リカレントニューラルネットワークにおける BPTT 法を用いたアトラクタの形成

高瀬 治彦*・郷原 一寿**・内川 嘉樹***

Creating Attractors Using BPTT Algorithm with Recurrent Neural Networks

Haruhiko TAKASE*, Kazutoshi GOHARA** and Yoshiki UCHIKAWA***

We have already demonstrated that, by regarding the time variable external input of finite length as one cycle of the wave form of a periodic function, RNN's can be described as forced vibration systems. It was also pointed out that to explain behaviors of RNNs we must consider attractors created in the state space of RNNs corresponding to training patterns. This paper describes our recent systematic simulations to investigate into behaviors of additive nets, typical models of RNNs. In the case where multiple combinations of simple wave forms of limited length are employed as time sequential training patterns. The results showed that (1) a cyclic trajectory is created in the state space of the RNN corresponding to each of training patterns, and (2) transition time from one cyclic trajectory to another induced by switching from one input pattern to another can be shortened remarkably by selecting appropriate combinations of training patterns.

Key Words: recurrent neural network, time sequential pattern transformation, attractor

1. はじめに

これまで提案されてきた、ニューラルネットワーク (NN) のモデルは、(1) 内部に時間を含むモデル、(2) 含まないモデルの二つに大別することができる。前者の代表的なモデルは、ホップフィールドモデルであり、後者のそれは、階層型ネットワークモデル (多層パーセプトロン) である。ホップフィールドモデルでは、外部入力を時間的に変化しない一定値とし、ユニット間の結合を対称に限定することにより、ポテンシャル関数が定義できる。その動作は、状態空間内のある一点から、漸近安定な平衡点に収束するまでの状態変化であると考えることができる。他方、階層型ネットワークモデルは、

フィードバック結合を持たず、信号の流れが入力から出力への一方方向である。その動作は入力から出力への写像ととらえることができる。従来、これら2種類のモデルは、全く異なるものとして、理解されてきた。

これらに対して、時間遅れを持つユニットが、任意に結合されているリカレントニューラルネットワーク (RNN) は、上述した二つのモデルを包含する一般的なモデルである。RNN の教師あり学習の定式化は、ここ数年来、相次いで行われてきた。これら定式化においては、RNN の動作を時系列入力信号から時系列出力信号への時系列パターン変換の観点からとらえたものがほとんどである。しかし、RNN は数学的には力学系 (非線型ダイナミカルシステム) であり、RNN の動作の理解には、状態変化の解析という観点を忘れてはならない。

RNN の学習の対象には、一般に時間軸上で有限長の入出力パターンが与えられる。この入出力変換ができるためには、学習する有限長の時系列入出力パターンを、無限に続く周期信号の一周分とみなし、学習パターンに対応して、状態空間内にアトラクタが形成される必要があることを報告した^{1)・2)}。

* 三重大学工学部 津市上浜町 1515

** 北海道大学工学部 札幌市北区北 13 条西 8 丁目

*** 名古屋大学工学部 名古屋千種区不老町

* Faculty of Engineering, Mie University, Tsu

** Faculty of Engineering, Hokkaido University, Sapporo

*** Faculty of Engineering, Nagoya University, Nagoya
(Received January 13, 1995)

(Revised January 10, 1996)

換言すれば、RNN による時系列パターン変換が、入出力パターンに対応したアトラクタ、およびそれに付随する引き込み領域（ベイスン）の位置・形状により決定されることになる。したがって、RNN の学習に関する問題は、以下の 2 点に集約される。

- (1) 複数ある学習パターンそれぞれに対応して、アトラクタを状態空間内に形成することが可能か。
- (2) 異なった入力パターンが時間的に連続して入力される場合、RNN の状態はそれぞれのパターンに対応したアトラクタ間を遷移することになるが、その遷移時間を学習により短くすることが可能か。

本論文では、これら 2 点の問題に対する検討を詳細な数値実験により行なう。学習法としては、理論的に学習の収束が保証されている BPTT (Back Propagation Through Time) 法を用いた。

本論文の構成を以下に示す。2 で学習パターンと状態空間のアトラクタとの関係について、文献 1), 2) の内容を簡潔にまとめる。3 で数値実験の結果を、4 でその結果から、上記の問題がいずれも可能であることが結論できることを示し、5 で以上をまとめる。

2. 時系列パターン変換とアトラクタ

2.1 リカレントニューラルネットワークモデル

RNN の i 番目のユニットの動作は、連続系の場合、一般に次式のように表すことができる。

$$\begin{aligned} \frac{dx_i(t)}{dt} &= f_i(w, x(t), X(t)) \\ y_i(t) &= g_i(x_i(t)) \end{aligned} \quad (1)$$

