

## 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

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）に対して、（１）「制御系の学習機能」、（２）「制御系の状況認識と定量的表現」および（３）「自律エージェントの知識獲得」の３つの分担研究を行っている。平成12年度は、それぞれの分担研究によってつぎのような研究成果が得られた。

### （１）制御系の学習機能に関する研究

昨年度における本分担研究では、自律移動ロボットに対する学習システムとして、「強化学習」法のシミュレーション研究を行い、「強化学習」は周辺環境の変動がない静的環境下における学習法として有効であることが分かった。そこで本年度は、強化学習、特にQ-learningが環境変動に対してどのような学習対応を示すかについてのシミュレーション研究を行うこととした。

昨年度と同一の行動領域に対して、２つの障害物が１エピソード終了毎に１マス移動することとし、①スタート点とゴール点を結ぶ線上の往復移動、②スタート点とゴール点を結ぶ線と平行に往復移動、③環境１と環境２の同時変化、④スタート点からゴール点への通路を遮るように巡回移動、⑤スタート点とゴール点を結ぶ線と直角方向に往復移動、の５つの移動方法で考察した。その結果、動的環境に対する行動学習法としてQ-learningを用いる場合、移動障害物の経路上に障害物が定常的に存在すると見なして、適正な行動戦略を学習生成することがわかった。すなわち、定常的と見なされた移動障害物位置によって目標地点への経路が寸断される場合には、経路探索に失敗することがわかった。

### （２）制御系の状況認識と定量的表現

ビークルの走行制御については、昨年度までに16あるいは32ビットSHシリーズ制御用CPUボードによる制御を試みたが、本年度はDA変換によるDCモータの台形型走行パターンによる、こまやかな走行制御についての研究を行った。

一方、状況判断については、各目的を持った複数台の各移動ロボット（ビークル）からのロケーション情報を無線によってサーバに送り、そのサーバの情報を各クライアントとしてのビークルが共有する無線LAN走行システムを構築している。その障害物検知情報について、昨年度は人間検知（温度差）センサを移動するビークルに搭載したために、対象物をすべて人として判断してしまった。そこで本年度は、物を検知する一般的な距離センサ

を搭載し、ビークルが障害物を検知して停止した後の一定時間後に再走査を行なわせ、その障害物が移動物体であれば、再走査時には検知信号が変化するため、移動体であるかどうかを検出することができる。また、停止後に人間検知センサを作動させれば、人間と物体の識別が可能となり、走行制御をシンプルにすることができるようになった。もう一つの試みとして、音を制御系に利用できないかという観点から、原理的な検討と騒音計、FFTによる基礎的実験も行っている。

### (3) 自律エージェントの知識獲得に関する研究

本分担研究では、車の自動運転の第一歩として、障害物を避けながら前進して後退で車庫に入れる車庫入れ問題と前進した後に後退で縦列駐車する縦列駐車問題に対する運転操作を自動的に生成する方法を検討した。すなわち、昨年度開発した複数の遺伝的操作を並列的かつ適応的に施す進化計算法を適用して運転操作パターンの生成を試みた。

生成された操作パターンの良否を、①運転操作終了時点における車体位置と理想駐車位置の差、②障害物と車体との接触回数、③舵角変化の絶対値の累積値、の3つの評価項目で計り、各項目の重み付き合計値の逆数を個体の適合度としてシミュレーションを行った。その結果、車庫入れ問題では、逆位、突然変異、交叉の順で遺伝的操作の選択される確率が高くなることがわかり、約120世代で良好な解を生成した。また、縦列駐車問題では、遺伝的操作は突然変異、逆位、交叉の順で選択確率が高くなる場合が多いことがわかり、約150世代で良好な解を生成した。

## 制御系の学習機能に関する研究

竹口知男（工学部）

強化学習法は、設計者がロボットの動作・行動結果に対して報酬や罰（強化信号）を設定するのみで、ロボットが自律的に行動学習を行い、不確実性環境下での行動学習を取り扱うことが可能であるといわれている。そのため本分担研究では、自律移動ロボットに対する学習システムとしての強化学習法の有効性について、シミュレーション研究を行っており、昨年度は環境変化のない静的な環境下において強化学習法が有効であることを検証した。そこで本年度は、強化学習、特にQ-learningが環境変動に対してどのような学習対応を示すかについて考察することとした。

動作環境としては、昨年度と同様、Fig. 1 に示すような14×14マス目の二次元平面空間を移動ロボットの行動領域とし、任意に設定したスタート点から出発して、途中ロボットが通過できない壁や障害物を避け、最短ステップ数でゴール点に向かわせる最短経路探索問題を設定した。ただし、移動ロボットには事前に障害物などの環境に関する情報は一切与えず、強化信号としては、ゴール点到着時にそれまでの試行よりもステップ数が1ステップ減るごとに+1の報酬（Reward）を与え、壁や障害物に衝突するたびに-3の罰（Penalty）を与えるようにした。また、移動ロボットは、隣接する8方向の内の一方向へある確率で

