カオスニューラルネットを用いた強化学習におけるカオス性の影響

大分大学　○後藤祐樹　柴田克成

Influence of Chaotic Property on Reinforcement Learning Using a Chaotic Neural Network

Yuki Goto and Katsunari Shibata, Oita University

Abstract: Aiming for the emergence of higher complicated dynamic function such as "thinking", our group has set up a hypothesis that internal chaotic dynamics in an agent's chaotic neural network grows from "exploration" to "thinking" through reinforcement learning, and proposed a new learning method for that. However, even after learning in a simple obstacle avoidance task, the agent sometimes moved irregularly and collided with the obstacle. By reducing the scale of the recurrent connection weights, which is expected to have a deep relation to the chaotic property, the problem was reduced. Then in this paper, the learning performance depending on the recurrent weight scale is observed. The scale has an appropriate value as can be seen in FORCE learning in reservoir computing.

1　序論

自律的な機能創発に基づく汎用人工知能 (Artificial General Intelligence: AGI) の実現を目指すために、我々の研究室ではセンサからモータまでを柔軟に学習できるニューラルネット（NN）で構成し、強化学習を行うことで様々な機能を創発する “End-to-End 強化学習” のアプローチを提唱してきた [1, 2]。近年では DeepMind が TV ゲームや囲碁などで驚くべき結果を示しており [3, 4]、このことは我々のアプローチの妥当性を強く支持している。

また、高次機能獲得の観点から、センサからモータまでを静的にマッピングするだけでなく、ダイナミクスを扱えるリカレント NN に強化学習を通して、記憶が必要な機能が創発され、さらに外部入力による内部状態の遷移もある程度学習できることを示した [5, 6]。しかし、典型的な高次機能である「思考」は自律的、そして合理的に内部状態を遷移する必要があるため、実現は困難であった。

一方、強化学習に必要な「探索」は単なるランダムな行動選択のように見えるが、自律的に状態遷移するダイナミクスという点で「思考」に似ている。例えば我々は分かれ道では道路の状況や交通標識など様々なことを考えながら探索を行なっており、「探索」と「思考」は切り離して考えられないことがわかる。このことから我々は、カオス NN（ChNN）の自律的に状態遷移をしていく内部ダイナミクスの中に学習を通して合理的に遷移するアトラクタを形成することにより、よりランダムに近い「探索」からより合理的な「思考」へ成長するという仮説を立てている。この中で「閃き」や「発見」は連想記憶で見られるカオス的遍歴 [7] に近い現象として発現し、またアトラクタが形成されていない未知の状況では探索的な行動を自律的に再開する。このようにカオスダイナミクスによる発散と学習によって形成される収束のダイナミクスの共存によって「思考」が創発されると期待できる。

以上のダイナミクスを実現するために、探索として外部から乱数付加せずに ChNN を学習させる新しい強化学習を提案し、ゴール到達タスクを学習できることを確認した [2, 8]。またエージェントに視覚センサ入力、車輪出力を備えて障害物回避タスクを行わせ、障害物を回避しながらゴール到達できることも確認した [9]。しかしながら学習を通しても、エージェントが突然不自然な行動を起こしたり、障害物に衝突してそのまま動けなくなることが何度も現れていた。この問題を解決するために、本論文では ChNN のカオス性に深く関わっていると思われるリカレント結合重み値のスケールを変更し、学習パフォーマンスを確認する。

2　カオスニューラルネットを用いた強化学習

強化学習は自律的により多くの報酬を得るための行動を学習することができる。本論文では連続的な動作を扱うために Actor-Critic を用いた。Fig.1 に示すように、一般的な 3 層の階層型 NN を Critic 部として、カオス NN（ChNN）を Actor 部として用いる。センサからの入力 $\mathbf{S}_t$ を両ネットに送り、Actor ChNN の出力 $\mathbf{A}_t$ を動作信号、Critic NN の出力 V_t を状態評価値としている。

