrid-learning が静的環境での自律移動ロボットの行動生成に有効な手法であることを示した。ただし、時々刻々と状況が変化する動的環境下での限界性についての検討に関しては、参考文献 7) を参考にしていきたい。

2.4 おわりに

行動ベース型知能ロボット処理系の概念である包摂アーキテクチャ (Subsumption Architecture : SA) の自律移動ロボットに対する有効性を検証するために、実験モデル機を製作し、事前に障害物などの環境に関する情報をまったくロボットに与えず、点光源をゴールとした二次元平面の障害物回避経路探索問題と先導ロボットと追尾ロボットの 2 台のロボットによる協調物体搬送問題に対して実験を行った。実験結果として、環境ならびに行動目的が比較的単純な場合には条件反射的行動(昆虫の自律行動)を実行することから非常に効率良く動作し、自律移動ロボットの行動生成機能として有効である。しかし、SA そのものの概念では、装備する行動モジュールならびに行動モジュール間の競合解消(発火順位付け)方法については設計者があらかじめトップダウン的に決定してやらねばならず、より複雑な環境に対応させ、より高度な知的処理を行わせようとした場合、それらの問題点を含めた自律的な学習機能が必要であることが分かった。

そのため、自律的な学習機能として、目標状態を与えれば、そこに至る動作・行動列の試行錯誤を通じて、自律的に行動戦略を獲得する学習システムとして、強化学習(Reinforcement Learning : RL)法を採り入れることとした。そこで強化学習の代表的な評価手法である経験強化型の利益共有法(Profit sharing plan : PSP)ならびに環境同定型のQ-learning についての有効性検証を 2 次元平面上の障害物回避最短経路探索問題に対するシミュレーション実験によって行った。そのシミュレーション実験結果によって得られたPSPおよびQ-learningの得失を踏まえて、それらをハイブリッド化した“Hybrid-learning”と呼ぶ強化学習法を本研究において提案し、同一の障害物回避最短経路探索問題のシミュレーション実験によって、本提案手法がPSPおよびQ-learningより獲得経路の最適性ならびに学習速度の両方の点で優れていることを示した。

参 考 文 献

- 1) J. L. Jones, A. M. Flynn, “Mobile robot – Inspiration to implementation”, AK Peters, Ltd, 1993 [邦訳：熊切康雄, “移動ロボット 基礎科学と応用”, トップラン, 1996
- 2) Newton Research Labs : Interactive C User’s Guide
- 3) Brooks, R. A. “Intelligence Without Reason”, MIT AI Lab Memo 1293, April, 1991
- 4) R. S. Sutton, A. G. Barto : “Reinforcement Learning”, The MIT Press, 1998
[邦訳：三上貞芳, 皆川雅章, “強化学習”, 森北出版, 2000
- 5) 堀内匡, 藤野昭典, 片井修, 榎木哲夫, “経験強化を考慮した Q-learning の提案とその応用”, 計測自動制御学会論文, Vol.35, No. 5, pp. 645-653, 1999
- 6) 木村元, 宮崎和光, 小林重信, “強化学習システムの設計指針”, 計測と制御, Vol.38, No.10, pp. 618-623, 1999
- 7) 竹口知男, 小尻一憲, 大橋美奈子, 今井弘之, 能勢和夫, “強化学習によるロボット移動経路の探索”, 日本機械学会関西支部第77回定期総会講演会講演論文集, No.014-1, pp. 5-47-5-48, 2001

3. 制御系の状況認識と定量的表現

3.1 はじめに

自律走行をする移動体（ビークル）が指定された経路あるいは経路を選択して目的地に到達する過程において、適切な判断を行うためには走行に必要な周辺状況を的確に得る必要がある。本分担研究ではその条件となるルート、ロケーション、分岐点、障害物に対して数的表現としての情報を得ることを目的とした。まず、第一の課題として、どのようなビークルをいかなる環境で走行させるか、第二の課題は周辺状況を定量的にどう表現するかということが主たる内容となる。しかし、ターゲットをそこにおきながらも付随的な検討が必要でもあったので、これらも含めて以下に報告する。

3.2 ビークルと走行環境

過去において、ビークル自体は色々と試作し、誘導方式についても種々検討を行った。本研究においては 2 軸独立駆動の動輪を円形車体の左右に配置し、その中央部の前後にキャスターを取り付けたビークルを採用した。これによって前進後進はもとより極地旋回によって全方向への誘導が動輪の回転制御のみで行える利点を得た。このビークルに後述する走行制御用のセンサおよび CPU ボード、無線 LAN によってサーバと通信を行うクライアントとしての携帯型パソコンを搭載している。Fig. 3.1 に今回使用した模型ビークルとそのシステム構成を示す。

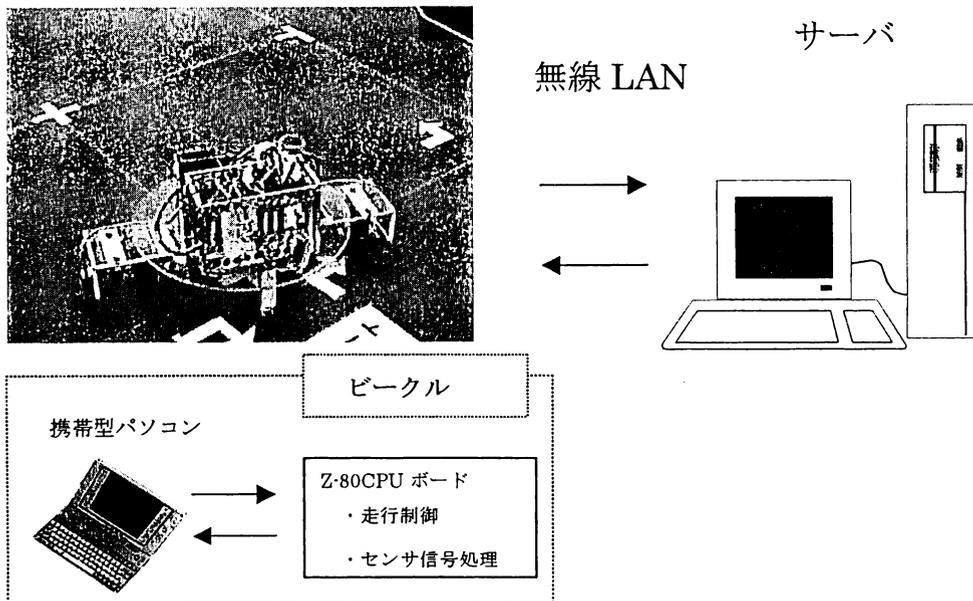


Fig. 3.1 模型ビークルとシステム構成

走行する場所の環境として、マトリックス状に走行路が敷設できる工場、作業場所、置き場などを想定した、この経路上の交叉点および交叉点と交叉点の間の走行路（経路要素）はサーバに格納され、クライアントとしての各ビークルの行き先および現在位置が無線LANによって送受信される。実際にはこのビークルに載せられた加工物が経路要素の加工ショップで受ける加工情報や搬送する物の情報などを統括するデータベースを支える搬送システムの基本となる部分である。

