

フィードバック学習に基づく順序機械の自己組織化に関する研究

5AG-8

神谷 健志 岡部 洋一

東京大学 大学院 工学系研究科 超伝導工学専攻

kamiya@okabe.rcast.u-tokyo.ac.jp

1 はじめに

生物が実際の世界の中で行動を獲得する際に働いている原理として、フィードバック学習が考えられる。フィードバック学習とは、生物が試行錯誤的に行動を変化させていく際、その行動の変化が正しかったのか間違っていたのかなど、自分のとった行動に対する環境の応答を見て、その結果を学習に取り入れていく、といった学習法である。そして、その際に働く学習アルゴリズムとして、岡部らはモデレーショニズム [1] を提案した。モデレーショニズムとは簡単に言えば、「生体は入出力信号ともに、適度な（モデレートな）信号レベルを好む」という仮説である。現在までにモデレーショニズムを用いて、人工腕において反射弓が形成されることなどがシミュレーションやロボットアームを用いた実験において確認されている。本研究では、このモデレーショニズムをデジタルの論理回路に適用することを試みる。そして、モデレーショニズムをもとにした学習アルゴリズムによって、組み合わせ論理回路および順序回路の学習が可能であることを示すことを目指す。

2 本システムの構成

2.1 可変論理回路

自らの出力と環境からの応答によって、自己組織化を行う論理回路を可変論理回路と呼ぶことにする。フィードバック学習を実現するために、この回路を「単調な入力が続くと自らの入出力関係を変えて、多様な入力が来るようになるまで試行錯誤する」という特性を持ったニューロンによって構成する。この学習法においては、特別な教師信号や、グローバルな（外部から与えられた）評価などはいっさい用いず、ただそれぞれのニューロンの信号だけを見ることによって学習が進め

られる。

構造変化は、それぞれのニューロンの出力が単調であるときに、そのニューロンへの入力の結合荷重や閾値を変化させる、という形で行われる。この時の構造変化は基本的にはランダムに行われ、ニューロン i からニューロン j への結合をみて、下の表のように変更している。

x_i	x_j	構造変化
単調	単調	閾値の変更、結合荷重の変更
多様	単調	結合荷重を大きくする
単調	多様	閾値の変更、結合荷重の変更
多様	多様	OK

2.2 環境

今回の研究では、可変論理回路にある入出力関係を獲得させることを目標とするので、次のようなものが環境としてふさわしいと考えられる。すなわち、環境は、可変論理回路が（目標とする入出力関係と照らし合わせて）正しい出力を返してくるまで、同じ入力パターンを入力し続ける。そして、正しい出力を返している間は多様な入力を与えるのである。これによって、可変論理回路が誤った出力を出しているときは、単調な入力が続くことになり、構造変化を促すことができる。

2.3 評価関数

今回、シミュレーションに用いた評価関数は次の二つである。

評価関数 1

ある一定時間 N_{watch} の間のニューロンの出力の平均 m を計算し、これを用いて、

$$M = 1 - 4(m - 0.5)^2$$

という関数を考える。この関数は $m = 0.5$ で最大値 1 をとり、 $m = 0, 1$ で最小値 0 をとる。つまり、この関数は信号の平均が 0.5 からどれだけ離れているかを表す尺度となる。

評価関数 2

$$\begin{aligned} & \dots 0001101101\dots \\ P_{00} &= 1/5 & P_{10} &= 1/5 \\ P_{01} &= 2/5 & P_{11} &= 1/5 \end{aligned}$$

図 1: パターンの出現頻度

ある一定時間 N_{watch} のニューロンの出力をみる。この長さ N_{watch} の 0,1 からなる文字列の中に、長さ $L_{pattern}$ のパターンがいくつ現れるかを数える。長さ N_{watch} の文字列の中には長さ $L_{pattern}$ のパターンが ($N = N_{watch}/L_{pattern}$) 回登場するが、あるパターン i が n_i 回現れたとき、このパターンの出現頻度 p_i を、 $p_i = n_i/N$ と表す。(図 1) この時、

$$M = \sum_{i=1}^m -p_i \log p_i$$

という関数を考える。 M は、すべてのパターンが等確率で現れたときに最大値 $\log N$ となり、一つのパターンしか現れなかったときに最小値 0 となる関数である。つまり、この関数は、様々なパターンが現れるほど大きな値をとる関数であり、信号の多様さを表す一つの基準となりうる。

3 組み合わせ論理回路の学習

実際に 2 入力 OR の動作を可変論理機械に学習させることを試みた。環境は、(0,0),(0,1),(1,0),(1,1) の入力の組を可変論理回路に入力する。それぞれに対して正しい出力 (この場合 0,1,1,1) を返してきたら次のパターンを入力する。誤った出力を返してきた場合は正しい出力を返してくるまで同じ入力が続ける。このとき、評価関数 1、2 を用いた場合に差は見いだせなかった。

4 順序回路の学習

1 入力 1 出力のリカレントネットワーク (図 2) に対して、1[timestep] 前の入力を出力として返すような順序回路を学習させることを試みた。環境は、 $SIZE[timestep]$ の入力をパターンとして持っていて、順々に可変論理回路に入力する。すなわち、 $SIZE = 2$ の場合は、パターンとして 00,01,10,11 の 4 つを持っていて、正しい答えを返してきたときには次のパターンを

与えるが、誤った答えを返してきたときは同じパターンを与え続ける。(図 3) この場合、評価関数 1 を用いると簡単に学習が終わってしまうが、正しい入出力関係は学習されない。しかし、評価関数 2 ($L_{pattern} = 2$) を用いると正しい入出力関係が獲得されることがわかった。

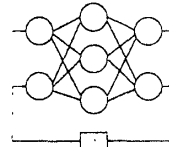


図 2: 1 入力 1 出力の可変順序機械

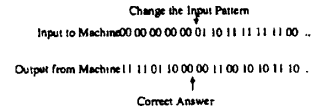


図 3: 順序機械への信号の与え方

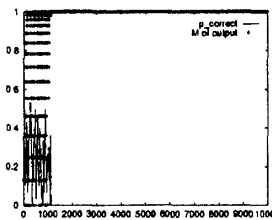


図 4: 正解率と出力の評価関数の推移 (評価関数 1 を用いた場合)

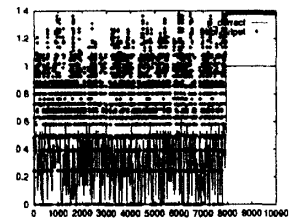


図 5: 正解率と出力の評価関数の推移 (評価関数 2 を用いた場合)

5 おわりに

評価関数を適当に与えることによって、順序回路においても単純なルールで学習が可能であることがわかった。とは言うものの、基本的に試行錯誤に基づく学習であるので成功率が高いとはいえない。成功率を向上することと、より複雑な回路への適用が今後の課題である。

参考文献

- [1] Okabe Yoichi "Moderationism: Feedback Learning of Neural Networks", Proc. of 1988 Intn'l Industrial Electronics Conference (IECON'88), IEEE, pp. 1028-1033(1988)
- [2] 酒井 千絵 "母子関係における対話をモデルにした可変順序機械の研究", 1995 年度国際基督教大学卒業論文 (1996)