両ネットワークのニューロンモデルはここでは静的ニューロンであり、次式で表される。

$$u^l_{j,t} = \sum_{i=1}^{N^{l-1}} w^l_{j,i} x^{l-1}_{i,t} \left(+ \sum_{i=1}^{N^l} w^{FB}_{j,i} x^l_{i,t-1} \right) \quad (1)$$

ここで、$u^l_{j,t}$, $x^l_{j,t}$ はそれぞれ時刻 t での l 層 j 番目ニューロンの内部状態と出力を表し、$w^l_{j,i}$ は $(l-1)$ 層 i 番目ニューロンから l 層 j 番目ニューロンへの結合重み値、N^l は l 層のニューロン数を表す。右辺第 2 項は ChNN の中間層のみに用いられ、$w^{FB}_{j,i}$ は中間層の i 番目ニューロ

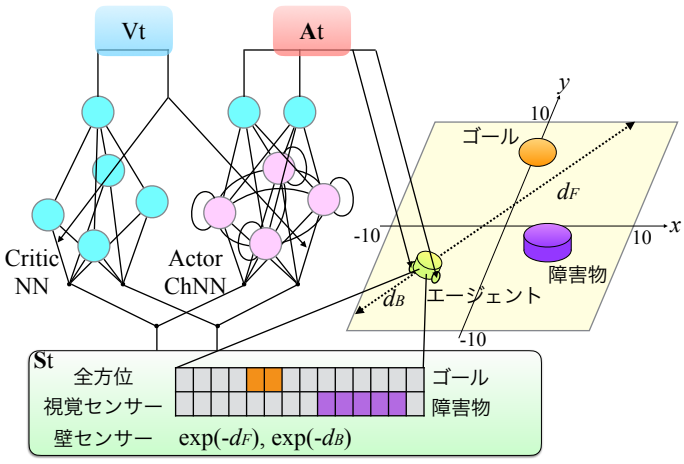


Fig. 1　障害物回避タスク

ンから j 番目ニューロンへのリカレント結合重み値である。出力 $x^l_{j,t}$ は値域 [-0.5,0.5] のシグモイド関数 $f()$ を用いて、$x^l_{j,t} = f(u^l_{j,t})$ としている。

学習で用いる TD 誤差 $\hat{r}_t$ と Critic 部の教師信号 T_{V_t} は次式で表される。

$$\hat{r}_t = r_{t+1} + \gamma V_{t+1} - V_t \quad (2)$$

$$T_{V_t} = V_t + \hat{r}_t = r_{t+1} + \gamma V_{t+1} \quad (3)$$

ここで r_{t+1} は時刻 $t+1$ で与えられる報酬、γ は減衰率を表す。Critic NN は誤差逆伝搬法を用いて学習する。

提案手法では Actor 出力に探索成分として外部から乱数を加えていない。ChNN の重み値 $w^l_{j,i}$ のみ、過去の入力を保持する因果トレース $c^l_{j,i,t}$ と学習係数 η を用いて次式で更新される。

$$\Delta w^l_{j,i,t} = \eta \hat{r}_t c^l_{j,i,t} \quad (4)$$

$$c^l_{j,i,t} = \left(1 - |\Delta x^l_{j,t}|\right) c^l_{j,i,t-1} + \Delta x^l_{j,t} x^{l-1}_{i,t} \quad (5)$$

ここで $\Delta w^l_{j,i,t}$ は時刻 t での $w^l_{j,i}$ の更新量を表し、因果トレース $c^l_{j,i,t}$ は出力の変化量 $\Delta x^l_{j,t} = x^l_{j,t} - x^l_{j,t-1}$ に基づいて入力の取り込みと保持の割合が変化する。