3.3 ロケーションの認識

経路上を走行するビークルが自分のロケーションを認識する方法は経路要素内、交叉点にセンサに反応するものを置くことによって得られるが、他のビークルの位置を認識するためには何らかの通信手段を講じなくてはならない。本研究ではネットワークを利用することによって問題解決を行った。すなわち、ビークルに搭載したセンサが標識となる磁気あるいはマークを検出して、その信号をCPUボードから携帯型パソコンに転送する。更にその信号はサーバに送信され、すべてのビークルが他のビークルの経路要素上の位置と交叉点のファイル情報を共用できる。携帯パソコンは次に入る交叉点とそれに繋がる経路要素の情報をサーバから得て走行指示をCPUボードに送る。交叉点への進入はルールを作り追突、衝突を回避させる。このようにして複数台のビークルが互いのロケーションを認識しながら走行できる上、携帯型パソコンの採用によってCPUボードの負担が軽減されるので路面情報や周辺情報が得やすいメリットも生じた。

3.4 誘導路の認識

ビークルの製作上の整合性、走行時の路面とのスリップ等によってビークルは誘導ラインから外れて行くのは普通である。路面に白色テープあるいは磁気テープを貼り付け、それに沿って走行させる場合でも、センサの取り付け場所、左右の間隔、駆動輪の調整回転数の差によって蛇行の周期と振幅が異なる。ビークルの走行目的によって蛇行幅も変わると思われる。例えば、レールのように左右にラインを貼り、その中でふらつきながら走行させる場合の制御は比較的やりやすいが、分担研究の目的から外れ、量を認識して制御を行うことに少々固執した。結論として、前後のキャストの部分に8個の光センサを持つユニットを取り付け、前後8ビットのポート(8255)入力として外れ量と外れ角度を量的に検出した。

3.5 経路上の障害物の認識

想定したクローズなルートにおいて、進路を妨げる物が何であるかを知るより、障害物までの距離を知る方が实际的であり、現時点では最も実用的と思われる。しかし当初は上記誘導路の認識も含めて障害物を認識してそれを回避する方法について検討を行った。最初の発想として、次の3方式を考えたが、結論として設備、費用、力量など諸々のリソー

ス不足から最初の二つは断念したが一応紹介しておく。第一は、ビークル上方に円錐形あるいは双曲面のミラーを取り付け、直下からビデオカメラで 360 度のパノラマを撮影して全方向の風景を画像変換し、周辺情報を獲得することであった。将来的にこの方法は周辺監視などのメカニカルな方法に取って代わると考えるが、これを実現して更に障害物を認識するまでにはもう少し時間が必要である。第 2 の方法は、ビークルの上面に何らかの認識物を取り付け、経路の上方から撮影することによって障害物の認識およびその回避を検討したが、現在利用できる携帯型パソコン（ノート型で CPU：200MHz）程度ではどうにもならないこともあり、第三の方法について基礎的な実験を行った。

その内容は黒いマットの上に 1 台の模型ビークル（後述）と 1 つの障害物を置き、上方からデジタルビデオカメラで撮影し、この障害物を画像処理手法を利用して回避するものである。ビークルはラグウォーリア（LagWarrior：移動ロボットの品名で Motorola 68 系プロセッサ搭載）を使用した。このビークルには対話型の C 言語が用意されており、パソコンで作成したプログラムのダウンロードや実行の操作が簡単である。Fig 3.2 にシステムの概略図を示す。

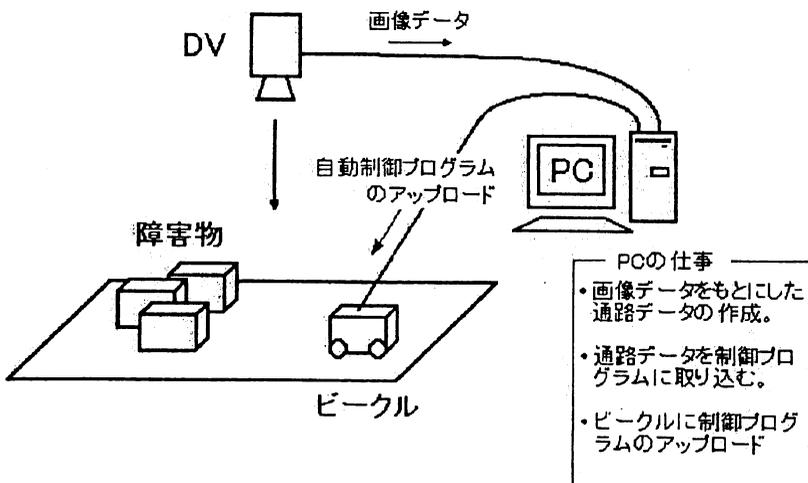


Fig. 3.2 基礎実験システム概略図

実験手順はまずビークルと障害物をマットの上に置き、デジタルビデオカメラで撮影し、カメラ付属のソフトでパソコンに取り込む。次は画像処理であるが、近年画像解析の方法は日進月歩の状態で、パソコン性能の向上と相まってますます加速すると考える。ここで使用した方法は最もポピュラーで、精密な画像情報を必要としないためにハードに負担のかからない手法を選んだ。「白黒化」「縮小化」のために 1 画素の RGB を R の値に統一し、画面を 1/9 に縮小するために縦 3×横 3 の 9 つの画素の中心に統一した。また、床面との差別化のためにしきい値を設けて「2 値化」を行った。これによって得られた 2 値化画像は「1」

と“0”の幾何学的特徴を表現できる。Fig. 3.3 にビデオ画像を Fig. 3.4 に縮小画面を Fig. 3.5 に 2 値化処理結果を示す。

なお、原画 640×480 の 1/9 の縮小時間は 15 秒、2 値化処理時間は約 147 秒であった。次に 2 値化された対象を具体的な数値にするための境界線追跡を行うが、目的が画像処理でなく「ビークルの障害物回避」であるため、2 値の境界を追跡すると同時に境界の最大座標と最小座標からオブジェクトを四角形で表現し、処理速度の向上を図った。Fig. 3.6 に境界線追跡結果を Fig. 3.7 に四角化結果を示す。なお、これに要する処理時間は 10 秒であった。

最後はパターン認識と回避走行である。本来、パターン認識には膨大な時間と労力を要するが、ここでは四角化を行っているのでそのオブジェクトの比較は評価式を作るまでも無く安易に処理できる。また、回避もビークルを直角に移動させる前提からこれも簡単である。以上のように非常に単純なモデルで障害物の定量的表現を試みたが、最初に述べたように実現までにはハードの性能向上をもうしばらく待たねばならない。

