

小鳥の歌文法を再現するニューラルネットワーク Neural network model generating symbol sequence for songs of Bengalese Finch

小谷 潤一郎† 森 康久仁† 松葉 育雄†
Junichirou Kotani Yasukuni Mori Ikuo Matsuba

1. まえがき

小鳥の歌の神経生物学は、1960年代にアメリカで創始され、現在では世界各国60を超える研究室で研究が行われるようになってきている。そして、近年の研究では、ジュウシマツという種の小鳥は他の種に比べ複雑な歌をうたい、その歌の形式は確率有限オートマトンによって記述できることが報告されている。さらに、ジュウシマツの歌中に見られる階層的な構造と、ジュウシマツの脳内のいくつかの神経核(神経細胞の密集部)には、はっきりとした対応関係が見られるという実験結果も報告されている[1]。

本研究では、このような小鳥の歌の神経生物学の研究結果に基づいて、簡単な人工ニューラルネットワークを作成し、そのネットワークを使って、ジュウシマツの歌に見られるような確率有限オートマトンで記述される文法に従った記号列の学習を行う。そして、学習後のネットワークを動作させることで記号列を発生させ、その結果が元々の確率有限オートマトンにしたがっているかどうかを検証する。

2. 小鳥の歌の構造と神経回路

小鳥の歌は、いくつかの種類の特定の周波数パターンを持った音が、無音区間で区切られた音節の形で順番に出現するという形式になっている。このような音節は「歌要素」と呼ばれている。歌要素はランダムな順序で発せられるのではなく、学習により得られた特定の順番に従って発せられる。

歌要素の発声順序は、多くの種では毎回同じ順序で発せられるステレオタイプなものとなっているが、ジュウシマツという種の歌の順序は、他の種と比べて非常に複雑な形になっている。図1にオートマトンで表わされたジュウシマツの歌の文法の一例を示す。図中では、歌中の異なる歌要素をアルファベットで表している。このオートマトンは、可変長 N グラムモデル[2]により実際のジュウシマツの歌をもとにして作られたものである。このようなオートマトンで表現された文法は、「有限状態文法」と呼ばれる。この図に見られるように、ジュウシマツの歌は複雑な形をしているものの、その中には決まった歌要素の並びが出現していることがわかる。このような特定の歌要素のまとまりは「チャンク」と呼ばれている。ジュウシマツの歌は、歌要素・チャンク・有限状態文法という階層構造をもっているといえる。

このようなジュウシマツの歌の構造は、脳内における神経核の階層構造と対応関係がみられることが知られている。小鳥の歌生成において主要な役割を果たしているとされる神経核には、RA、HVC、Nifがあり、RAは歌要素単位での生成を担い、HVCはチャンク単位での生成を

行い、Nifはジュウシマツの文法全体に固有の複雑性を与えていると考えられている。これらの事実は、各神経核の損傷実験によって確かめられている。

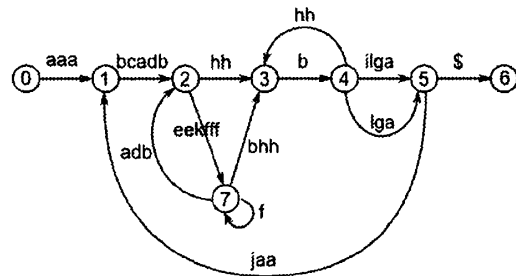


図1 ジュウシマツの歌文法

3. ネットワーク構成

本研究では、小鳥の神経回路を参考にした図2のような人工ニューラルネットワークで、確率有限オートマトンから生成される記号列の学習を試みる。モデル化は、ジュウシマツのNif神経核を損傷すると、歌文法が単純なものへと変化してしまうという実験事実をもとに行う。この事実より、基本的な歌生成はHVCでの状態変化により行われ、そこへNifからの信号が加わることにより、歌にバリエーションがでるとする仮定をとる。また、Nifからの信号は、前時刻までに発声された歌要素の履歴を反映したものであると仮定し、その履歴信号はNifでの自己フィードバックによって実現されるとする。ネットワークの実装では、HVC神経核に相当する部分に、代表的な人工ニューラルネットワークである多層パーセプトロン[3]を用い、時刻 t での入力ベクトル I_t と、時刻 $t-1$ までの履歴を反映した文脈ベクトル C_{t-1} から、次時刻の記号の確率分布を示す O_{t+1} を出力させる。文脈ベクトル C_{t-1} は、Nif神経核に相当する記憶層での重み d ($0 \leq d \leq 1$) の自己フィードバックにより実現され、時間的な減衰を伴った入力履歴の積算という形式となっている。

4. 実験方法

実験では、ジュウシマツの歌を可変長 N グラムモデルによって解析して得られた図1の確率有限オートマトンより記号列を生成し、その記号列を学習データとして用いる。オートマトン中の状態遷移の分岐は等確率で起こるとし、初期状態から状態遷移を繰り返して複数の記号列を生成する。記号の種類は13種なので、ネットワークへの入力の際には、各記号を13bitのビット列に変換し、1bitのみが1で他がすべて0となるようにする。このビット列が入力層、記憶層への入力ベクトルとなる。

