カオスニューラルネットを用いた記憶タスクの強化学習におけるカオス性の影響

大分大学　○大石将人　柴田克成

Effect of Chaos on Reinforcement Learning of Memory Tasks Using Chaotic Neural Networks

Oita University ○ Masato OOISHI, Katsunari SHIBATA

Abstract: As an essential technique for the emergence of thinking, we have proposed a new reinforcement learning (RL) framework using a chaotic neural network. In this framework, learning of feedback connection weights is critical, but has not been well done. Then in this paper, intentionally in the conventional RL framework, mutual connection weights, which control the chaoticity, and positive self-feedback connection weights, which are good for learning of memory tasks are both varied, and chaoticity and learning performance are observed. As a result, when the mutual connection weight scale is large, chaoticity is large and the learning performance becomes worse not depending so much on the self-feedback connection weight value. However, in the new RL, since chaoticity is required for exploration, more detailed investigation is necessary.

1　序論

現在、深層学習をはじめとした人工知能技術が脚光を浴び、様々な状況に柔軟に対応できる汎用人工知能の実現が注目されている。われわれは以前よりこのような知性をニューラルネット (NN) と強化学習の組み合わせによって形成することを目指している。そして「記憶」、「予測」といった機能の創発を簡単なタスクで確認してきた [1][2]。しかし、典型的な高次機能である「思考」については創発が確認できていない。

頭の中で考えを巡らせる「思考」とあれこれ迷う「探索」はともに自律的な多段階状態遷移の内部ダイナミクスという点で類似しており、両者の違いは合理性の有無にあると考えた。そこで「ランダムに近い状態遷移である探索が学習によってより合理的な思考へと成長する」という仮説を立てた。そして元々不規則な状態遷移をもつカオスダイナミクスを NN に導入したカオスニューラルネット (ChNN) を使用する新しい強化学習の枠組みを提案した [3]。

しかし、この学習法ではネットワーク内部で作り出す探索成分を陽に取り出せないため、強化学習により教師信号を生成できず、従来用いてきた BPTT(Back Propagation Through Time) 法が適用できないという問題点がある。新しく提案した学習法 [3] ではフィードバック結合 (FB) 部の学習ができず、また、何とか教師信号を作り出す方法も確立できていない。さらに、カオスダイナミクス生成のために FB 部の重み値スケールを大きくすると、そもそも BPTT 法において勾配発散問題が生じてうまく学習できなくなるという懸念も残る。

一方、われわれは、従来型の強化学習で BPTT 法を用いて記憶タスクを学習させる際、各ニューロンのセルフフィードバックの結合重み値を正の値に設定することで、記憶のためのアトラクタの形成を容易にし、同時に勾配消失、発散の問題を回避してきた [1][2]。

そこで本研究では、ChNN を用いて強化学習を行う際、そもそも BPTT 法を用いて学習できるかを知るとともに、その際の FB 部の重み値設定に関する知見を得ることを目的とする。そのため、あえて従来型の強化学習を用い、カオス性を左右するニューロン相互のフィードバック結合の重み値（相互結合）と、記憶の学習を容易にするセルフフィードバック結合の重み値（セルフ結合）を同時に変化させ、ネットワークのダイナミクスを観察するとともに、強化学習の性能を比較する。

2　カオスニューラルネットの強化学習

ChNN における中間層と出力層の内部状態 $\boldsymbol{u^{hid}}$, $\boldsymbol{u^{out}}$ および各層の出力の計算式を以下に示す。

$$\boldsymbol{u}_t^{hid} = \boldsymbol{W}^{hid}\boldsymbol{o}_t^{in} + \boldsymbol{W}_{FB}\boldsymbol{o}_{t-1}^{hid} \quad (1)$$

$$\boldsymbol{u}_t^{out} = \boldsymbol{W}^{out}\boldsymbol{o}_t^{hid} \quad (2)$$

$$\boldsymbol{o}_t = \tanh(\boldsymbol{u}_t) \quad (3)$$

$\boldsymbol{W}$, $\boldsymbol{W}_{FW}$ はそれぞれ各層の下層からの入力に対する重み値、FB 結合の重み値を表す。中間層の内部状態を計算する式 (1) においてランダムに決定する FB 結合 $\boldsymbol{W}_{FB}$ のスケールを大きくすることで ChNN とすることができる。重み値の更新は強化学習によりエージェント自身で教師信号を生成し、BPTT 法による教師あり学習を行う。本研究では強化学習法として、連続的な行動を扱うことができる Actor-Critic 法を用いる。時刻 t におけるエージェントの状態 S_t での Actor の出力を $\boldsymbol{A}(\boldsymbol{S}_t)$ とすると、実際の動作ベクトル $\boldsymbol{A'}(t)$ は探索成分ベクトル $\boldsymbol{rnd}_t$ を加えるため、式 (4) のように求める。

$$\boldsymbol{A'}(t) = \boldsymbol{A}(\boldsymbol{S}_t) + \boldsymbol{rnd}_t \quad (4)$$

