

行動に基づく情報獲得に向けた 形態と制御系の同時設計

杉浦 孔明^{†‡}, 赤羽 真[†], 塩瀬 隆之^{†‡}, 下原 勝憲^{†‡}, 片井 修[†]

[†] 京都大学大学院 情報学研究科
[‡] ATR ネットワーク情報学研究所

Autonomous Design of Morphology and Controller towards Behavior-Based Information Acquisition

Komei Sugiura^{†‡}, Makoto Akahane[†], Takayuki Shiose^{†‡},
Katsunori Shimohara^{†‡}, and Osamu Katai[†]

[†] Graduate School of Informatics, Kyoto University
[‡] ATR Network Informatics Laboratories

Abstract: This paper describes a system which automatically designs sensory morphology of an autonomous robot. In this system, a Genetic Algorithm is used to evolve sensor layout that is appropriate for learning line following task. We made the system to design a line following robot, and then constructed physical one. We have carried out experiments on how the robot learns to solve the task. The results have shown that the robot outperforms man-made robots on average lap time and robustness against changes.

Keywords: sensor evolution, ecological balance, embodiment, learning and evolution

1. はじめに

行動学習に有利な形状を持つロボットは、どのように設計すればよいのであろうか？

ロボットの行動学習に関する多くの研究において、学習器はセンサ空間とモータ空間のマッピングを自律的に行なうものとされている。実環境で行動するロボットは、センサおよびアクチュエータ（モータ）を用いて環境と相互作用するため、センサの位置が変化しただけであっても、信号が表す物理的变化は異なるものとなる。その結果、ロボットの知覚世界が変化するだけでなく、学習すべきマッピングも変化するといえる。

一方、アクチュエータやセンサを含めロボットの形態は設計者により与えられ、固定されていることが多い。形態の設計と、ロボットの制御系の設計とは、別々の観点から行なわれてきたためである。例えば、センサ形態の設計においては、設計者が物理量を計算しやすいような設計が重視されてきた反面、制御器が用いる学習則との関連は考慮されることが少なかった。

しかしながら、環境の変化に適応的なロボットを構築するためには、形態・制御系・環境のバランスを考慮しなければならない [3]。Pfeifer はこの設計原理を「生態学的バランス」と呼んだ。近年、形態と制御系をうまく組み合わせ、ロボットに適応的な振る舞い

を獲得させる研究が注目されている [5]。

このような考えに基づき、Sims はエージェントの形態と制御系をソフトウェア上で進化させることに成功している [4]。Lipson らは、この手法と迅速成形技術を組み合わせることによって、ソフトウェア上で得られた形態を、ハードウェアとして実現した [2]。しかし、これらの研究では個体発生的なスパンにおける適応、すなわち学習にとって有利な形態を自動設計させたものではない。一方、Jung は学習の結果を用いて、系統発生的なスパンにおいて状態空間の再配置を行なっているものの、形態の設計システムとして実現したわけではない [1]。

そこで本研究では、ロボットの形態のうちセンサに着目することによって、ロボットの行動学習に有利な形態を自動設計するシステムを構築する。すなわち、本システムは、ロボットの学習能力に適した状態空間を構築するために、学習結果に基づいてセンサ形態を自律的に設計する。センサ形態に着目する利点は、1) 「物理世界」と「情報世界」のインタフェースであるセンサをボトムアップ的に構築できる、2) アクチュエータを固定するため、ハードウェア上で実現しやすい、という2点である。