1ステップずつ移動し、壁や障害物と衝突した場合には元の座標に戻る。8方向それぞれの移動確率は、それまでの学習およびその学習試行で与えられた強化信号によってリアルタイムに変化させるようにした。

環境変動としての障害物の移動は、1エピソード終了毎に2つの障害物が1マス移動するものとし、移動のさせ方はFig. 1 (a)~(e)に矢印で示すつぎの5つの環境で行った。

環境1：2つの障害物がスタート点とゴール点を結ぶ線上を往復移動する。

環境2：2つの障害物がスタート点とゴール点を結ぶ線と平行に往復移動する。

環境3：環境1と環境2の環境変化が同時に起こる。

環境4：2つの障害物がスタートからゴール点への通路を遮るように巡回移動する。

環境5：2つの障害物がスタート点とゴール点を結ぶ線と直角方向に往復移動する。

環境1~5に対して学習をさせた結果、エージェント（移動ロボット）が獲得したそれぞれの環境での最短経路をFig. 1内の菱形で示し、そのときの学習率、割引率、ステップ数およびエピソード数をTable 1に示す。環境4においては学習が成立しなかった。環境1~3および5においては、移動障害物の経路には定常的に移動障害物が存在するものとして、スタート点からゴール点までの最短経路を探索しており、特に環境5では、定常的に移動障害物が存在するものと見なした上で、障害物間の間隙通過路を探索しているという興味深い結果をもたらしている。しかし、環境4のようにタイミングを見計らって通過すれば最短経路を探索できるような動的環境に対してまでは、Q-learningでは対応できないことがわかった。

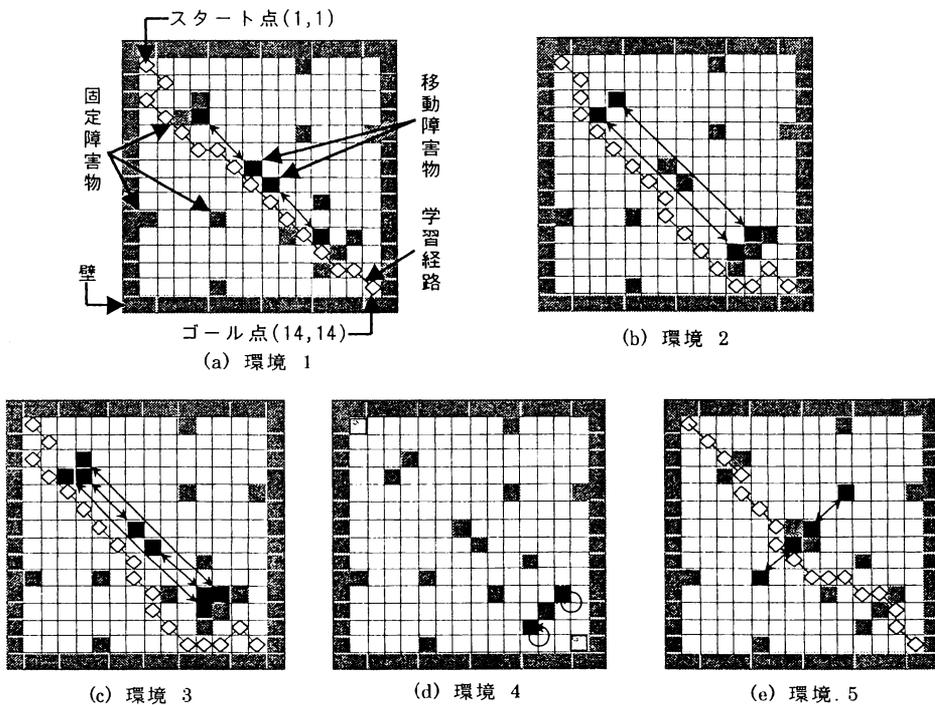


Fig. 1 障害物移動とQ-learningによる学習移動経路

Table 1 各環境における学習結果

	学習条件		学習結果	
	学習率 $\alpha$	割引率 $\gamma$	最小ステップ数	エピソード数
環境 1	0.6	0.9	15	59
環境 2	0.1	0.1	16	45
環境 3	0.8	0.8	17	68
環境 4	—	—	—	—
環境 5	0.3	0.8	16	53

## 制御系の状況認識と定量的表現

岩佐哲夫 宮井義裕 (工学部)