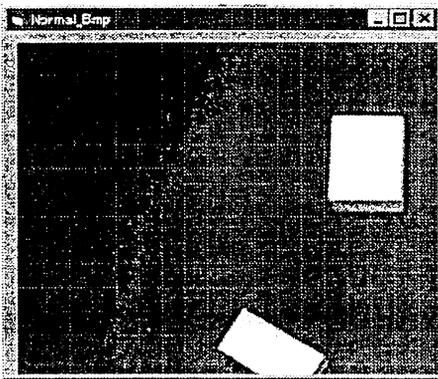


Fig 3.3 ビデオ画像

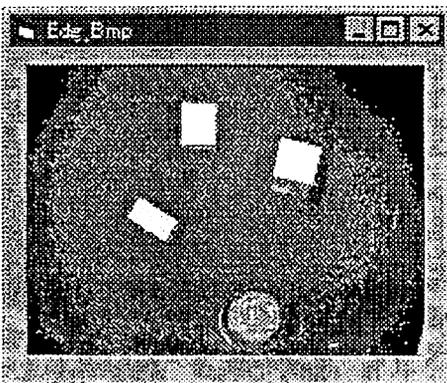


Fig 3.4 縮小画像

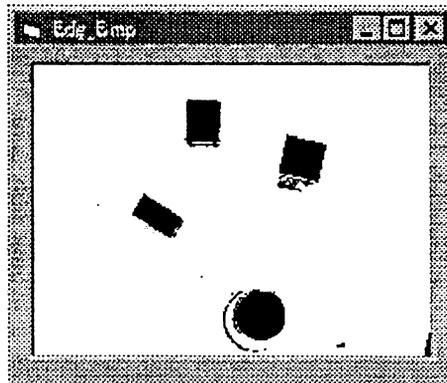


Fig 3.5 2 値化処理結果

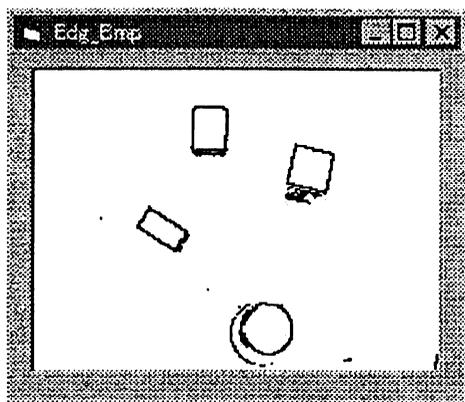


Fig. 3.6 境界線追跡結果

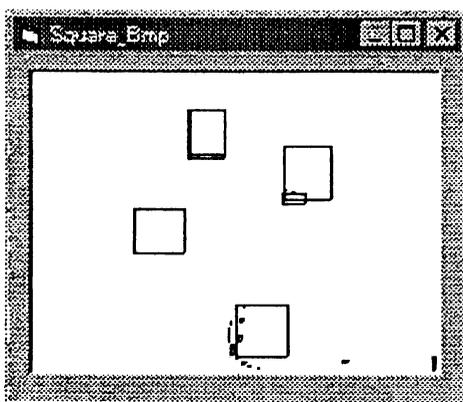


Fig. 3.7 四角化結果

3.6 考察およびその他の実験について

移動するビークルの制御に関する色々な対象について量的な表現による情報の獲得目的として実験を行ってみた。その経験から将来の移動体の制御は視覚すなわち画像処理に負うところが大きいと考える。また聴覚の情報を得るための基礎的な実験として、騒音計、FFT による色々な試行も確証の得られる方向は残念ながら見出せなかった。しかし指向性のある音を取り出す方法、例えば電波に音声を乗せるように超音波に音声を乗せればアンプも不要で筒の中を通るような音が空間で獲得できる。この技術が実用できれば光と同様に有効な応用が期待できると考える。

参考文献

- [1] 宮井, 岩佐: 磁気テープによるビークルの制御, 日本機械学会創立 100 周年記念北陸信越支部講演会講演論文集, pp. 269-270, 1997
- [2] 宮井, 岩佐, 藤原: ネットワークを利用した自律型ビークルの誘導に関する研究, 第 42 回システム制御情報学会研究発表講演会講演論文集, pp. 505-506, 1997

4. 自律エージェントの知識獲得

「複数車両の協調運転操作の生成に関する研究」

4.1 はじめに

制御知識の獲得の第一歩として、車の運転操作の自動生成問題に取り組んだ。車の運転操作の生成手法には、いくつかの研究がある。Ferbach[1]は、運転操作の生成問題を最適制御問題として捉え、動力学に基づく車両の運動方程式を導入し、動的計画法を用いる解法を提案している。車両の運動を記述する決定変数の数が多く、解の探索に必要な計算量も多い。梶谷ら[2]は、ファジィ理論を用いて操作パターンを求める手法を提案している。この手法では、予め教示コースを入力する必要があるため、教示コースにあてはまらない走行環境が出現すると、解が求まらない可能性がある。また、走行環境が複雑になると教示コースも複雑になり、結果として膨大なデータベースが必要で、解を得るまでの計算量も増大してしまう。井上ら[3]は、トレーラ・トラックのルールに基づく後退運転制御問題を取り上げている。走行途中に障害物がある場合など、予め想定しているルール以外の状況が発生すると解が求まらない。

これらに対して、本研究では教示コース設定などの必要がなく、また、走行環境に依存することなく、前進、後退や停止などの運転操作を巧みに活用し、障害物を回避しながら最適な運転操作を生成することが出来る手法について考察した。Ackerman[4]によると、車両が低速度走行するとき、舵角と速度を与えると車両の移動軌跡は幾何学的な関係で求められる。そこで、車の運転操作生成問題を最適化問題として捉え、単位時間幅毎の運転操作量を決定変数とするとともに、車両の移動軌跡に対して評価関数を設定し、進化計算手法で最適化をはかることを考えた。この方法は、数値的にはあるが、走行環境に依存せずに解を求めることができる。したがって、後退でハンドルを切り返しを行う車庫入れや、巧みな運転操作を必要とする縦列駐車問題にも適用可能であることが、ここでは、とくに、複数台の車の交差点や狭い道路での同時協調運転操作について考察する。以下の記述は、文献[5][6]の内容を記述したものである。

なお、この手法を用いた二輪操舵車や四輪操舵車の車庫入れや縦列駐車、トレーラ・トラックの車庫入れやプラットフォーム駐車については文献[7][8]を参照されたい。

4.2 運転操作自動生成の方針

運転操作最適化問題とは、図 4.1 に示すように、出発位置と目的位置を与えたとき、移動に必要な運転操作時系列を自動生成する問題である。ここで考える同時協調運転操作問題では、車両は道路上に 2 台存在するものとする。計算に必要な初期設定条件として走行速度、車両寸法、初期位置、目標位置、時間刻み幅、障害物位置、道路環境を与える。それぞれの車両の時間刻み幅毎の実舵角と進行方向を決定変数として設定する。すなわち、運転操作の時系列表現を決定変数とする。ただし、運転操作問題では、時系列長さも決定