ここで、 $x_i(t)$ 、 $y_i(t)$ はそれぞれ、ユニット i の時刻 t における内部状態、出力を表す。また、 w は、結合荷重ベクトルであり、時間に依存しない可調節パラメータである。 $x(t)$ は、RNN のすべてのユニットの内部状態を、 $X(t)$ は入力ユニットに入力される信号をベクトル表記したものである。 f_i 、 g_i は、一般に非線形関数であり、 g_i としては、シグモイド関数が用いられることが多い。以下、全ユニット数は N 個とする。

時刻 t における RNN の状態は、 $x_1, x_2, \dots, x_N$ で張られる N 次元空間内の一点として表すことができる。この空間が状態空間である。RNN の状態変化は、式 (1) の形を持つ N 個の連立非線形微分方程式の解として得られ、状態空間内の一つの軌道として表わされる。外部入力 $X(t)$ が時間に依存しない一定値であれば、RNN は自励系となる。また、 $X(t)$ が時間依存性を持つ非自励系の場合にも、その導関数が陽に記述できる場合には、式 (1) の微係数の次数を一次上げるにより自励系に帰着できる。しかし、実際の応用の際には、 $X(t)$ は

離散的な点列 ($X(t_1), X(t_2), \dots$) として与えられる場合がほとんどであり、導関数を正確に求めるのが極めて困難であるため、上記の手法を適用できない場合が多い。

2.2 時系列パターン変換

学習の対象としては、次の P 組の有限時間長の時系列パターン変換を考える。

$$\tilde{X}^{(p)} \rightarrow \tilde{Y}^{(p)} \quad (p = 1, 2, \dots, P) \quad (2)$$

ただし、

$$\begin{aligned} \tilde{X}^{(p)} &\equiv (X_0^{(p)}, \dots, X_t^{(p)}, \dots, X_{T-1}^{(p)}) \\ \tilde{Y}^{(p)} &\equiv (Y_1^{(p)}, \dots, Y_t^{(p)}, \dots, Y_T^{(p)}) \end{aligned} \quad (3)$$

ここで、 $X_t^{(p)}$ 、 $Y_t^{(p)}$ はそれぞれ時刻 t におけるパターン p の入力信号ベクトル（入力ユニット数の次元を持つベクトル）、教師信号（出力ユニット数の次元を持つベクトル）である。これらは、 $X_{i,t}$ 、 $Y_{k,t}$ をそれぞれ時刻 t におけるユニット i の入力信号、ユニット k の教師信号とすると、次式のように表される。

$$\begin{aligned} X_t^{(p)} &= (X_{1,t}^{(p)}, \dots, X_{i,t}^{(p)}, \dots, X_{N,t}^{(p)}) \\ Y_t^{(p)} &= (Y_{1,t}^{(p)}, \dots, Y_{k,t}^{(p)}, \dots, Y_{N,t}^{(p)}) \end{aligned} \quad (4)$$

本論文では「パターン」を、式 (3) で定義したような、時系列のベクトル全体を表す語として用いる。

本論文では、式 (1) で表される力学系において、所望の時系列パターン変換を行うことを目的とする。RNN において、そのダイナミクスが微分方程式で記述されることから、同じ入力パターンが与えられても、異なる初期状態により、異なる出力パターンが得られる。RNN を長時間動作させる場合、その内部状態は時間と共に変化するため、一般には、異なる時刻から与えられる、同一の入力パターンに対して、同一の出力パターンが得られることはない。以下に示すことを考慮することにより、この目的が実現可能であることを示す。

2.3 時系列パターン変換とアトラクタ

まず、RNN に周期的な外部入力を与えられる場合を考える。すなわち、 $\tilde{X}^p$ が無限に繰り返し与えられる場合である。このとき、式 (1) は強制振動系を表しており、したがって、状態空間内には、アトラクタが形成されることが知られている。すなわち、RNN の内部状態は $t \rightarrow \infty$ では、ポイントアトラクタ、周期アトラクタ、カオスアトラクタのいずれかに引き込まれる。

次に、有限時間内で定義された入力信号 $\tilde{X}^{(p)}$ を考える。この場合でも、次の条件が満足されるならば、上記入力に対応したアトラクタが状態空間内に形成されたと、近似的に考えることができる。すなわち、内部状態がアトラクタに引き込まれるまでの時間が、外部入力パターン $\tilde{X}^{(p)}$ の時間長 T に比べて、十分に短いという条件

Fig. 1 Attractors corresponding to time sequential input patterns

が満足される場合である。Fig. 1 に、これを模式的に示す。(a), (b) は、時系列入力パターン $\tilde{X}^{(1)}$, $\tilde{X}^{(2)}$ アトラクタ $a^{(1)}$, $a^{(2)}$ が、それぞれ対応しているようすを示している。