状態 $\boldsymbol{S}_t$ における Critic の状態評価出力を $V(\boldsymbol{S}_t)$ として、学習に用いる TD 誤差 $\hat{r}_t$ は式 (5) で求める。

$$\hat{r}_t = r_{t+1} + \gamma \cdot V(\boldsymbol{S}_{t+1}) - V(\boldsymbol{S}_t) \quad (5)$$

ここで r_{t+1} はエージェントが時刻 $t+1$ でもらう報酬であり、本研究ではスイッチを通ってゴールへ到達した場合に $r_{t+1} = 0.8$ が与えられる。γ は割引率で、ここでは 0.96 としている。TD 誤差 $\hat{r}_t$ を用いて、Actor, Critic それぞれの教師信号 T_{Vt}, $\boldsymbol{T}_{At}$ は以下のように求める。

$$T_{Vt} = \hat{r}_t + V(\boldsymbol{S}_t) \quad (6)$$

$$\boldsymbol{T}_{At} = \boldsymbol{rnd}_t \cdot \hat{r}_t + \boldsymbol{A}(\boldsymbol{S}_t) \quad (7)$$

この教師信号と出力の誤差を BPTT 法により、時間を遡って学習する。

本研究で用いる記憶タスクとは目標の到達に記憶が必要なタスクのことである。ここではエージェントがスイッチを通ってからゴールすることで報酬を与えるスイッチタスクを行う。エージェントは、スイッチ上にいる間だけ *switch-flag* =1 の入力を受け取るため、それをネットワーク内部で保持し、行動を切り替える必要がある。ChNN でスイッチタスクが学習できれば、カオスダイナミクス上に記憶のためのアトラクタが作られたと考えられる。

中間層ニューロン自身からのフィードバックであるセルフ結合、つまり FB 結合重み値行列 $\boldsymbol{W}_{FB}$ の対角成分を正の値とすることで、記憶タスクの学習が容易になる。外部入力の影響を除いてダイナミクスを決める式 (1), (3) を原点近傍で線形近似すると、以下のようになる。

$$\begin{aligned} \boldsymbol{o}_t &= \tanh'(0)\boldsymbol{W}_{FB}\boldsymbol{o}_{t-1}^{hid} \\ &= \boldsymbol{W}_{FB}\boldsymbol{o}_{t-1}^{hid} \end{aligned} \quad (8)$$

ここで、対角成分以外の相互結合を 0 とし、対角成分であるセルフ結合の値を 1 とすると、遷移を表す行列は単位行列となる。これによって状態がそのまま伝播し、重み値を少し増加させるだけで容易にアトラクタの形成ができるようになる。さらに BPTT 法適用時の勾配消失、発散の問題も解決できる。しかし、カオスダイナミクスを生成するには、ランダムに決める相互結合の重み値のスケールを大きくする必要がある。

本研究ではカオス生成のために相互 FB の重み値としてセットする乱数のスケールと、記憶タスクの学習を容易にするためのセルフ結合の重み値の値を両方変化させてダイナミクスと学習の様子を観察する。

3 シミュレーション

本研究では、ChNN の強化学習によってフィードバック部の学習が可能かどうか調べるため、前述のスイッチタスクを行う。Fig.1 にネットワークの構成とスイッチタスクの概要を示す。

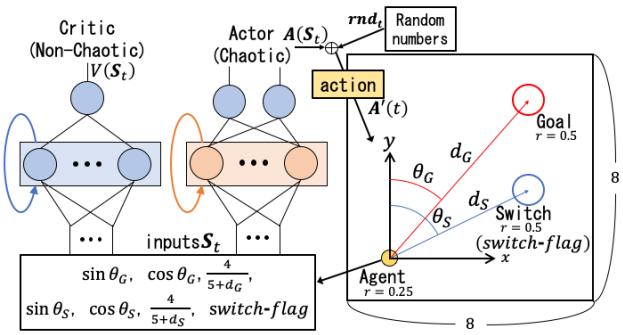


Fig. 1: Learning system and Switch task

ネットワークへの入力は Fig.1 よりエージェントとスイッチ、ゴールそれぞれとの距離、角度の情報 (sin, cos) に *switch-flag* を加えた 7 つである。また、ここでは動作生成部 Actor のカオスダイナミクスが状態評価部 Critic の学習に影響を与えないよう Fig.1 のようにネットワークを二つ用意する (ActorNN, CriticNN)。