変数である。決定変数を入力，車両の移動を出力とする車両の移動モデルを設定し，これらの決定変数を移動モデルに入力することにより車両の移動軌跡を求める。また，道路形状および対向車線車両は障害物と見なす。そして，得られた車両の運転軌跡に対する評価式を設定する。評価規範としては，移動軌跡の滑らかさ，移動距離，対向車との接触回数などである。この時，協調運転操作最適化問題とは，各車両の出発位置と目標位置が与えられた時，障害物と衝突することのない評価値の高い移動軌跡を生成する実舵角と進行方向の時系列（運転操作パターンという）を求める問題となる。そして，協調運転操作パターンは，最適解を探索ことで数値的に求められる。

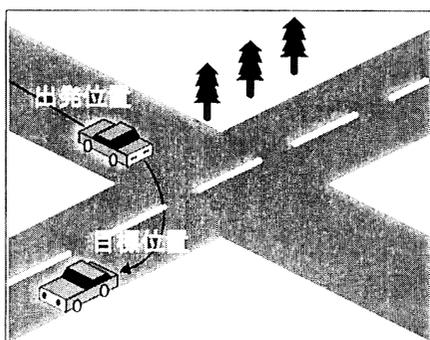


図 4.1 運転操作最適化問題

4.3 最適化問題

問題の非線形特性を考慮して，探索方法としては，4.4 で述べる適応機能を有する進化計算手法を用いる。ここでは，まず，制約条件である車両の移動モデルについて述べ，ついで，進化計算手法に適した決定変数と評価関数について述べる。

4.3.1 車両の移動モデル

車両として二輪操舵車を考える。低速度走行のとき，車両の移動軌跡はアッカーマンステアリングジオメトリ[4]に従う。車両は実舵角 δ で極めて低速度 V で走行を行っているものとする。この時車両は図 4.2 に示すような円旋回運動を行い，式(4.1)(4.2)の関係が成り立つ。

$$\delta = \frac{\ell}{R} \tag{4.1}$$

$$\Delta\theta = \frac{V \cdot \Delta t}{R} \tag{4.2}$$

ここに， ℓ は車両のホイールベース長， R は回転半径， O は回転中心， Δt は時間刻み幅である。 Δt ごとに舵角 δ と進行方向を与えると，式(4.2)によって，車両の移動軌跡を計算することができる。

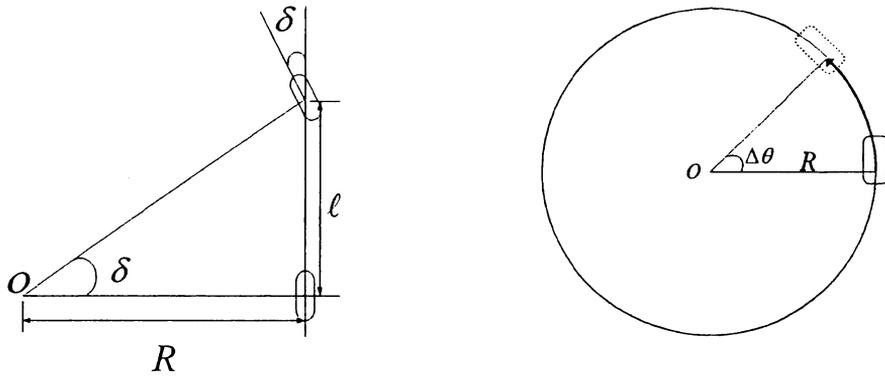


図 4.2 車両の二輪モデルと円巡回運動

4.3.2 決定変数

決定変数である操作パターンは、表 4.1 に示すように、時間幅 τ 毎の実舵角の変更量、進行方向および時間幅の 3 組の遺伝子の系列とする。舵角変更量は

$$\Delta\delta \in \{-\theta, 0, \theta\} \quad (4.3)$$

とする。舵角に対する上下限制約として $\delta \in [-D, D]$ を設ける。進行方向は

$$\rho \in \{\text{forward}, \text{stop}, \text{backward}\} \quad (4.4)$$

とする。時間幅は

$$\Delta t \in \{0, \tau\} \quad (4.5)$$

とする。このとき、時刻 $t = \sum_{i=0}^n \Delta t_i$ における実舵角は $\delta(-1) = 0$

$$\delta(t) = \delta(t-1) + \Delta\delta_n \quad (4.6)$$

で与えられる。

表 4.1 車両の運転操作パターン

操作 No.	0	1	2	...	N
$\Delta\delta$	$\Delta\delta_0$	$\Delta\delta_1$	$\Delta\delta_2$...	$\Delta\delta_N$
ρ	ρ_0	ρ_1	ρ_2	...	ρ_N
Δt	Δt_0	Δt_1	Δt_2	...	Δt_N

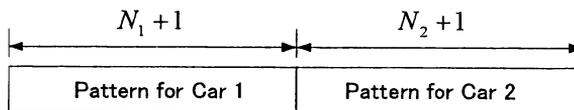


図 4.3 染色体

そして、図 4.3 に示すように 2 台の車の運転操作パターンを合わせて個体の染色体とする。各車両の運転時間長は事前には分からない。そこで、探索の結果として運転時間長は求まるものとし、便宜上、染色体長 N_i に上限 (4.7) を設ける。なお、初期染色体長は、車両 1、車両 2 の初期位置と目標位置の距離に比例した長さを参考にして、適当に設定する。

$$N_i \leq \frac{L}{2} \quad (i = 1, 2) \quad (4.7)$$

4.3.3 評価関数

2台の車の移動軌跡に対して、個体の適合度 (*fitness*) を式 (4.8) で評価する。

$$fitness = \frac{1.0}{\alpha_a \cdot a + \alpha_b \cdot b + \alpha_c \cdot c + \alpha_d \cdot d + \alpha_e \cdot e} \quad (4.8)$$

ここに、

- a* : 運転操作終了時点における車体位置と目標位置の差
- b* : 障害物と車体の接触回数
- c* : 舵角変化の絶対値の累積値
- d* : 移動距離の累積値
- e* : 車同士の接触回数

である。また、 $\alpha_a \sim \alpha_e$ は評価項目の重み係数である。したがって、適合度の高い個体が優れた移動軌跡を生成することになる。

4.4 進化計算

提案する進化計算は、図 4.4 に示す適応機能を有する遺伝的アルゴリズムである。このアルゴリズムの特徴は、複数の遺伝的操作を、単純遺伝的アルゴリズムのように直列的に施すのではなく、並列的に配置し、かつ、その選択を適応的に施す点にある。すなわち、解の探索段階に応じて効果的な遺伝的操作を確率的に多く施し、結果として探索効率を上げようと意図するもので、交通流制御問題でこの方法の有効性を確認している[9]。ここでの新たな試みは、時系列長の適正值をも探索できるように並列配置の遺伝的操作に挿入、削除処理、ならびに協調運転操作実現のためにゼロ時間幅の挿入処理を新たに導入した 2点である。主な処理内容を以下に示す。