前年度までの研究成果はシステム・制御・情報学会、自動車技術会、ビークル研究会等の講演会において発表し、まず、ビークルの走行制御については一応の結論を得たと考えている。また制御用CPUボードも8ビットから16あるいは32ビットSHシリーズによる制御を試みた。特に本年度はDA変換によるDCモータの制御ソフトの試作によって、スタート時の増速および停止時の減速、すなわち台形型走行パターンによる、こまやかな走行の制御も行った。ビークルの試作はこの辺で一段落とし、状況判断の対象について、もう少し検討を行うことにした。本分担テーマで考えている走行ルートはマトリックス状に設置した経路上を複数台の移動ロボット(ビークル)が目的を持って走行するものである。その状況の定量的把握は、各経路上のロケーション、他ビークルの位置、進行方向等はクライアントを兼ねるビークルから無線によりサーバと送受信され、各クライアントとしてのビークルは、このサーバの情報を共有し、無線LANによる走行システムを構築している。問題は前年度の報告でも課題となった障害物の判断である。その障害物が何であるかではなく、人か物のいずれかを判断することである。前回では人間検知(温度差)センサを移動するビークルに搭載したために、対象物をすべて人として判断する失敗をした。そのため、このセンサの搭載をあきらめ、人・物を物として検知する一般的な距離センサを搭載し、もしルート上の障害物を検知してビークルが停止すると、一定時間後の走査を行ったとき、それが人あるいは移動中の障害物であれば、ビークルの前方から障害物が無くなっている、すなわち検知されない。これでは芸がないという訳ではないが、人検知センサを何とか活用出来ないだろうかという事にこだわってみた。その結果、ビークルが停止するまでは前述の通りで、停止後にこのセンサを作動させれば、移動する人である事の検知は可能で、走行制御をシンプルにすることも出来るようになった。もう一つの試みとして、音を制御系に利用出来ないだろうかという、学生の提案もあって、非常に原理的な検討と騒音計、FFTによる基礎的実験も行った。しかし、実際の使用にいたる大前提として、ビークルおよびその周辺の音量が一律でないため、特殊な状況、例えば特定の周波数帯を検知する場合には有用となるが一般的(日常の生活音)では使用出来ないと思われる。

本論ではないが、耳の不自由な人に後方の音量を増幅し、バイブレータによる認識に少々興味を感じた。これからの集大成として、近年、非常にインテリジェントな付加機器が搭載可能な模型が市販されるようになったので、共同研究者の夫々の課題をシミュレーションあるいは実験的に具体的なものとして要約しようと云うことになり、13年度の課題としている。

## 自律エージェントの知識獲得に関する研究 能勢和夫（工学部）

3年目は、車の自動運転の第一歩として、車庫入れや縦列駐車のための運転操作を自動的に生成する方法を検討した。すなわち、昨年度開発した複数の遺伝的操作を並列的かつ適応的に施す進化計算法を適用して運転操作パターンの生成を試みた。

車両の運動モデルにはアッカーマンステアリングジオメトリを用いた。すなわち、実舵角  $\delta$  で極めて低い一定速度  $V$  で走行している車の運動軌跡は、時間刻み幅  $\Delta t$  毎に繰り返して (1) 式の計算を行って求めることが出来る。

$$\delta = \frac{l}{R} \tag{1}$$

ただし、 $\delta$  は実舵角、 $l$  はホイールベース、 $R$  は旋回半径である。

時間刻み幅  $\Delta t$  毎の実舵角の変更量と進行方向の組をパラメータとし、パラメータの並びで操作の時系列パターンを表現した。舵角変更量は  $\Delta\delta \in \{\theta, 0, -\theta\}$  とし、舵角に対する上下制限約  $\delta \in [-D, D]$  を設けた。また、進行方向は  $\rho \in \{\text{forward, stop, backward}\}$  とした。これらを用いて、表1に示すパラメータの並びで、運転操作パターンを構成した。そして、これを進化計算での個体の染色体とした。

表1 パラメータの並び

操作No.	0	1	2	...	$N$
$\Delta\delta$	$\Delta\delta_0$	$\Delta\delta_1$	$\Delta\delta_2$	...	$\Delta\delta_N$
$\rho$	$\rho_0$	$\rho_1$	$\rho_2$	...	$\rho_N$

生成された操作パターンの良否を次の評価項目で計り、各項目の重み付き合計値の逆数を個体の適合度とした。

- ① 運転操作終了時点における車体位置と理想駐車位置の差
- ② 障害物と車体との接触回数
- ③ 舵角変化の絶対値の累積値

遺伝的操作は、交叉 ( $C$ )、突然変異 ( $M$ )、および逆位 ( $I$ ) の各操作を並列的、適応的に施すものとし、その際、適合度向上に効果的な操作を高い確率で選択した。そのために、各遺伝的操作毎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 (2) ~ (4) 式で求める。

$$f_r^j = \frac{O_r^j}{P_r^j} \quad r = C, M, I \quad (2)$$

ここに、 $O_r^j$  は操作  $r$  を施して生成した子集団の平均適合度、 $P_r^j$  はその子集団に対する親集団の平均適合度である。この適合度変化のトレンドを指数平滑式

$$\bar{f}_r^j = \lambda f_r^j + (1 - \lambda) \bar{f}_r^{j-1} \quad \bar{f}_r^0 = 1 \quad (0 < \lambda < 1) \quad (3)$$