以下では、本システムを用いた設計実験および得られた設計解、加えて実機による検証実験について述べる。

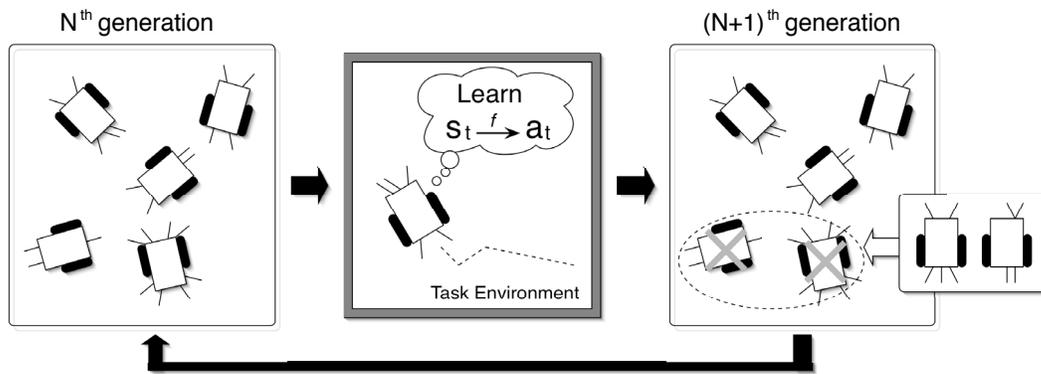


Figure 1: 設計

2. センサ形態設計システム

2.1 「学習しやすい」センサ形態の自動設計

行動学習に有利なロボット形態を設計するためには、実機を用いて実験と調整を繰り返してもよい。しかしながら、このような手法では、1) コスト (時間、回路の再設計など) が大きい、2) 設計にバイアスが入ってしまう、という二つの問題がある。

これに対し、本研究ではシミュレータを用いて実環境を模擬し、ソフトウェア上でロボットの形態を自動設計させる。このようにして得られた形態を設計指針として、ハードウェアを組み立てることにより、上に挙げた二つの問題を解決する。本研究で構築した設計システムは、ソフトウェア上でロボットの形態を書き換えることにより、様々な形態のロボットに適用可能である。なお、本システムでは、シミュレータとして Cyberbotics 社製の Webots¹を使用した。

本設計システムでは、学習に有利な形態を自動設計するために遺伝的アルゴリズム (GA) を用いる。各 GA 個体には、センサ形態がコーディングされ、学習結果に応じて評価値が与えられる。設計のおおまかな流れを以下に示す (Fig. 1 参照)。

1. 初期世代をランダムに生成する
2. 行動学習ループ
 - (a) 各試行ごとに報酬に基づき学習を行なわせる
 - (b) 一定の試行回数終了後、結果に応じて評価値を与える
3. 遺伝的操作を加えて次世代を生成する
4. 終了条件が満たされるまで 2. と 3. を繰り返す

¹<http://www.cyberbotics.com>

ただし、個体発生的時間スケールで行なわれる学習に関しては、世代間で学習結果の継承は行なわない。すなわち、ラマルクの進化は用いない。

2.2 タスク環境

本研究では、設計システムの適用対象としてライントレーサを採用する。ライントレーサとは、床に引かれたラインに沿って移動しゴールを目指すロボットのことであり、用いるコースを Fig. 2 に示す。コースは 2m×5m の長方形で、4 分の 1 ずつの地点にチェックポイント p_1, p_2, p_3 を設けてある。

タスクとしてライントレーサを用いる理由として以下の 2 点が挙げられる。

- 人手による設計との比較が可能
ライントレーサは多くの大会が開かれ、かつ多くの専用キットが販売されているため、人手によるセンサ形態のサンプルが多く存在している。そのため、本研究の設計システムによるセンサ形態との比較が可能である。
- センサ配置の変化が情報の意味を変化させる
例えば、2 個の床センサをラインに対して平行に配置する場合と、垂直に配置する場合では、ロボットが得る情報の意味は大きく異なる。そのため、より高いパフォーマンスをあげるには、センサ配置を効果的に配置する必要がある。一方、大会で用いられるロボットの多くでは、センサ配置が単純かつ規則的である。