ActorNN の出力は x 方向、y 方向のベクトルでこれに探索成分となる大きさ [-1.0, 1.0] の乱数を加えた分だけ移動する。CriticNN の出力は式 (5) で用いられる状態評価出力 V($\boldsymbol{S}_t$) として扱われる。

Table 1 にシミュレーションで用いたネットワークのパラメータを示す。

Table 1: Parameters for two networks(Actor, Critic)

Name		ActorNN	CriticNN
		Chaotic	Non-Chaotic
Number of hidden neurons		60	60
Initial weight	output layer	0	0
	hidden Layer	[-0.25,0.25]	[-0.25,0.25]
	self-feedback	**varied**	1.0
	other feedback	**varied**	[-0.25,0.25]
Learning rate	output layer	0.005	0.005
	hidden layer	0.005	0.005
	feedback	0.00025	0.00025
Traceback times in BPTT		20	20

本研究では、カオス性を測る指標としてリアプノフ指数を用いる。リアプノフ指数は、ActorNN の 60 個のニューロンに [-1.0, 1.0] の一様乱数を内部状態として与えたネットワークを二つ用意し、一方に大きさ 10^{-10} の摂動を加え、100 ステップ回した後の 500 ステップで測定した。また、リアプノフ指数を測る際に二つのネットワークの距離を毎ステップ正規化している [4]。

まず学習前のネットワークで、相互結合の重み値を決める一様乱数のスケールとセルフ結合の重み値をそれぞれ変化させてダイナミクスのカオス性を調べた。その結果を Fig.2 に、学習前で入力を与えない場合の中間層ニューロン 3 つの出力の時間変化を Fig.3 に示す。

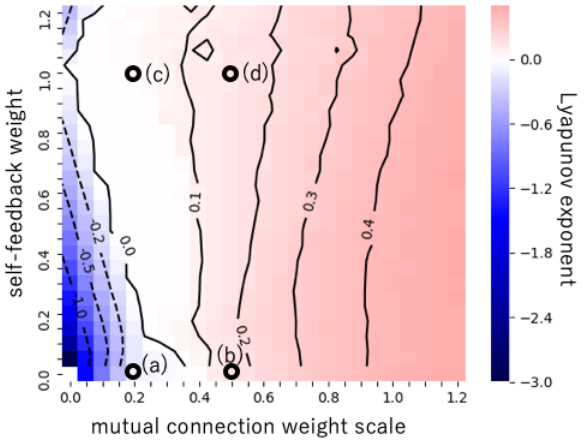
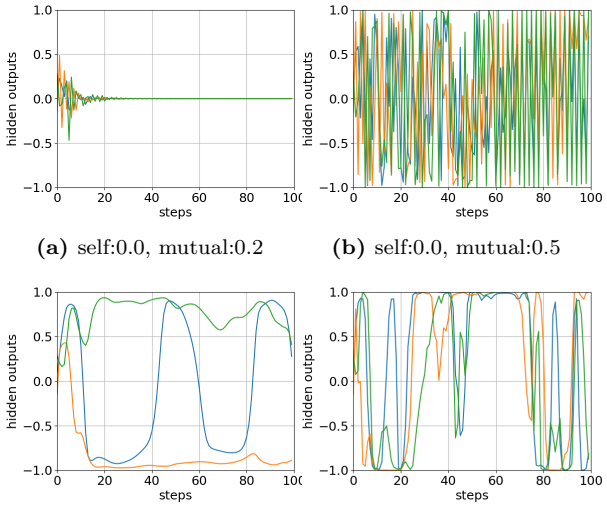


Fig. 2: Lyapunov exponent depending on the feedback connection weights.
(Vertical axis : self-feedback connection weight
Horizontal axis : mutual connection weight scale)