(a) 交叉処理

ルーレット選択により親個体群の中から 2 個体を選択する。確率的に交叉処理位置を決めて一点交叉を施す。2 個の子個体が生成される。交叉処理は車両 1 と車両 2 の部分染色体で別々に行う。各個体の染色体長は異なるが、各個体の交叉点は同じ位置とする。すなわち、2 個体の染色体長さを N_1, N_2 、交叉点を $p-1, p$ 間とすると、 $1 \leq p \leq \min\{N_1, N_2\}$ とする。

(b) 突然変異処理

ルーレット選択により親個体群の中から 1 個体を選択する。確率的に突然変異処理位置を決めて突然変異処理を施す。ただし、時間幅は $\Delta t = \tau$ とする。1 個の子個体が生成される。突然変異処理は車両 1 と車両 2 の部分染色体で別々に行う。

(c) 挿入処理

ルーレット選択により親個体群の中から 1 個体を選択する。確率的に挿入処理位置を決めて挿入処理を施す。ただし、時間幅は $\Delta t = \tau$ とする。1 個の子個体が生成される。挿入処理は車両 1 と車両 2 の部分染色体で別々に行う。

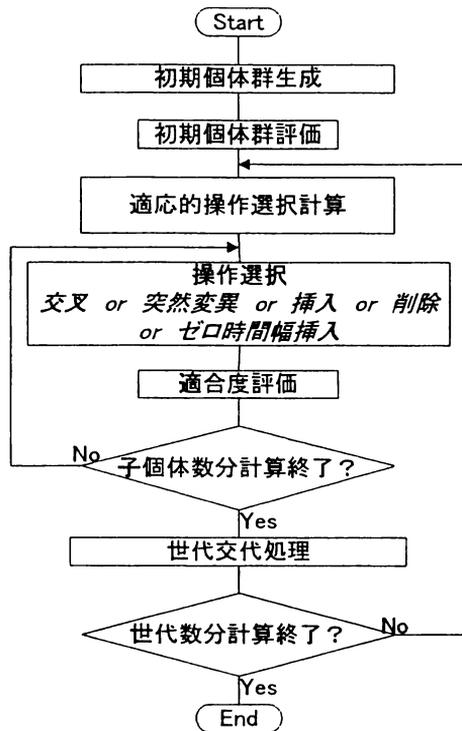


図 4.4 進化計算のフローチャート

(d) 削除処理

ルーレット選択により親個体群の中から 1 個体を選択する。確率的に削除処理位置を決めて削除処理を施す。1 個の子個体が生成される。削除処理は車両 1 と車両 2 の部分染色体で別々に行う。

(e) ゼロ時間幅の挿入処理

複数台の車の同時協調運転では、他の車が通り過ぎるのを待つという動作が発生する。一方の車の時間幅が $\Delta t = 0$ でもう一方の車の時間幅が $\Delta t \neq 0$ であるなら相対的に車の停止動作が生成されると考え、つぎのゼロ時間幅の挿入を導入した。

ルーレット選択により親個体群の中から 1 個体を選択する。確率的に挿入処理位置を決めて $\Delta t = 0$ なる遺伝子の挿入処理を施す。1 個の子個体が生成される。但し、車両 1 と車両 2 どちらか一方にのみ挿入処理を施す。この時、対である ρ, δ の染色体には何も挿入しない。

(f) 適応的操作選択

図 4.4 に示したように、上述の遺伝的操作としては交叉 (C)、突然変異 (M)、挿入 (I)、削除 (D)、およびゼロ時間幅の挿入 (ZI) を並列的に並べ確率的に選択するものとする。各遺伝的操作は適合度向上に効果的な操作を高い確率で選択する。そのために、各遺伝的操作別に親集団と子集団の平均適合度を計算し、各操作の $j-1 \sim j$ 世代間の適合度向上率を式 (4.9) (4.10) (4.11) で求める。

$$f_r^j = \frac{O_r^j}{P_r^j} \quad r = C, M, I, D, ZI \quad (4.9)$$

$$\bar{f}_r^j = \lambda f_r^j + (1 - \lambda) \bar{f}_r^{j-1} \quad \bar{f}_r^0 = 1 \quad (4.10)$$

$$S_r^j = \frac{\bar{f}_r^j}{f_C^j + f_M^j + f_I^j + f_D^j + f_{ZI}^j} \quad S_r^0 = \frac{1}{5} \quad (4.11)$$

ここに、 $\lambda \in [0,1]$ は平滑係数である。この S_r^j を基に各遺伝的操作をルーレット選択することにより、適応的操作選択機能を実現する。

(g) 世代交代

世代交代は、エリート戦略を用いる。すなわち、個体群の集団数を μ とするとき、 $\mu - 1$ 個生成した子個体と適合度の最も高い親個体 1 個とで次世代の親個体群を形成する。

4.5 シミュレーション

4.5.1 シミュレーション条件

実際の走行環境に近い車両寸法、走行環境を設定した。図 4.5 は車両サイズを示したものである。図 4.6 は狭い道路 1 の走行環境である。普通車両と大型車両が存在する。お互いが衝突しないよう走行するには普通車両は大型車両が大きな道路から出てくるのを待たなくてはならない。図 4.7 は狭い道路 2 の走行環境である。車が 1 台入れる位の退避場所を設定した。普通車両と大型車両が存在する。お互いが衝突しないよう走行するには普通車両は大型車両が通り過ぎるのを設定された退避場所で待たなくてはならない。図 4.8 は信号機のない交差点の走行環境である。普通車両が 2 台存在する。1 台の車両は交差点を右折し、もう 1 台の車は交差点を直進するものとする。お互いが衝突しないように走行するには、通常、直進車が交差点を通過後、右折車が交差点に進入することになる。狭い道路 1, 2 および交差点の計算に用いた入力パラメータを表 4.2 に示す。

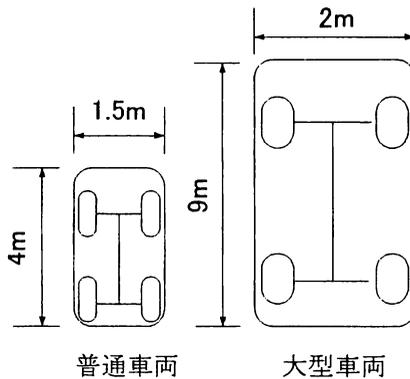


図 4.5 車両のサイズ

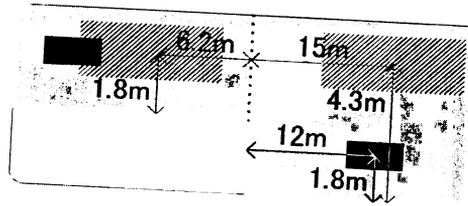
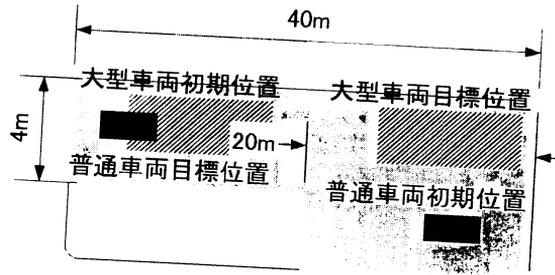