3　シミュレーション

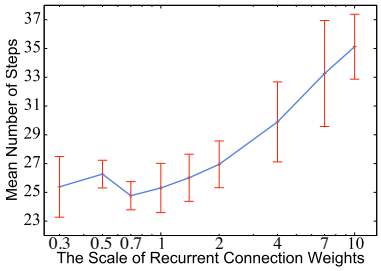
本論文では Fig.1 に示す障害物回避タスクを用いる。半径 1.0 のゴールは (0,5) の位置に固定し、半径 0.5 のエージェントと半径 1.5 の障害物は毎試行ランダムに配置する。5° ずつの 72 セルの全方位視覚センサー 2 つがそれぞれゴールと障害物の位置を取得し、エージェントの正面前後の壁との距離信号と合わせ、計 146 の信号が両 NN の入力 ($= \mathbf{S}_t$) として与えられる。エージェントの左右車輪は Actor 部の出力 ($= \mathbf{A}_t$) によって回転する。ゴールエリアに到達すると 0.4 の報酬を、障害物または壁に衝突すると-0.1 の罰を与える。ゴールするか 1,000 ステップ以内にゴールできない場合、その試行を終了する。シ

Table 1　シミュレーションに用いたパラメータ

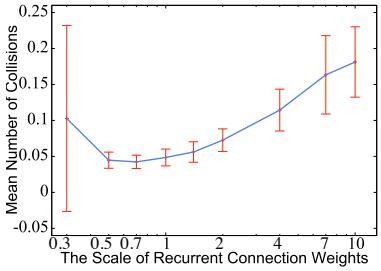
名前		Actor	Critic
試行回数		1,000,000	
中間層ニューロン数		100	
シグモイド関数の	出力層	1	
ゲイン: g	中間層	2	1
学習係数: η		0.001	1.0
初期重み値 $w^l_{j,i}$ の値域		[-1,1]	
減衰率: γ		—	0.99

ミュレーションに用いたパラメータを Table 1 に示す。

$\mathbf{w}^{FB}$ のスケール $|\mathbf{w}^{FB}|$ を 0.3 から 10 までの 8 ケース変化させ、各ケースで初期重み値を 10 パターン分学習させる。学習最後の 100,000 試行のゴールまでのステップ数と、障害物や壁との衝突回数の平均と標準偏差を Fig.2 に示す。スケールを小さくするとゴールまでのステップ数と衝突回数は減少し、スケールが 0.7 のときに最小となる。また、スケールが 0.1 の時はエージェントは探索を行えず、最終的には動けなくなることが確認された。