2.3 設計対象

設計対象として、Fig. 3 に示すようなロボットを選んだ。このロボットは、マイコンカーラーにおいて標準的に用いられている文化総研社²のマイコンカー

²<http://www.bun-net.co.jp>

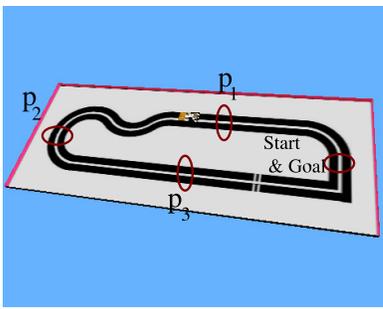


Figure 2: タスク環境

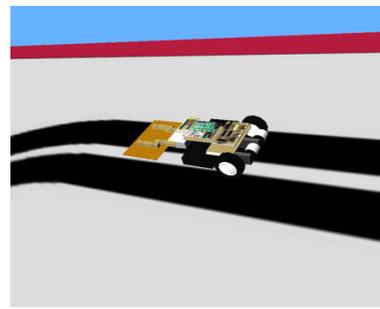


Figure 3: ライントレーサ

を元にしたものである。ただしこのロボットには、シミュレーションで扱いやすいように、2輪型へと改造を加えている。

ロボットは床センサを用いてラインを検出する。床センサは赤外発光 LED と受光素子から構成され、床面の色に応じた値をコントローラに渡す。本研究では、床センサの配置および個数に対して 2.1 節で述べた設計システムを適用し、自律的にセンサ形態を獲得させる。本来、センサの位置は連続値であるが、今回は簡単のため、床センサを配置する位置の候補を 32 個用意し、32bit の遺伝子型としてコード化した。これにより、センサの配置と個数を変更可能にしている。

2.4 行動学習

個体発生的時間スケールにおいて、ロボットは Q-learning を用いて行動を学習する。ただし、ロボットが観測できる情報はセンサ値のみとした。すなわち、絶対座標などの情報は与えていないため、ロボット自身の視点から得られる情報を用いて行動を学習する必要がある。また、ロボットの行動として、直進・左折(大)・左折(小)・右折(大)・右折(小)の 5 種類を用意した。なお、行動選択には ϵ -greedy 戦略を採用した。

ロボットには、100 回の試行機会 (エピソード) を与え、エピソードごとに位置および向きを初期化して最大 T_{max} ステップまで行動を行なわせた。ロボットは、各ステップごとにラインの中心からの近さに応じて報酬を得る。学習の結果、ゴールへ最も早く到達できるようになった個体ほど、評価が高くなるように適応度を与えた。

3. システムの適用実験

3.1 実験設定

本節では、提案するシステムを用いて行なった 2 種類の実験について述べる。実験の概要は以下の通りである。

- 実験 A : 本システムを用いたセンサ形態の設計
前節で説明したシステムを用いて、ラインレースの学習に有利なセンサ形態を自動設計させる。
- 実験 B : 人手によるセンサ形態との比較
実験 A において得られたセンサ形態を固定し、人手によるセンサ形態と比較する。すなわち、両者のセンサ形態に対して強化学習による学習結果を調べる。

個体の適応度は、各試行でのゴールまでに要したステップ数から算出される。すなわち、適応度 ϕ は以下に示す ϕ の平均とした。

$$\phi = 1 - \frac{t}{T_{max}}$$

ここで、 t はそのエピソードでゴールまでに要したステップ数、 T_{max} は最大行動ステップ数 (2000 回) である。ただし、ゴールできなかった場合は $\phi = 0$ とし、チェックポイントまで到達した場合には、 ϕ に補正を加えた。

3.2 設計実験の結果 (A)

まず、本研究で提案したシステムの設計解を定量的に検討した。Fig. 4 は、各世代における適応度と、その世代の最良個体のセンサ個数の変化を示したものである。ただし図において、実線は最大適応度を、破線は平均適応度を、点線はセンサ個数をそれぞれ表している。実験を 10 回行なって、その平均値をプロットした。

Fig. 4 より、最良個体においては、センサ個数が減少するに従って適応度が増加していることがわかる。例えば、第 5 世代 (センサ数 7 個前後) の適応度が約 0.2 であるのに対し、50 世代目 (センサ数 5 個前後) の適応度は約 0.4 に上昇している。この結果から、ラインレース環境においては、5 個前後のセンサを有していることが学習に有利であるといえる。これは、学

習器を用いることによって、マイコンカーラーで標準的に用いられるセンサ個数(8個)を、減らすことができることを示唆している。

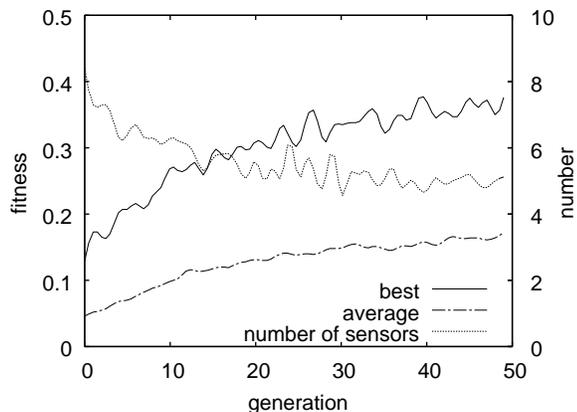


Figure 4: 最大適応度, 平均適応度, 最良個体のセンサ個数の変化 (10 回の実験の平均)