ここで、(c) のように、入力信号として $p = 1, 2$ のパターンが連続して入力された場合を考えると、RNN の動作は以下の 5 つのステップに分解することができる。

- (1) アトラクタ $a^{(1)}$ への引き込み。
- (2) アトラクタ $a^{(1)}$ 上での状態変化。
- (3) 入力パターンの切り替えによるアトラクタ $a^{(1)}$ の消滅・アトラクタ $a^{(2)}$ の形成。
- (4) アトラクタ $a^{(2)}$ への引き込み。
- (5) アトラクタ $a^{(2)}$ 上での状態変化。

この場合にも、 $\tilde{X}^{(p)}$ に対応して所望の時系列出力 $\tilde{Y}^{(p)}$ が得られると共に、アトラクタ間の遷移のための時間が入力パターン $\tilde{X}^{(p)}$ の時間長 T と比較して十分短ければ、「時系列パターン変換 $\tilde{X}^{(p)} \rightarrow \tilde{Y}^{(p)}$ 」に対応したアトラクタが形成されたことになる。

2.4 RNN のモデルと学習法

RNN の具体的なモデルとして、additive net を取り上げる⁴⁾。このモデルは、外部入力が無い場合について任意の力学系を近似する能力を持つことが証明されてい

る³⁾。その動作は、次式で表現される。

$$\begin{aligned} \tau \frac{dx_i(t)}{dt} &= -x_i(t) + \sum_{j=1}^N w_{ij} y_j(t) + X_i(t) \\ y_i(t) &= g(x_i(t)) = \frac{1}{1 + \exp(-x_i(t))} \end{aligned} \quad (5)$$

ここで、 w_{ij} はユニット j からユニット i への結合荷重である。

additive net の入出力信号は、 N 個のユニットのうち、 L 個を入力ユニットと定め、外部からの入力パターンを与え、 K 個を出力ユニットと定め、この出力を additive net の出力パターンとして得る。additive net を用い、時系列パターン変換を行う場合、時刻 t から $t+T$ に渡り与えられた、時系列入力パターンに対して得られる、時刻 t から $t+T$ の時系列出力パターンを、その出力として用いる。これは、その可調節パラメータ w を適切に定めること（学習）により、様々な、時系列入出力変換を行うことができる。

数値実験では、式 (5) を時間軸上で離散化する。本論文では、離散化の時間間隔 $t_{\lambda+1} - t_{\lambda} = \delta t_{\lambda} \equiv \text{const}$ を 1 として正規化した。

式 (5) で表現される RNN に対して、種々の学習則が提案されているが、誤差の収束性が理論的に保証されている、Back Propagation Through Time (BPTT) 法を用いた。以下に概要を説明する^{5)~7)}。詳細は、文献 7) を参照されたい。

RNN の誤差 E は、式 (6) のように、出力値 y_k と教師信号 Y_k との差の 2 乗積分として定義する。

$$\begin{aligned} E &= \sum_{p=1}^P E^p \\ E^p &= \int_0^T \sum_k \frac{1}{2} (y_k^p(t) - Y_k^p(t))^2 dt \end{aligned} \quad (6)$$

ここで、結合荷重の 1 回の更新量 Δw_{ij} を

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} \quad (\eta > 0) \quad (7)$$

で与えれば、これによる誤差の変化 ΔE は

$$\begin{aligned} \Delta E &= \sum_{i,j} \frac{\partial E}{\partial w_{ij}} \Delta w_{ij} = -\eta \sum_{i,j} \left(\frac{\partial E}{\partial w_{ij}} \right)^2 \\ &\leq 0 \end{aligned} \quad (8)$$

となる。式 (6) で定義される誤差 E は、離散系でのリヤプノフ関数であるので、式 (7) に従い w_{ij} の更新を繰り返すことにより、減少することが保証されている。

具体的には、 w の 1 回の更新につき以下の手順を踏む。

- (1) 時刻 0 から時刻 T^p までの各時刻におけるユ

ユニット出力 $y = (y_1, \dots, y_N)$ を、式 (5) に基づき求める。

(2) y を用いて、時刻 T から時刻 0 までの、ラグランジュの未定係数 p_i を式 (9) より求める。

$$\tau \frac{dp_i}{dt} = p_i - \sum_j p_j w_{ji} g'(x_i) - \delta_i (y_i - Y_i) g'(x_i) \quad (9)$$

ただし、 δ_i はユニット i に教師信号が与えられている場合 1、そうでない場合 0 をとる変数であり、 $g'(x_i)$ は $g(x_i)$ の x_i に関する一階微分を表す。