図 4.6 狭い道路 1

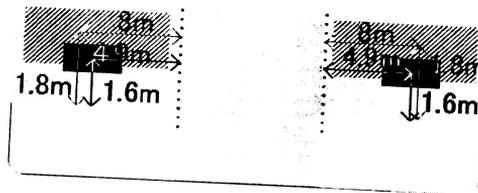
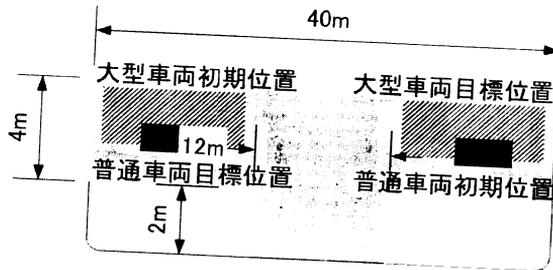


図 4.7 狭い道路 2

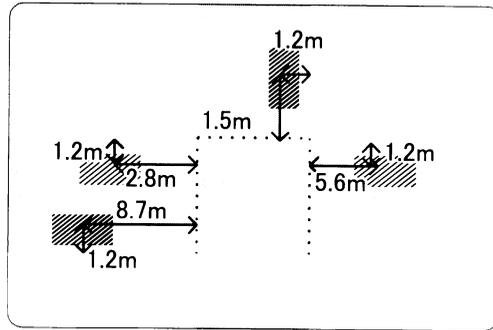
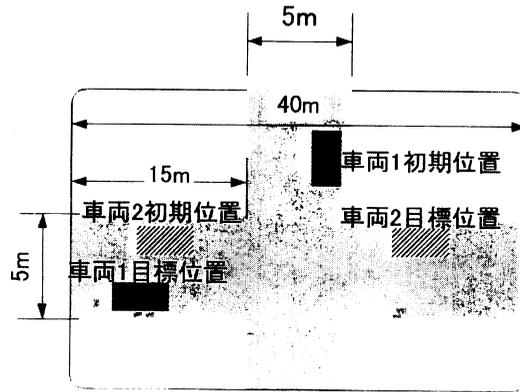


図 4.8 交差点

表 4.2 パラメータ設定

	狭い道路 1	狭い道路 2	交差点
親個体数	50	50	50
子個体数	49	49	49
世代数	350	350	350
$V(cm/sec)$	4	4	4
$\Delta t(sec)$	4	4	3
$\Delta \theta(deg)$	5	5	5
$D(deg)$	± 30	± 30	± 30
α_a	1	1	1
α_b	3	3	3
α_c	1	1	1
α_d	0.5	0.5	0.5
α_e	2	2	2
λ	0.045	0.045	0.045

4.5.2 狭い道路1

図4.9は退避場所の有る狭い道路において普通車(車両1)と大型車(車両2)がすれ違う時の移動の経過を示したものである。大型車が左側から出発し、直進する一方、普通車は右側から出発して待避場所で大型車が通過するのを待ち、直進するという運転操作が得られた。

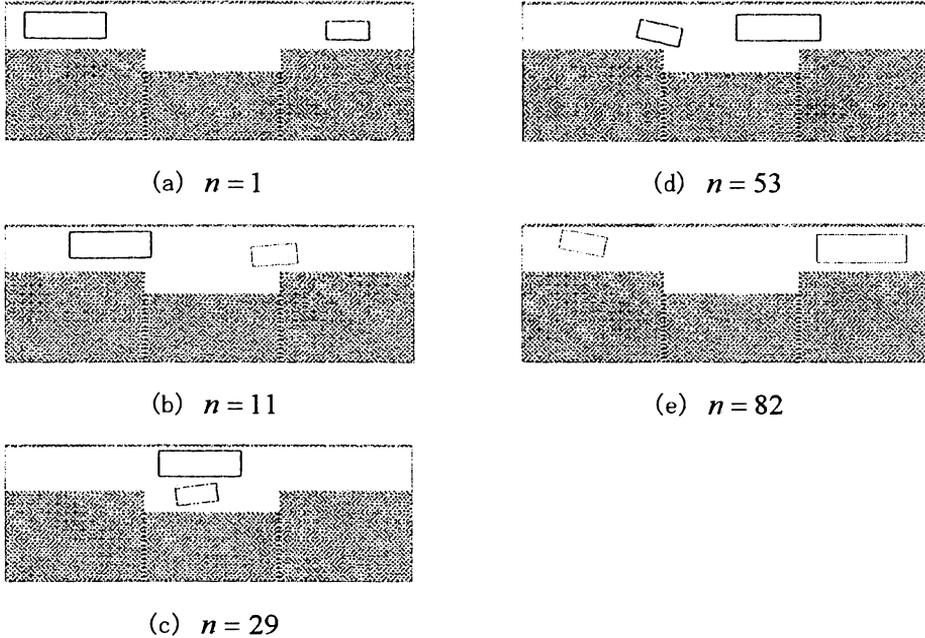


図4.9 狭い道路1における車両1, 2の移動

4.5.3 狭い道路2

図4.10は狭い道路2へ進入しようとする普通車(車両1)と、狭い道路から進んでくる大型車(車両2)がすれ違う時の移動の経過を示したものである。まず、大型車が左側の狭い道路から出発し、広い道路に出てきた後、右側から出発した普通車が狭い道路へ進入するという運転操作が得られた。

4.5.4 交差点

図4.11は信号のない交差点において直進する車(車両1)と右折する車(車両2)の移動経過を示したものである。右折車両は交差点の手前で停止し、直進車両が交差点を通り過ぎるのを待って交差点に進入し、右折するという運転操作が得られた。