本システムによって設計されたセンサ形態のうち, 代表的なものを Fig. 5 に示す. ただし, Fig. 5 には, センサの数が 4, 5, 6 個の 3 通りを示した.

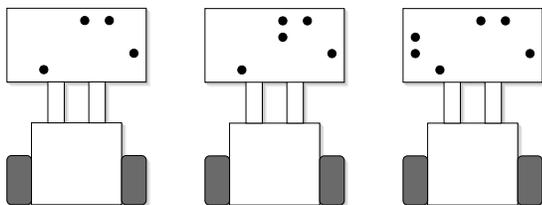


Figure 5: 設計された形態の例: 4 個 (右), 5 個 (中), 6 個 (左)

3.3 人手によるセンサ形態との比較結果 (B)

実験 A において得られたセンサ形態と, 人手によるセンサ形態を用いて学習結果の比較を行なった. 比較対象のセンサ形態を Fig. 6 に示す. Fig. 6 左図がマイコンカーのデフォルトのセンサ形態 (M1) であり, 右図が本システムの設計解 (M3) である. この 2 種類に加え, 設計解と状態空間の次元が等しいマイコンカー M2 (Fig. 6 中図) を用意し, 3 者で比較を行なった.

Fig. 7 にエピソードに対する ϕ の変化, すなわちゴールに到達する早さの変化を示す. 図において破線・点線・実線は, それぞれ M1・M2・M3 の結果を表す.

Fig. 7 より, 本システムが設計した M3 が, 全てのエピソードにおいて高い値を示していることがわかる. このことを確認するために, エピソード数を増やして同様の実験を行なったところ, Fig. 7 と同じ傾向が確

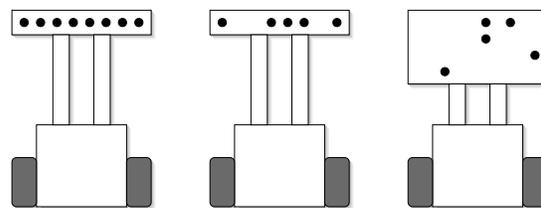


Figure 6: 左図: デフォルト配置のマイコンカー M1, 中図: マイコンカー M2, 右図: 本手法で設計されたマイコンカー M3

認された. また, 状態空間が大きい M1 に対しても, 収束速度とパフォーマンスの面で M3 が勝っていることから, 物理世界を適切に観測することによって, より効果的な学習を行なえることが示唆される.

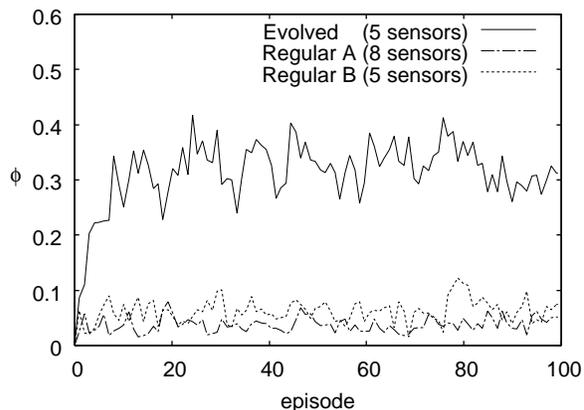


Figure 7: 学習結果の比較

3.4 考察

本システムによって獲得されたセンサ形態に関して, 配置と個数の点から考察を行なう.

センサの配置 Fig. 5 に示すように, 獲得されたセンサ形態は以下のような特徴を持つ.

- 配置が左右非対称である

今回のタスクは, コースを半時計回りするようにしたため, S 字カーブを除けば左カーブが多い. そのため左カーブを得意とするセンサ形態の適応度が高くなったと考えられる.