(3) Δw_{ij} を式 (10) より求める。

$$\Delta w_{ij} = -\eta \int_0^T p_i y_j dt \quad (10)$$

この手法は、初期内部状態を固定したオフラインの学習法であるが、計算量、記憶容量などの観点で、初期内部状態が変動するオンライン学習則と比較して、優れている。

しかし、オフライン学習を用いた結果、RNN は学習時の初期内部状態からの動作のみ、保証されることに注意しなければならない。これは、微分方程式の解の初期値への依存性に起因する。RNN の内部状態は、時間に伴い変化するため、通常用いる際に、時系列パターンの入力開始時の内部状態が、学習時の初期状態と一致するとは限らない。このため、常に、学習した時系列パターンを出力することは、保証されない。しかし、以下の章に示す方法により、この問題を解決することができる。

3. 数値実験

本章では、2.3 で述べた条件が実際に満足されることを、数値実験により実証する。具体的には、次の2点を示す。

(1) RNN の学習により、複数の時系列入出力変換を、それぞれアトラクタに対応付けることが可能である。

(2) 学習パターンを適切に与えることにより、アトラクタへの引き込み時間の調節が可能である。

すなわち、RNN による有限時間長のパターン変換の学習は、学習パターンに対応するアトラクタの形成として理解されることを示す。

3.1 実験条件

RNN は、全ユニット数 $N = 5$ 、入力ユニット数 $L = 1$ 、出力ユニット数 $K = 1$ 、時間遅れ定数 $\tau = 4.0$ として実験を行った。また、学習前の結合係数 w_{ij} ($1 \leq i, j \leq N$) は、区間 $[-1, 1]$ の一様乱数により設定した。

Table 1 Learning Patterns

	Learning Pattern		Simulation		
	Input	Teacher	I	II	III
1	$\tilde{X}^1$	$\tilde{Y}^1$	○		
2	$\tilde{X}^2$	$\tilde{Y}^2$	○		
3	$\tilde{X}^1 \triangleright \tilde{X}^1 \triangleright \tilde{X}^1$	$\tilde{Y}^1 \triangleright \tilde{Y}^1 \triangleright \tilde{Y}^1$		○	○
4	$\tilde{X}^2 \triangleright \tilde{X}^2 \triangleright \tilde{X}^2$	$\tilde{Y}^2 \triangleright \tilde{Y}^2 \triangleright \tilde{Y}^2$		○	○
5	$\tilde{X}^1 \triangleright \tilde{X}^2$	$\tilde{Y}^1 \triangleright \tilde{Y}^2$			○
6	$\tilde{X}^2 \triangleright \tilde{X}^1$	$\tilde{Y}^2 \triangleright \tilde{Y}^1$			○

学習パターンの一例として、異なる2種類の変換を取り上げる(付録参照)。

変換 1: $\tilde{X}^{(1)} \rightarrow \tilde{Y}^{(1)}$

$$\text{Input } X^{(1)}(t) = 1 - \sin\left(\frac{\pi}{24}t\right)$$

$$\text{Output } Y^{(1)}(t) = 0.65 - 0.1 \cos\left(\frac{\pi}{12}t\right) \\ t = 0, 1, 2, \dots, 23$$

変換 2: $\tilde{X}^{(2)} \rightarrow \tilde{Y}^{(2)}$

$$\text{Input } X^{(2)}(t) = -1 + \sin\left(\frac{\pi}{24}t\right)$$

$$\text{Output } Y^{(2)}(t) = 0.45 + 0.1 \cos\left(\frac{\pi}{12}t\right) \\ t = 1, 2, 3, \dots, 24$$

ここで、変換 1 および 2 の6種類の組み合わせを、学習パターンとして定義する(Table 1)。ただし、 $\tilde{X}^{(p)} \triangleright \tilde{X}^{(q)}$ は、 $\tilde{X}^{(p)}$ の後に $\tilde{X}^{(q)}$ が、切れ目なく続くことを表わすものとする。実験は、用いた学習パターンの組み合わせにより、実験 I ~ III に分けられる。実験 I ~ III において、Table 1 の、○を付した学習パターンを、学習時に RNN に提示するパターンとして用いた。すなわち、実験 I では、 $\tilde{X}^{(p)} \rightarrow \tilde{Y}^{(p)}$ をそのまま、実験 II では、これらを3回繰り返したものを、実験 III では、実験 II に加えて、異なるパターンの組み合わせが連続するものを提示した。これらのパターンに対し2.4 に示した BPTT 法と、探索を高速化する谷探索法⁸⁾を用い学習をおこなった。