(a) self:0.0, mutual:0.2 **(b)** self:0.0, mutual:0.5

(c) self:1.0, mutual:0.2 **(d)** self:1.0, mutual:0.5

Fig. 3: Comparison of Network dynamics before learning

Fig.2 より、相互結合のスケールが大きいほどリアプノフ指数が大きくなっており、カオス性が強いことが分かる。Fig.3a と Fig.3b, Fig.3c と Fig.3d をそれぞれ比較すると、相互結合のスケールが大きい Fig.3b、Fig.3d の方が中間層ニューロンの出力の変化が激しくなっている。

また Fig.2 より、相互結合重み値が 0~0.4 程度の小さい部分では、セルフ結合が増加して 1 に近づくとリアプノフ指数が増加し、0 に近い Edge of Chaos の状態になり、セルフ結合がさらに増えるとリアプノフ指数は再び減少することがわかる。一方、相互結合重み値が 0.4 より大きい部分では、セルフ FB が大きいほどリアプノフ指数が少し小さくなる傾向が見られる。また、Fig.3 を見ると、セルフ結合を 1.0 とすることでニューロンの出力

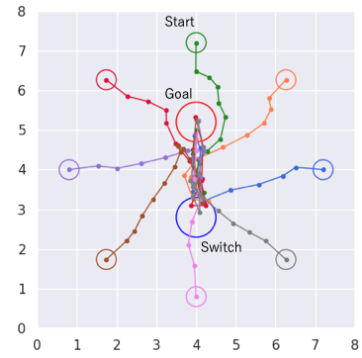
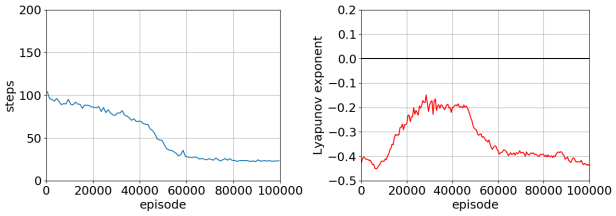
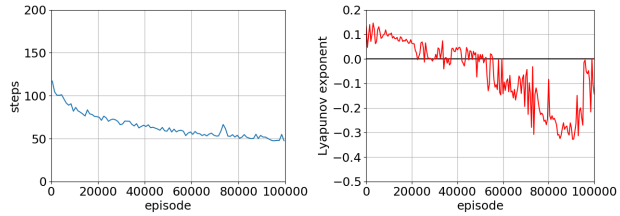


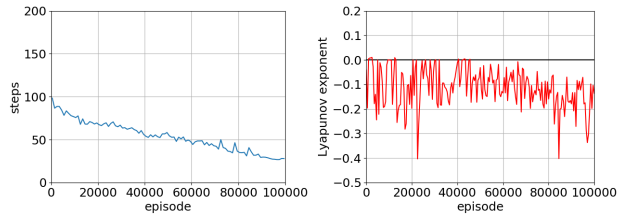
Fig. 4: 8 trajectories after learning without adding exploration component((c)self:1.0, mutual:0.2)



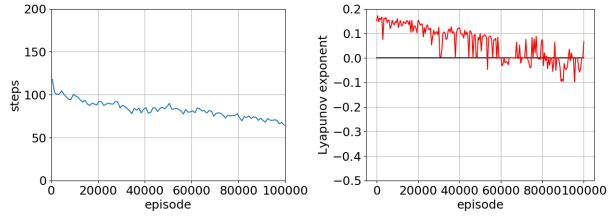
(i) Learning curve **(ii)** Lyapunov exponent
(a) self : 0.0, mutual : 0.2



(i) Learning curve **(ii)** Lyapunov exponent
(b) self : 0.0, mutual : 0.5



(i) Learning curve **(ii)** Lyapunov exponent
(c) self : 1.0, mutual : 0.2



(i) Learning curve **(ii)** Lyapunov exponent
(d) self : 1.0, mutual : 0.5

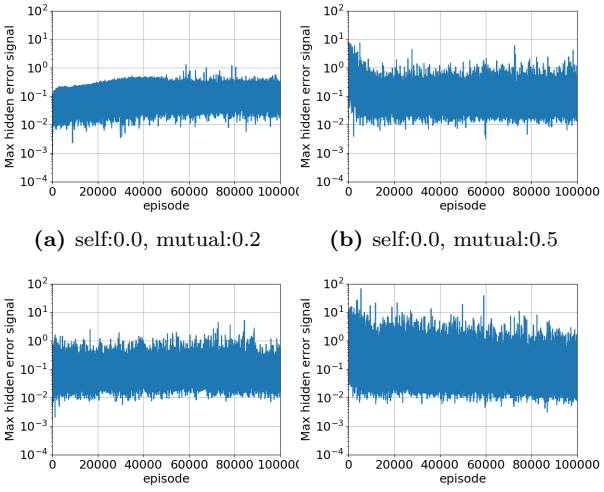
Fig. 5: Comparison of Learning result

の値の変化がゆっくりになり、特に相互結合のスケールが小さいときは 0 への収束がなくなっていることが分かる。また、Fig.3a と Fig.3c の場合では、Fig.2 より両者のリアプノフ指数はそれほど変わらず、セルフ結合を 1.0 にしてもカオス性が確認できる。