- 前後に分散して配置している

横一列に並んだ配置は, ラインの前後関係を読み取ることができない. そのため, ロボットがカーブの上にいるのか, あるいは直線の上にいるのか, をロボット自身が判断することが難しい. 前後に

センサが分散することで、ラインの前後関係から直線やカーブといったラインの形状をロボットが読み取ることが出来るようになったと考えられる。

センサの個数 自動設計されたセンサ形態は、5個前後のセンサを持つものが多い。センサの個数は状態空間の次元を決定するため、学習速度と取得できる情報量に影響を与える。一般的に、次元が少ない方が学習の収束が早いですが、学習収束後のパフォーマンスは次元が多い場合よりも低くなると考えられる。しかし今回の実験では、学習が収束した後であっても、センサ数が少ないロボットのパフォーマンスが高い、という結果が得られた。これは、少ない状態空間を効果的に利用すれば、収束速度だけでなく、学習結果も改善できることを示している。

以上のことから、本手法を用いることにより、タスク環境とロボットの学習能力に適した状態空間を構成させることができ、より適応的な行動が可能になることを示せた。

4. 実機による検証実験

4.1 設定

本システムにより得られた設計解を実機に実装して、既存手法との比較実験を行なう。本実験では、3種類のロボット M1, M2, M3 に対してライントレースを行なわせる。これらは、Fig. 6の左、中、右のロボットにそれぞれ対応する。用いたロボットの比較表を Table 1 にまとめた。

Table 1: ロボットのセンサ形態とコントローラ

マイコンカー	センサ形態	コントローラ
M1 (既存手法)	デフォルト	ハンドコーディング
M2 (既存手法)	デフォルト	シミュレーションの学習結果
M3 (本手法)	本システムによる形態	シミュレーションの学習結果

M1 のセンサ形態は、マイコンカーのデフォルトのものを用いた。また、M1 のコントローラは、付属のサンプルプログラムを改良したものを用いた。M2 のセンサ形態は、M1 と同様にデフォルトのものを用いた。M2 のコントローラは、設計実験によって得られた Q 値に基づいて行動を決定する。すなわち、センサが観測した状態に対して、最大の Q 値を持つ行動を選択する。ただし、本実験では Q 値をオンライン

で更新することはない。M3 のセンサ形態は、本システムが自動設計したセンサ形態のうち、センサ数 5 個のものを実装した (Fig. 8 参照)。M3 のコントローラは、M2 と同様に設計実験で得られた Q 値を用いたものである。

各マイコンカーは 5 回試行することができ、1 回の試行はコースを 10 周するかコースアウトすると終了する。走行性能を検討するため、M1, M2, M3 の平均ラップタイムを比較する。また実機では、バッテリーの消耗のため、ラップを重ねるごとに同じコントローラでの走破が困難になるという特徴がある。そこで、マイコンカーのロバスト性を調べるために、平均滞在平均ラップ数についても比較を行なう。

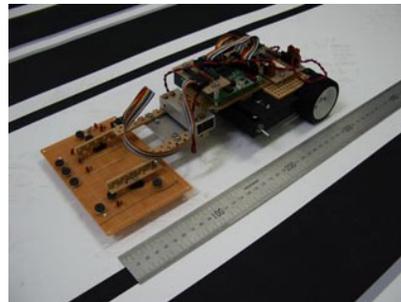


Figure 8: 製作したライントレース (M3)

4.2 実験結果

実験結果を Table 2 に示す。表より、平均ラップタイムを比較すると M3 が最も速いことがわかる。加えて平均滞在ラップ数は M3 が最も大きいことから、M3 は M1, M2 に比べてコースアウトの少ないロボットであるといえる。すなわち本システムは、実機においても、人手による設計より優れた設計を行なうことができると考えられる。

次に、状態空間が等しい M2 と M3 に対して、センサの配置の影響について検討する。M2 と M3 では同じ学習則を適用したにも関わらず、M3 のパフォーマンスが顕著に高い。このことから、身体形状が固定されたロボットであっても、センサ配置を変更することでパフォーマンスを改善できることを示唆している。