これらのパターンの定め方により、状態空間内における軌道のようなことが変化することを示す。なお、ここで用いた教師パターンの組み合わせを選んだ理由は、4節で述べる。

RNN の初期状態は次のように定めた。

$$\left\{ \begin{array}{l} \text{出力ユニット以外} \\ y_i(0) = 0.5 \quad (i = 1, 2, 3, 4) \\ \text{出力ユニット} \\ y_5(0) = \left\{ \begin{array}{ll} 0.45 & : \text{時刻0から変換1(11)} \\ & \text{を期待する場合} \\ 0.55 & : \text{時刻0から変換2} \\ & \text{を期待する場合} \end{array} \right. \end{array} \right.$$

3.2 実験結果

実験結果を以下に示す。Figs. 2~4 は、それぞれ実験 I ~ III の結果を表している。左図 (a) は入力 $X(t)$ および実際の出力 $y(t)$, 教師 $Y(t)$ の時間変化を表し、右図 (b) は RNN の状態変化を状態空間の軌道として描いたものを表す。状態空間は 5 次元でありすべてを図示できないため、(b) では横軸を入力ユニットの出力 y_1 , 縦軸を出力ユニットの出力 y_5 と定めた 2 次元平面に軌道を投影した。図中の四角は、時刻 0 における RNN の状態を表している。なお、出力 y_i は状態 x_i と 1 対 1 に対応しているため、便宜上、これを「状態」と呼ぶことにする。

ほかの y_i の組み合わせによる 2 次元投影図も観察した結果、以下の議論に本質的な違いのないことを確認したことを、付け加えておく。

3.2.1 実験 I

Fig. 2 に、実験 I の結果を示す。

I-1 周期的入力

(i), (ii) は、入力パターン $\tilde{X}^{(1)}$, $\tilde{X}^{(2)}$ それぞれの、繰り返し入力に対する応答を表している。

学習パターンは、Table 1 のパターン 1, 2 である。(i) - (a), (ii) - (a) より、学習時の t の定義域 ($0 \leq t \leq 24$) では、出力信号が教師信号と一致している。しかし、学習した区間以外 ($t > 24$) では、教師信号と一致していない。状態空間内に描かれている軌道は、(i) - (b), (ii) - (b) より、初期状態から 20 ~ 30 ステップをかけて徐々に、それぞれの周期軌道に引き込まれている。

I-2 入力パターンの切り換え

(iii), (iv) に、時刻 48 において、入力パターンが $\tilde{X}^{(1)}$ から $\tilde{X}^{(2)}$ へ切り換えられた場合の動作を示す。

(iii) は、先に $\tilde{X}^{(1)}$ が、入力された場合の応答である。 $t = 0$ から $t = 47$ までは、(i) と同じである。 $t = 48$ で、入力パターンが切り替わった後、約 30 ステップの過渡応答 (図中 A) を経て、(ii) の場合とほぼ同じ出力を示している。このときの状態空間の軌道 (iii) - (b) は、初期状態から周期軌道 (i) - (b) に近づいてゆき、入力パターンの切り替えと同時に、周期軌道 (ii) - (b) に近づいてゆく。このとき、二つの周期軌道の間で見られる過渡状態は、30 ステップ程度であることも確認できる。

(iv) においても、同様の傾向が確認できる。

I-3 ランダム入力

入力パターン $\tilde{X}^{(1)}$, $\tilde{X}^{(2)}$ を、ランダムに切り替えた (v) では、 $0 \leq t \leq 24$ のみで所望の出力が得られているが、 $t > 24$ では、(i) ~ (iv) の場合とは、応答が異なる。(v) - (b) の、状態空間では、(i) - (b), (ii) - (b) の周期軌道がいずれも見い出せない。周期軌道への引き込

みは、30 ステップが必要であることを、上で述べた。このステップ数は、入力パターンの長さである 24 ステップを上回っており、周期軌道に引き込まれる前に、入力パターンが切り換わることになる。

3.2.2 実験 II

Fig. 3 に実験 II の結果を示す。

II-1 周期的入力

(i), (ii) では、実験 I の場合と異なり、学習パターンに相当する $0 \leq t \leq 72$ 以降でも、所望の出力が得られている。状態空間の軌道は、実験 I の場合と同様に、初期状態から約 20 ステップをかけて、実験 I の場合とは形状の異なる周期軌道に引き込まれている。

II-2 入力パターンの切り換え

(iii), (iv) では、 $t = 48$ で入力パターンが変化した後、約 30 ステップの過渡状態 (図中 B) を経て、所望の出力が得られていることが分かる。状態空間の軌道は、約 30 ステップをかけ、(i) (ii) で得られた周期軌道間を遷移する結果となっている。