で算出する。そして

$$S_r^j = \frac{\bar{f}_r^j}{\bar{f}_C^j + \bar{f}_M^j + \bar{f}_I^j} \quad S_r^0 = \frac{1}{3} \quad (4)$$

を  $j \sim j+1$  世代間における操作  $r$  の選択確率とした。

交叉によって必ず親個体と異なる子個体が生成されるようにした。すなわち、選ばれた親個体の 2 つの染色体に、対応する遺伝子の値が 2 ヶ所以上で異なっている場合のみ交叉を施すことにした。なお、突然変異、逆位は通常の処理方法とした。

世代交代の操作としては、子個体群は全て次世代に残し、親個体群の中で適合度が高いものから順に次世代に残す方法を用いた。

表 2 に示すパラメータを用いて車庫入れと縦列駐車 of 運転操作パターンを求めた。

表 2 パラメータ設定

	車庫入れ	縦列駐車
染色体長	65	110
親個体数	200	200
子個体数	120	120
世代数	250	250
$V(cm/sec)$	50	20
$\Delta t(sec)$	5	4
$\theta(deg)$	5	5
$\lambda$	0.05	0.05

図1は後進での車庫入れ問題の結果である。右の初期位置から出発して前進で障害物を避け、切り返しを行い、後進で車庫入れを行っていることが分かる。

図2は縦列駐車問題の結果である。右の初期位置から出発して前進で駐車位置を通り過ぎた後、ハンドルを切り返し、後進で縦列駐車を行っている。縦列駐車では細かなハンドル操作が必要である為  $V, \Delta t$  を小さく取った。これは操作パターンが車庫入れの場合と比べ短いステップ数で変化する為である。

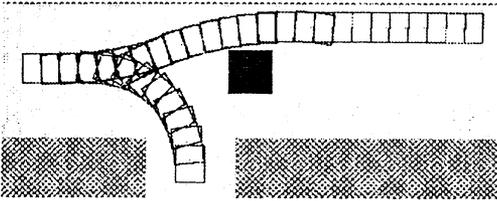


図1 車庫入れ

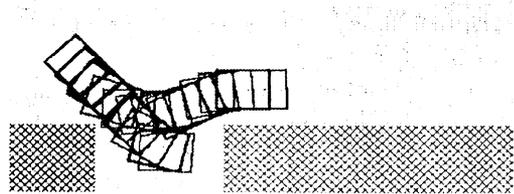


図2 縦列駐車

図3に車庫入れにおける操作の選択確率の推移を示した。遺伝的操作は、逆位、突然変異、交叉の順で選択される確率が高い事が分かる。同図より約150世代以降に交叉はほとんど施されていないことが分かる。図4は図3のときの最大適合度と平均適合度の推移を示したものである。同図より約120世代で良好な解が見つかったことが分かる。

なお、縦列駐車問題においては、遺伝的操作は突然変異、逆位、交叉の順で選択確率が高くなる場合が多く、また、約150世代で良好な解が見つかった。

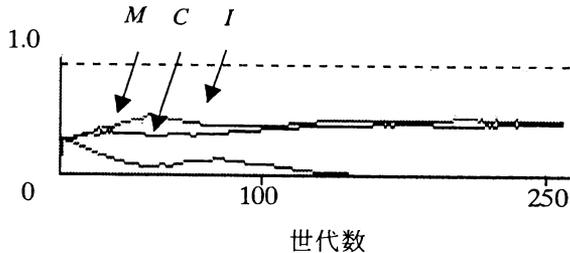


図3 操作の選択確率の推移

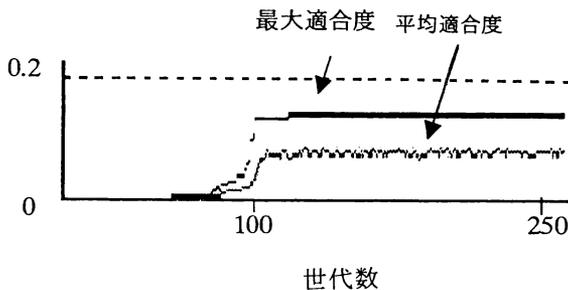


図4 最大適合度と平均適合度の推移

本研究では遺伝的操作を並列的、適応的に施す進化的計算法を、車の運転操作パターンの生成問題に適用した。今後は、トレーラトラックの運転や複数台車の協調運転の自動生成について研究していく予定である。