以上を踏まえて、Fig.3 で用いた ChNN にスイッチタスクの学習を行わせる。なお、Fig.2 の測定の際ではバイアスを加えていないが、ここでは、学習しやすくするため各ニューロンに [-1.0, 1.0] の初期バイアスを加え、学習によって変化させた。セルフ結合 1.0, 相互結合 0.2(Fig.2 の (c)) における、スイッチタスクの学習成功後のエージェントの軌道を Fig.4 に示す。その後、セルフ結合と相互結合をそれぞれ変化させ、スイッチタスクを行なった場合の学習曲線とリアプノフ指数を Fig.5 に示す。

Fig.4 の学習後のエージェントの軌道を見ると、エージェントが各スタートからまずスイッチを踏み、その後、ゴールへ向かっていることが分かる。学習性能を示す学習曲線を比較すると、相互結合が 0.5 でカオス性が大きい Fig.5b(i), Fig.5d(i) に比べ、相互結合が 0.2 でカオス性が小さい Fig.5a(i), Fig.5c(i) の方が、最終的により少ないステップ数でゴールできおており、学習性能が良いことが分かる。また、学習によるリアプノフ指数の推移を見ると、リアプノフ指数が常に負でカオス性が全くない Fig.5a(ii) を除き、学習とともにリアプノフ指数が下がっており、学習によってカオス性が小さくなっていることが分かる。

ChNN で BPTT 法を適用した場合の過去に伝播する誤差信号の増大を見るために、1 試行における全ニューロン、全時刻における BPTT 法の誤差信号の中で、絶対値が最大のものを毎試行プロットしたものを Fig.6 に示す。



(a) self:0.0, mutual:0.2 **(b)** self:0.0, mutual:0.5

(c) self:1.0, mutual:0.2 **(d)** self:1.0, mutual:0.5

Fig. 6: comparison of maximum absolute value of error signal during learning

Fig.6 より、相互結合が大きい場合に学習初期の誤差信号の最大値が大きい。また、セルフ結合を 1.0 とすると、全体的に誤差信号の最大値が小さくなりにくいことが分かる。また、相互結合が大きい時には、BPTT で時間を遡るほど誤差信号が大きくなっていることも確認した。このことが学習性能の悪化に繋がった可能性がある。

4 結論

カオスニューラルネットの相互フィードバック結合重み値とセルフフィードバック結合重み値を変化させて、ネットワークのダイナミクスを観察するとともに、探索成分を外部から加える従来の手法によるスイッチタスクの強化学習を行なった。その結果、セルフフィードバック結合重み値を正の値に設定してもカオス性は維持されたが、学習性能にはそれほど影響がなかった。さらに、相互フィードバック結合重み値スケールが大きいほどカオス性が強く、BPTT 法による誤差の増大が見られ、学習性能が低くなった。しかし、新しい強化学習では、カオスダイナミクスに基づいて探索を行う必要があるため、より細かい検討が必要である。

謝辞

本研究は JSPS 科研費 15K00360 の助成を受けたものである。ここに謝意を表する。

参考文献

[1] H. Utsunomiya, K. Shibata : Discovery of Pat- tern Meaning from Delayed Rewards by Reinforce- ment Learning with a Recurrent Neural Network, Proc. of IJCNN. 2011, pp. 1445-1452, 2011

[2] K. Goto, K. Shibata : Emergence of Flexi- ble Prediction-Based Discrete Decision Making and Continuous Motion Generation through Actor-Q-Learning, Proc. of ICDL-Epirob.ID 15, 2013

[3] Y. Sakashita, K. Shibata : Reinforcement Learning with Internal-Dynamics-based Exploration Using a Chaotic Neural Network, Proc. of IJCNN. 2015

[4] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, M. Asada : Information processing in echo state networks at the edge of chaos, Theory Biosci.(2012) pp. 207-208, 2012