II-3 ランダム入力

(v) では、 $0 \leq t \leq 24$ および $55 \leq t \leq 70$ 付近以外では、所望の出力が得られていない。これは、実験 I と同様、周期軌道間の遷移に 20 ステップ程度の時間を要するためである。実験 I と異なる点は、 $55 \leq t \leq 70$ 付近では遷移がほとんど終了し、所望の出力に近い波形が得られていることである。(v) - (b) において、状態空間に 2 つの周期軌道 (i) - (b), (ii) - (b) のいずれも見出せない結果となっている点は、実験 I と同様である。

3.2.3 実験 III

Fig. 4 に実験 III の結果を示す。

III-1 周期的入力

(i), (ii) では、実験 II の場合と同じく、すべての時刻で所望の出力が得られている。初期状態から周期軌道に引き込まれるまでの、ステップ数は 3 ステップ程度と、各パターンの時間長と比較して極めて短い点が、実験 I, II の場合と異なる。

III-2 入力パターンの切り換え

(iii), (iv) では、実験 II の場合と同じく、過渡状態 (図中 C) を除いて所望の出力が得られている。ただし、実験 III では過渡状態が 3 ステップ程度とパターン長の 24 ステップと比べ非常に短く、ほとんどの時刻において所望の出力が得られる。

III-3 ランダム入力

ランダムな順序で入力パターンを与えた場合、実験 I, II では所望の出力が得られなかった。しかし、過渡状態が短く所望の出力が得られていること、および、状態空間で周期軌道 (i), (ii) の間を数ステップで遷移する遷

Fig. 2 Results learned with pattern set I

移過程が確認できる。また、遷移はほとんど同じ軌道を通っていることも確認できる。この結果より、時刻 0, 24, 48, 72, ... における内部状態と、学習時に用いた初期状態 (式 (11)) が異なることから、以下のことが示される。学習の結果得られた RNN に、学習時の初期状態と異なる状態から、同一の時系列入力パターン (たとえば、時刻 0 および 72 から $\tilde{X}^{(1)}$) を与えた場合も、同一の時系列出力パターン (時刻 0 および 72 から $\tilde{Y}^{(1)}$) が得られる。

4. 考 察

3 で述べた 2 点の観点から、以上の実験結果を考察

する。

最初に、「RNN の学習により、複数の時系列入出力変換を、それぞれアトラクタに対応づけることが可能であるか」という問題に対して考察してみる。

実験 I~III により得られた状態空間内の軌道は、各図 (i), (ii) より、異なる二種類の入力パターンに対して、二種類の周期軌道を形成する。また、二種類の入力パターンを切り替えた (iii), (iv) では、最終的には、内部状態はそれぞれの周期軌道に引き込まれている。この結果より、これらの周期軌道は、周囲の状態を引き込む軌道、すなわち、アトラクタであると言える。実際、実

Fig. 3 Results learned with pattern set II

験 I~III を通じて、ネットワークの初期状態をこれらの周期軌道の周囲に対応する状態に定め、 $\tilde{X}^{(1)}$ を繰り返し与えた時、(i)、(ii) と同じ周期軌道に引き込まれることを確認した。

したがって、上記の問題に対する答えは、「可能」である。以降、実験で得られた周期軌道を、周期アトラクタと呼んでも差し支えなからう。

次に、「アトラクタへの引き込みが学習により調節可能であるか」に関して考察する。

実験 I では、(i)、(ii) より RNN が周期的な外部入力により、周期アトラクタに引き込まれてゆくことが示さ

れているが、出力は所望の波形が得られていない。また、(iii)、(iv) の網掛けの区間 A、A' は、内部が一つのアトラクタから離れ、別のアトラクタへの引き込まれていく遷移過程であるが、この過程を辿る約 30 ステップは、パターンの時間長 24 ステップに比べ長い時間を要している。

実験 II では、(i) (ii) については、所望の入出力関係が満たされている。これは、実験 I の 3 倍の長さの学習パターン $(\tilde{X}^{(p)} \triangleright \tilde{X}^{(p)} \triangleright \tilde{X}^{(p)}) \rightarrow (\tilde{Y}^{(p)} \triangleright \tilde{Y}^{(p)} \triangleright \tilde{Y}^{(p)})$, ($p = 1, 2$) を与えたためであると考えられる。すなわち、入出力パターンを繰り返し与えることにより、RNN の状態

Fig. 4 Results learned with pattern set III

空間内の軌道は、所望の入出力を満たすアトラクタを形成し、一旦そのようなアトラクタ形成されれば、学習区間に対応する時刻以降も、アトラクタから外れることはない。しかし、(iii) ~ (iv) 内の網掛け B, B' より、アトラクタ間の遷移に 30 ステップを要しており、基本パターンの長さ 24 ステップと比べ長い。