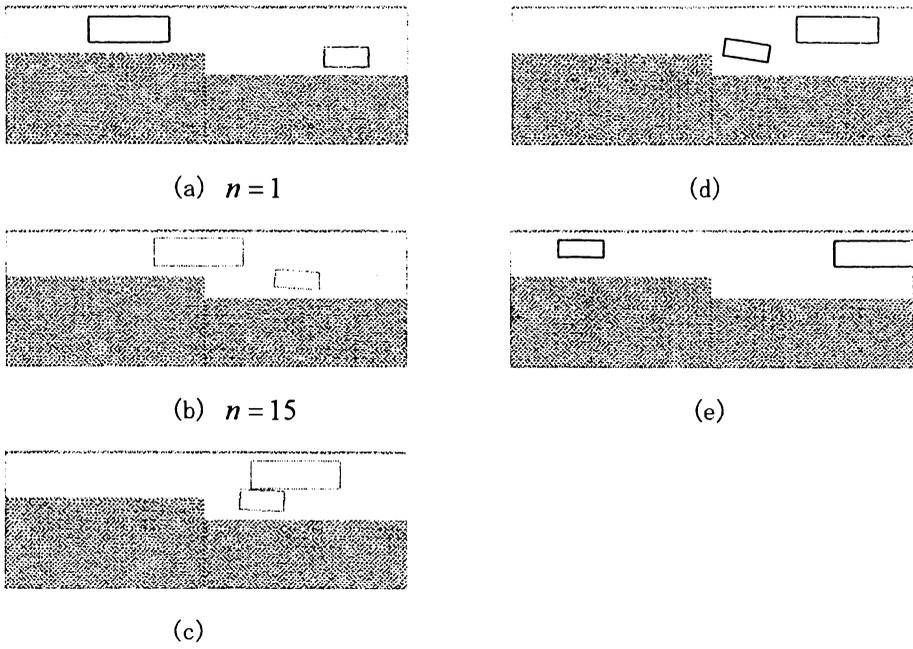


図 4.10 狭い道路 2 における車両 1, 2 の移動

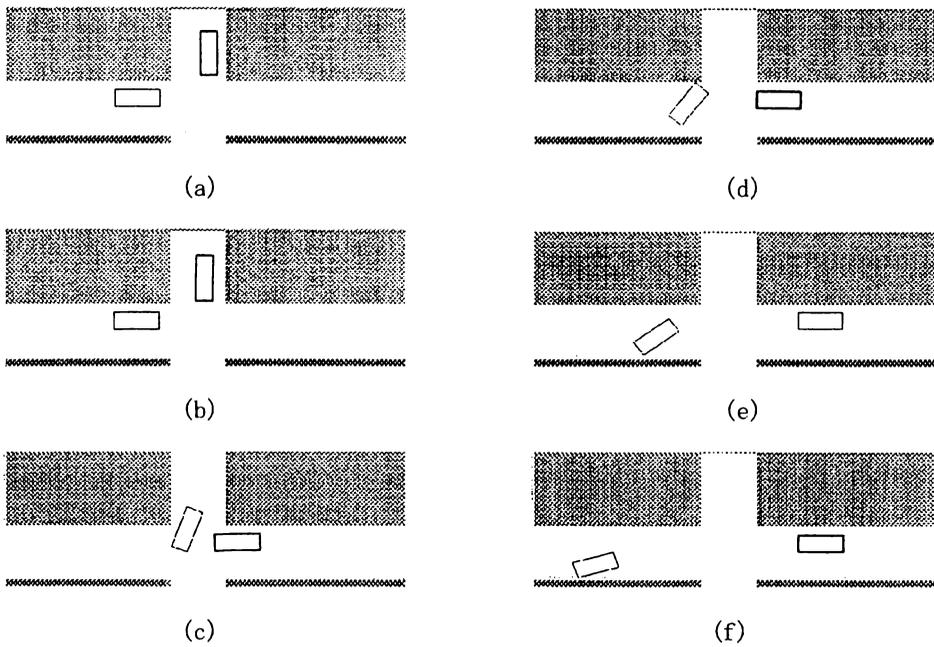


図 4.11 信号のない交差点における車両 1, 2 の移動

4.6 考察

図 4.12~4.14 は狭い道路 1 の計算で得られた遺伝的操作の選択確率，染色体長さ，最高適合度の 350 世代間におけるそれぞれの推移を示したものである。いずれも，10 回計算した平均値である。図 4.12 の操作選択確率の推移グラフによると，約 150 世代以降では遺伝的操作の選択確率はゼロ時間幅挿入処理が一番高く，ついで，交叉，削除，突然変異，挿入の順となっている。

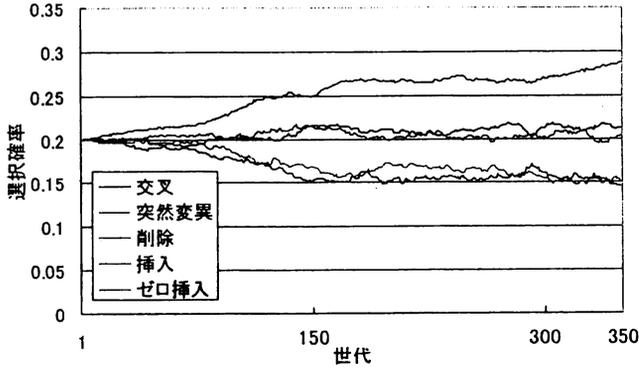


図 4.12 選択確率の推移(狭い道路 1)

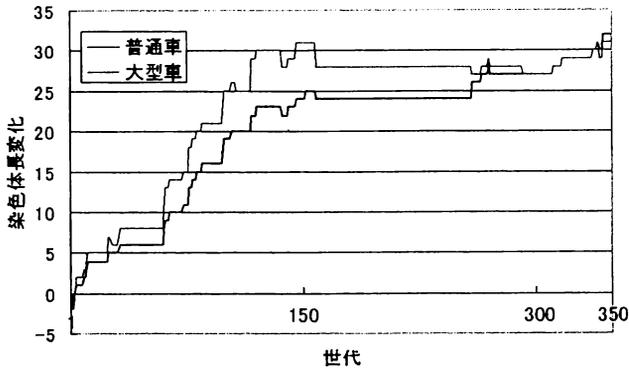


図 4.13 染色体長さの推移(狭い道路 1)

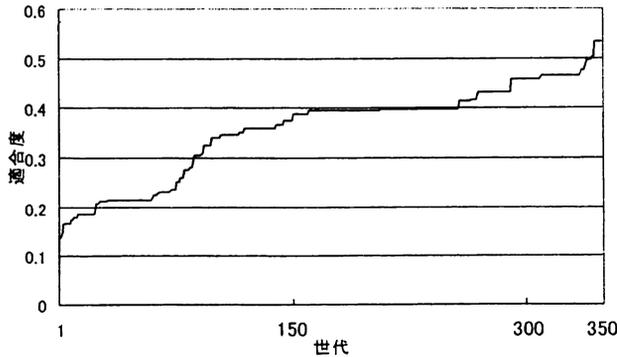


図 4.14 適合度の推移(狭い道路 1)

また、図 4.13 は染色体長さの推移を示したものであるが、2 台の車両の染色体長さは増加傾向にあり、約 150 世代以降に安定し始めている。図 4.14 の最高適合度の推移と対比してみると、染色体長さが長くなることにより適合度が向上していることがわかる。挿入の選択確率が低いことから、染色体長さが長くなるのは、ゼロ時間幅挿入によるものであると考えられる。狭い道路 2 および交差点の計算結果も同様な傾向であった。したがって、同時協調運転操作ではゼロ時間幅挿入が大変有効であると判断される。