Table 2: 実機における既存手法との比較結果

マイコンカー	平均ラップタイム [秒]	平均滞在ラップ数
M1	15.6	4.4
M2	16.0	3.0
M3	13.5	8.4

4.3 考察

ライントレースにおいて難易度が高いとされる直角コーナーの攻略を例として、M1, M2, M3 がどのようにセンサ配置を利用しているかを述べる。

M1 は直角コーナーの直前に存在するクロスラインを合図に用いる方法を採用している。すなわち、直角コーナーの直前のクロスラインを読み取ると、直角コーナー突破用のモードに移行し、直角コーナー専用の制御則を用いて直角コーナーをクリアする。このように、人手によるコントローラでは「クロスラインの後には直角コーナーが存在する」というコースに関する設計者の知識を利用できるため、通常のカーブとは異なる制御則を使って直角コーナーを突破できる。

しかし、M3 はコースに関する設計者の知識を利用できない。そのため、直角コーナーを通常のカーブと同じ制御則を用いてクリアしなければならない。M3 では、直角コーナーにさしかかると Fig. 9 左図に示すようにセンサが反応する。これは左カーブと同じセンサ状態であるので、M3 は少し左に曲がる。ところが、実際にはコースは直角であるため、Fig. 9 右図に示すようにコースから外れそうになる。ここで4番目のセンサがセンターラインを検出し、マイコンカーは大きく左折するため、直角コーナーをクリアすることができる。

一方、横一列に並んだセンサを持つM2は、直角コーナーでコースアウトをすることが多い。これは、M2のセンサが前後に分散していないため、コース上の前後関係を学習できなかつたことによると考えられる。

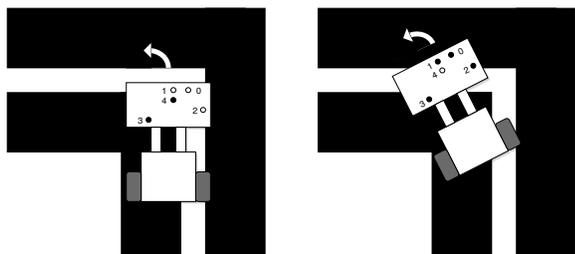


Figure 9: 直角コーナーにおける形態の利用

以上の定性的な比較と前節の定量的な比較から、本システムで設計されたロボットが、既存手法に比べて学習性能に優れることを、実機においても実証することができた。

5. おわりに

本研究では、ロボットの学習能力に応じた情報獲得を行なわせるために、行動学習に基づいてロボットのエ

ンサ形態を自動設計するシステムを構築した。本システムをライントレーサに対して適用し、得られた形態と人手による形態の学習結果の比較を行なった。この実験結果から、本手法で設計されたによるロボットは、物理世界を適切に観測することにより、人手による設計に比べて優れた学習能力を獲得できることが示せた。すなわち、本システムによって、環境・形態・制御系の相互作用を効果的に利用できることがわかった。さらに、本システムによって得られたセンサ形態を実機に実装して、既存手法と比較した。その結果、本手法で設計されたロボットは、平均ラップタイムおよび頑健性の点で優れた行動を行なうことを示した。

謝辞

本研究は独立行政法人情報通信研究機構の研究委託「人間情報コミュニケーションの研究開発」により実施したものである。

参考文献

- [1] Jung, T., Dauscher, P. and Uthmann, T.: Evolution and Learning: Evolving Sensors in a Simple MDP Environment, *Adaptive Behavior*, Vol. 11, No. 3, pp. 159–177 (2003).
- [2] Lipson, H. and Pollack, J. B.: Automatic design and manufacture of robotic lifeforms, *Nature*, Vol. 406, No. 6799, pp. 974–978 (2000).
- [3] Pfeifer, R. and Scheier, C.: *Understanding Intelligence*, MIT Press, Cambridge, MA. (1999). (石黒章夫, 小林宏, 細田耕監訳: 知の創成 – 身体性認知科学への招待, 共立出版, (2001)).
- [4] Sims, K.: Evolving 3D Morphology and Behavior by Competition, *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (Brooks, R. and Maes, P.(eds.)), Cambridge, MA., MIT Press, pp. 28–39 (1994).
- [5] 細田耕: 形態が学習にもたらすもの, 学習が形態にもたらすもの, *日本ロボット学会誌*, Vol. 22, No. 2, pp. 186–189 (2004).