すなわち、実験 I, II で与えた学習パターンでは、アトラクタは形成されるが、それらのアトラクタの性質として、入力パターンの切り替えにたいする応答が、満足されていない。

実験 III では (i) ~ (v) すべてについて、所望の入

出力関係を満たしている。また、(iii), (iv) 内の網掛け C, C' より、アトラクタ間の遷移過程が 3 ステップ程度と、実験 I, II と比較してかなり短いだけでなく、パターンの時間長と比べても、十分短い時間となっている。この結果は、実験 III で得られたアトラクタが、所望の入出力関係を満たすと共に、2 種類の入力パターンの切り替えに対しても、素早く反応することのできる性質を備えていることを示している。このような性質を備えたアトラクタが、これ以外にも存在するかどうかは、今後の研究課題として残されている。

実験 III と実験 I, II の差異については、次のよう

に考えることができる。実験 III は、実験 II で用いた RNN に対して、学習パターンとして $(\tilde{X}^{(p)} \triangleright \tilde{X}^{(q)}) \rightarrow (\tilde{Y}^{(p)} \triangleright \tilde{Y}^{(q)})$ を新たに追加して、学習を行った。前半の $\tilde{X}^{(p)}$ を順に入力に与えてゆくと、RNN の内部状態は、式 (11) で与えられた初期状態から変化してゆく。続けて、後半の $\tilde{X}^{(q)}$ を入力に与え学習することにより、 $\tilde{X}^{(q)}$ から $\tilde{Y}^{(q)}$ への変換を、式 (11) とは異なる内部状態を与えて、学習したことになる。

前論文では、異なる内部状態から同じ入出力変換を学習することにより、所望の入出力変換に対応したアトラクタを形成できることをすでに報告したが、初期状態の選定をどうするかが、残された問題として指摘されていた²⁾。実験 III の結果を見れば、上記のような学習パターンを用いることで、初期状態の選定の問題を解決できることが分かる。初期状態とは異なった状態を、RNN が自動的に選択し、学習するので、適切な学習が可能となった。

以上より、条件 (1), (2) を共に満たす学習パターンについて、次のようにまとめることができる。

- (a) 学習したい入出力関係を周期的な信号と見なし、これを複数周期分 (たとえば 3 周期分) 一度に提示する。これは、Table 1 における学習パターン 3, 4 に対応する。
- (b) 入力信号として考え得る組み合わせを作成し、それぞれ学習パターンとして提示する。これは、Table 1 における学習パターン 5, 6 に対応する。

すなわち、状態空間内のアトラクタを入出力パターンに対応付けることが、(a) により実現され、アトラクタ間の遷移を調節することが、(b) により可能であることを示せた。

なお、学習パターンの組み合わせ方としては、Table 1 以外の種々の組み合わせを用いて実験を行ったが、実験 III と同等な結果を得るためには、学習パターン 3, 4, 5, 6 の組み合わせが最も単純なものであった。

この結果は、本論文で用いた時系列入出力変換が、ほとんどの時刻で $X(t)$, $Y(t)$ が一対一対応していることから、RNN が、過去の履歴をほとんど保持しない各時刻 t における値の変換 $X(t) \rightarrow Y(t)$ を行うだけの、学習を行ったようにとらえることもできる。この点に関して、付録に示したような $X^{(p)}(t)$, $Y^{(p)}(t)$ が一対一対応していない時系列入出力変換を学習した場合も、同様の結果を得ることができることを、確認してある。

5. まとめおよび今後の課題

本論文では、RNN のモデルとして additive net を用いて、所望の時系列入出力関係と状態空間内のアトラク

タを対応付けることが可能であると共に、2 種類の時系列入力パターンの切り替えに対する、アトラクタ間の遷移を調節することが可能であることを、シミュレーションにより示した。さらに、この結果を実現するために必要な学習パターンの与え方、および、学習のメカニズムについて考察を行った。特に、同一時間長の 2 種の入出力変換を取り上げて行った数値実験の結果を、詳細に示した。なお、時空間パターンを外部入力とする RNN に対して、力学系の観点から理論的解析を行なう必要があるが、この点からの研究報告が最近行なわれたことを記しておく^{10), 11)}。