4.7 おわりに

狭い道路や交差点における 2 台の車両の同時協調運転操作の生成問題を考察し、従来の進化計算法に改良を加えたアルゴリズムを提案した。数値計算の結果ではあるが、改良点であるゼロ時間幅挿入処理の導入が同時協調運転操作の生成に有効であることが検証できた。ただ、現在の手法では、車両の台数が増加すると染色体長さも比例的に増加してしまうため、計算時間も増大してしまう。したがって、今後は、車両の台数に比例しない解の染色体表現方法を検討し、より多くの車両間での協調運転操作の生成に取り組んでいきたい。

参考文献

- [1] P. Ferbach : A Method of Progressive Constraints for Nonholonomic Motion Planning, IEEE Transactions on Robotics and Automation, Vol. 14, No. 1, pp. 172-179, 1998
- [2] 梶谷, 増田, 井上 : 異種個体集団が共存する GA によるファジィ推論ルールの獲得手法との比較および考察, 第 16 回ファジィシステムシンポジウム講演論文集, pp. 73-76, 2000
- [3] 井上, 亀井, 井上 : トレーラ・トラックのルールに基づく後退運転制御のための遺伝的アルゴリズムによるパラメータ決定, 第 38 回システム制御情報学会研究発表会講演論文集, pp. 39-40, 1994
- [4] 景山, 景山 : 自動車力学, 理工図書, 1984
- [5] 喜多, 能勢, 平松 : 適応的進化計算手法を用いた複数台の車の運転操作最適化, 第 44 回自動制御連合講演会前刷, pp. 548-551, 2001
- [6] M. Kita, K. Nose, A. Hiramatsu, H. Imai, T. Takeguchi : Cooperated Drive Operation of Two Cars Using Adaptive Evolutionary Computation, Proceedings of the Seventh International Symposium on Artificial Life and Robotics, pp. 401-406, 2002
- [7] 喜多, 能勢, 平松 : 適応的進化計算法を用いた車の運転操作の最適化, 第 43 回自動制御連合講演会前刷, pp. 225-226, 2000
- [8] 喜多, 能勢, 平松 : 適応的進化計算手法のトレーラ・トラック運転操作最適化への応用,

第 45 回システム制御情報学会研究発表講演会講演論文集, pp. 459-460, 2001

- [9] 能勢, 平松, 天目, 下浦: 高速道路の最適流入制御への進化計算手法の適用, システム制御情報学会論文誌, Vol. 15, No. 9, pp. 486-494, 2002

5. あとがき

以上の述べてきたように、本長期的共同研究の各分担研究によって、
分担研究「制御系の学習機能」においては、

行動ベース型知能ロボット処理系の概念である包摂アーキテクチャ (Subsumption Architecture : SA) の自律移動ロボットに対する有効性検証実験結果として、SA は環境ならびに行動目的が比較的単純な場合には条件反射的行動(昆虫の自律行動)を実行することから非常に効率良く作動し、自律移動ロボットの行動生成機能として有効である。しかし、装備する行動モジュールならびに行動モジュール間の競合解消(発火順位付け)方法については設計者があらかじめトップダウン的に決定してやらねばならず、より複雑な環境に対応させ、より高度な知的処理を行わせようとした場合、それらの問題点を含めた自律的な学習機能が必要であることが分かった

そのため、自律的な学習機能として、目標状態を与えれば、そこに至る動作・行動列の試行錯誤を通じて、自律的に行動戦略を獲得する学習システムとして強化学習(Reinforcement Learning : RL)法を採り入れることとし、強化学習の代表的な評価手法である経験強化型の利益共有法(Profit sharing plan : PSP)ならびに環境同定型の Q-learning についての有効性検証を 2次元平面上の障害物回避最短経路探索問題に対するシミュレーション実験によって行った。そのシミュレーション実験結果によって得られた PSP および Q-learning の得失を踏まえて、それらをハイブリッド化した “Hybrid-learning” と呼ぶ強化学習法を本研究において提案し、同一の障害物回避最短経路探索問題のシミュレーション実験によって、本提案手法が PSP および Q-learning より獲得経路の最適性ならびに学習速度の両方の点において優れていることを示した。

分担研究「制御系の状況認識と定量的表現」においては、

移動するビーグルの制御に関係する色々な対象について量的表現による情報獲得を目的として実験を行ってみた。その経験から将来の移動体の制御は視覚すなわち画像処理に負うところが大きであると考えられる。また聴覚の情報を得るための基礎的な実験として、騒音計および FFT による色々な試行も確証の得られる方向は残念ながら見出せなかった。しかし指向性のある音を取り出す方法、例えば電波に音声を乗せるように超音波に音声を乗せればアンプも不要で筒の中を通るような音が空間で獲得できる。この技術が実用できれば光と同様に有効な応用が期待できると考える。

分担研究「自律エージェントの知識獲得」においては、

狭い道路や交差点における車両 2 台の同時協調運転操作の生成問題を考察し、従来の進化計算法に改良を加えたアルゴリズムを提案した。数値計算の結果ではあるが、改良点であるゼロ時間幅挿入処理の導入が同時協調運転操作の生成に有効であること

が検証できた。ただ、現在の手法では、車両の台数が増加すると染色体長さも比例的に増加してしまうため、計算時間も増大してしまう。そのため、車両の台数に比例しない解の染色体表現方法を検討し、より多くの車両間での協調運転操作の生成について考慮する必要がある。

という研究成果を得た。また、本長期的共同研究の研究成果の一部については、つぎのように外部発表を行った。

- 1) 宮井, 岩佐: 磁気テープによるビークルの制御, 日本機械学会創立100周年記念 北陸信越支部講演会講演論文集, pp. 269-270, 1997
- 2) 宮井, 岩佐, 藤原: ネットワークを利用した自律型ビークルの誘導に関する研究, 第42回システム制御情報学会研究発表講演会講演論文集, pp. 505-506, 1997
- 3) 喜多, 能勢, 平松: 適応的進化計算法を用いた車の運転操作の最適化, 第43回自動制御連合講演会前刷, pp. 225-226, 2000
- 4) 竹口, 小尻, 大橋, 今井, 能勢: 強化学習によるロボット移動経路の探索, 日本機械学会関西支部第77回定期総会講演会講演論文集, No. 014-1, pp. 5-47 - 5-48, 2001
- 5) 喜多, 能勢, 平松: 適応的進化計算法を用いた複数台の車の運転操作最適化, 第44回自動制御連合講演会前刷, pp. 548-551, 2001
- 6) 喜多, 能勢, 平松: 適応的進化計算法のトレーラ・トラック運転操作最適化への応用, 第45回システム制御情報学会研究発表講演会講演論文集, pp. 459-460, 2001
- 7) M. Kita, K. Nose, A. Hiramatsu, H. Imai, T. Takeguchi: Cooperated Drive Operation of Two Cars Using Adaptive Evolutionary Computation, Proceedings of the Seventh International Symposium on Artificial Life and Robotics, pp. 401-406, 2002
- 8) 能勢, 平松, 天目, 下浦: 高速道路の最適流入制御への進化計算法の適用, システム制御情報学会論文誌, Vol. 15, No. 9, pp. 486-494, 2002