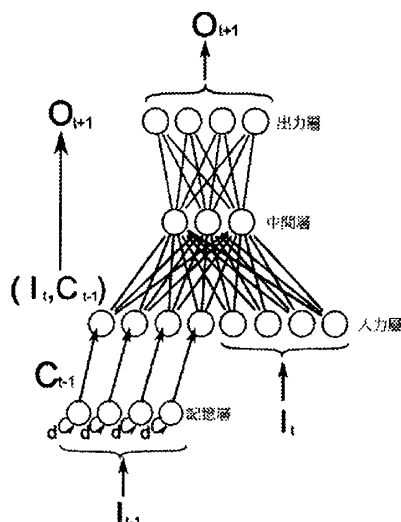


図2 ネットワーク構成

学習は多層パーセプトロン中の結合荷重をバックプロパゲーション法によって更新することで行った。多層パーセプトロンに入力される現時刻の記号を表すベクトル I_t と、これまでの入力履歴を反映した文脈ベクトル C_{t-1} の組から、次時刻の記号 I_{t+1} を出力させるように学習をさせる。ただし、元の小鳥の文法

中には分岐部が存在し、ひとつの入力に対して複数の出力候補があるため、分岐部においては学習を収束させることができない。よって、学習後のネットワークの出力は次時刻の記号そのものではなく、次時刻の記号の確率分布を反映した値となっているという風にとらえることにする。すなわち、出力層の各ニューロンの出力値に比例した確率分布をもとにして、各ビット値のどれか一つを確率的に1であると決定し、そのほかの値は0として次時刻の記号を表すベクトルを取得する。

学習後には、ネットワークの学習が成功したかを確認するために、ネットワークから記号列を発生させる。まず、歌文法の最初の記号である 'a' をネットワークに入力し、二番目の記号を出力から得る。そして得られた記号をフィードバックしてネットワークの入力に与え、次の三番目の記号を出力から得る。同様の手順を端末記号 \$ が出るまで繰り返すと、一つの記号列を得ることができる。そして、一連の手順を何度も行うことで、複数の記号列を得ることができる。その後、得られた記号列を使用して、元々の小鳥の歌に対して行われたのと同様に、可変長 N グラムモデルによってオートマトンを作成する。ここで作成されたオートマトンが学習データの生成元となったオートマトンと一致すれば、ネットワークは学習に成功したといえる。

また、以上の実験に加えて、オートマトン中の分岐部の分岐確率を変更した場合に、学習結果にどのような影響が出るかも調べた。この実験では、図1の状態2の分岐において、"hh"の遷移を行う確率を p 、"eekfff"の遷移を行う確率を $1-p$ として学習データを作成し、ネットワークがこの分岐確率の偏りを反映した学習を行うかどうかを確認した。

5. 結果

前節の実験を行った結果、図3のようなオートマトンを得ることができた。このオートマトンは図1のオートマトンと多少形が違うものの、受理する記号列は同じものとなっている。形が異なる原因は、可変長 N グラムモデルのパラメータの違いや、オートマトン生成の際の整形ア

ルゴリズムの違いによるものと考えられるが、本質的な問題ではない。

また、図1のオートマトンの状態2の分岐を変化させて学習を行った結果、学習後のネットワークが発する記号列の状態2に相当する部分での分岐確率は図4のようになった。図4では、横軸 p が学習データの"hh"への分岐確率を示し、縦軸 $q(p)$ が学習後のネットワークが生成する記号列の"hh"への分岐確率を示している。結果より、学習データ中の分岐確率の偏りもネットワークに反映されているといえる。

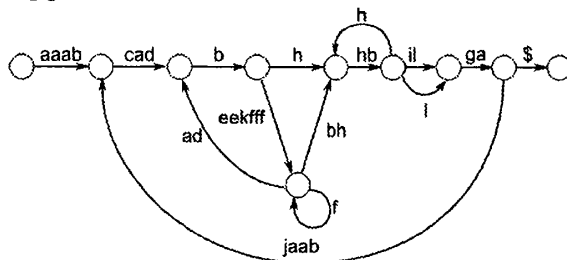


図3 ネットワークから生成された記号列の文法

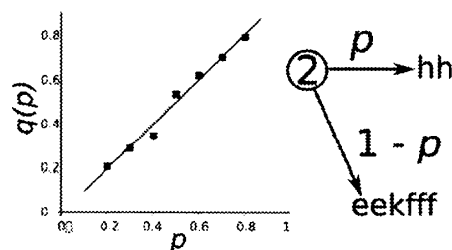


図4 ネットワークへの分岐確率の反映

6. まとめ

本研究では、小鳥の歌の神経生物学のこれまでの研究成果をもとに、簡単な人工ニューラルネットワークをモデル化し、そのネットワークによって小鳥の歌のような確率有限オートマトンで表現される文法に従った記号列の学習を行えるかどうかを検証した。その結果、現時刻の記号と短期的な記号履歴の組を多層パーセプトロンの入力とすることで、次時刻の記号の確率分布を得ることができ、元の学習データと同等の記号列の生成を可能とするネットワークを構成することができた。また、ネットワークが出力する確率分布は、学習データの分岐確率の偏りを反映したものとなった。

今後の課題としては、確率的な分岐を行うための機構をネットワーク内部に埋め込むことや、より生体に近い形でのネットワークのモデル化などが考えられる。

参考文献

- [1] 岡ノ谷一夫, "岩波科学ライブラリー92 小鳥の歌からヒトの言葉へ", 岩波書店, 2003.
- [2] 北研二, "言語と計算 4 確率的言語モデル", 東京大学出版会, 1999.
- [3] Simon Haykin, "Neural Networks: A Comprehensive Foundation", PRENTICE HALL, 1999.