実用的な観点から興味深いのは、入力、および出力のそれぞれの時間長が異なるパターンを取り扱う問題である。RNN の応用を広げるには不可欠である。これに対しても、本論分で展開したような RNN の動作を調べる研究が基本となる。ただし、本手法の欠点として、学習する入出力パターンの増大に伴い、学習パターンの組み合わせが急激に増加するという問題点がある。さらに、本学習法は、基本的に、各パターンの提示後に状態をリセットするオフライン学習である。このため、入力パターンが常にネットワークに入力され続けるようなオンラインの学習が必要な場合には、適用できない。これらの問題を解決する手法として、BPTT を応用した、時間前向き計算による実時間学習法が提案されている⁹⁾。しかし、その有効性は実験的に検証されているものの、学習の収束性の証明にまでには至っていない。収束性の証明には、本研究での学習パターンの生成法が大きなヒントになると考えられる。今後、本研究結果を踏まえ、収束性を理論的に証明することが、重要な研究課題である。

参考文献

- 1) 郷原 一寿, 高瀬 治彦, 内川 嘉樹: リカレントニューラルネットワークとアトラクタ I, 電子情報通信学会技術研究報告, **NC92-64** (1992).
- 2) Gouhara K., Takase H. and Uchikawa Y.: Attractor learning of recurrent neural networks, *Artificial Neural Networks*, **2**, 371/374 (1992).
- 3) Funahashi K., Ichikawa A.: Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks, *Neural Networks*, **vol. 6**, 801/806 (1993).
- 4) Seidl D.R. and Lorenz R.D.: A Structure by which a Recurrent Neural Network Can Approximate a Non-linear Dynamic System, *IJCNN-Seattle-WA*, **vol. 2**, 709/714 (1991).
- 5) Pearlmutter B. A.: Learning state space trajectories in recurrent neural networks, *Neural Computation*, **1**, 263/269 (1989).
- 6) Sato M.: A learning algorithm to teach spatio-temporal patterns to recurrent neural networks, *Biol. Cybern.*, **62**, 259/263 (1990).
- 7) 渡辺 辰巳, 郷原 一寿, 内川 嘉樹: リカレントニューラルネットワークの各学習則に関する検討及び学習曲面の

Fig. A.1 Another example of input-output pattern transformation

形状, 電子情報通信学会論文誌, Vol.J74-D-II No.12, 1776/1786 (1991).

- 8) 横井 邦雄, 郷原 一寿, 内川 嘉樹: リカレントニューラルネットワークに対する谷探索法の適用, 電子情報通信学会技術研究報告, NC92-37 (1992).
- 9) 高木 秀彦, 水野 政司, 郷原 一寿: リカレントニューラルネットワークの BPTT を応用した時間前向き計算による学習法とアトラクタ, 電子情報通信学会論文誌, Vol.J77-D-II No.2, 422/430 (1994).
- 10) 郷原一寿: 時空間パターンに励起されたニューロダイナミクス - 励起アトラクタモデル -, 電子情報通信学会技術研究報告, NC94-153 (1995).
- 11) 高木秀彦, 郷原一寿: 励起アトラクタの実験的検証, 電子情報通信学会技術研究報告, NC94-154 (1995).

《付 録》

本論文では結果の見やすさから, 3.1 に示した時系列パターン変換を用いた実験結果を示した。これは, ほとんどの時刻において, $X^{(p)}(t)$, $Y^{(p)}(t)$ が一対一対応しており, 時系列パターンの学習例としては単純である。しかし, ほかの多くの入出力パターン変換に関しても実験を行い, 定性的に同じ結果が得られることを確認した。Fig. A.1 に, このとき用いた学習パターンの例として, $X^{(p)}(t)$, $Y^{(p)}(t)$ が一対一対応していない入出力変換パターンを示す。ここで, $\tilde{Y}^{(1)}$, $\tilde{Y}^{(2)}$ は, 3.1 で示した変換 1, 2 のものと同一である。

[著 者 紹 介]

高 瀬 治 彦

1996 年名古屋大学大学院工学研究科博士課程後期課程修了。工学博士。同年, 三重大学工学部助手。現在に至る。ニューラルネットワークのダイナミクスに関する研究に従事。電子情報通信学会, 神経回路学会各学会会員。

郷 原 一 寿 (正会員)

1982 年名古屋大学大学院工学研究科修士課程修了。名大工学部助手, 講師を経て 1992 年中部大学工学部電気工学科助教授。94 年北大工学部応用物理学科助教授, 現在に至る。非線形動力学系としてのニューラルネットワークに関する研究に従事。応用物理学会, 電子情報通信学会, IEEE などの会員。工学博士。

内 川 嘉 樹 (正会員)

1969 年名古屋大学大学院博士課程満了。以後同大学助手, 助教授を経て, 現在同電子情報学科教授。ニューラルネットワーク、ファジィ理論、遺伝的アルゴリズム、自律分散システム、バーチャルリアリティー、電子計数法などを利用したシステム技術の基礎および産業への応用に関する研究に従事。工学博士。