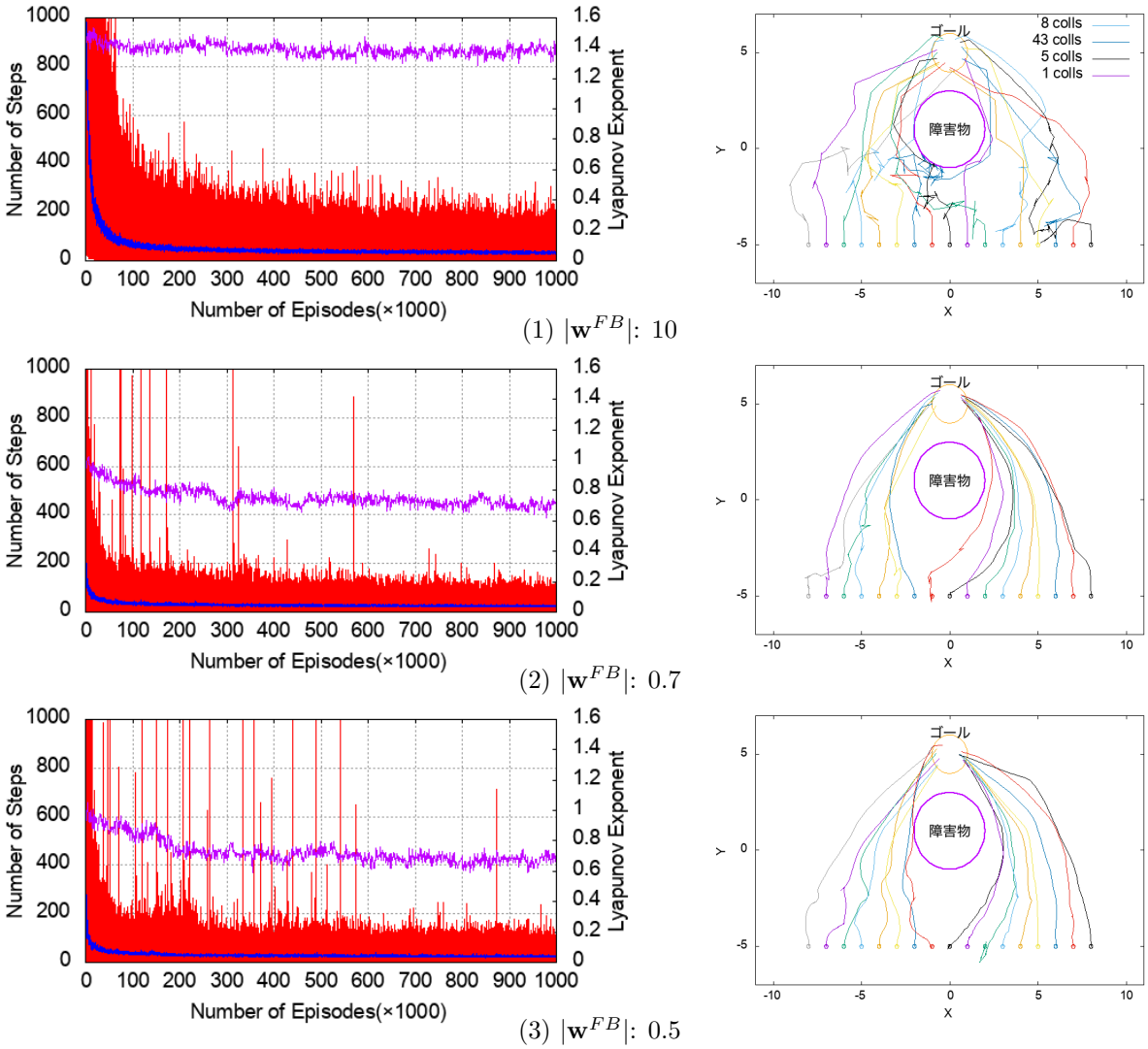
(a) ゴールまでのステップ数



(b) 障害物、または壁との衝突回数

Fig. 2　リカレント重み値のスケール $|\mathbf{w}^{FB}|$ による学習パフォーマンスの違い（青：平均、赤：標準偏差）

このうち、スケールが 10、0.7、0.5 の場合の結果を比較する。Fig.3 に (a) が学習曲線と擬似リアプノフ指数、(b) がエージェントの軌道を示す。(a) で、赤線と青線がそれぞれ、スタートからゴールするまでのステップ数とその 100 試行毎の平均、紫線は 1,000 試行毎の擬似リアプノフ指数を表す。擬似リアプノフ指数は正であればダイナミクスがカオス的である指標となる。本論文では ChNN の中間層ニューロンの内部状態に 0.001 に正規化した乱



(a) 学習曲線（赤、青：平均）
擬似リアプノフ指数の変化（紫）

(b) 学習後のエージェントの軌道
($x = -8, -7, \cdots, 8, y = -5$)

Fig. 3 $|\mathbf{w}^{FB}|$ 3 ケースにおける学習パフォーマンスの違い

数ベクトルを摂動として加え、摂動ありとなしそれぞれの場合で $\mathbf{A}_t$ に基づいて 5 ステップ回し、それぞれの中間層の内部状態のユークリッド距離 d を用いて、擬似リアプノフ指数 λ を次式で計算する。

$$\lambda = \frac{1}{P}\sum_{p=1}^{P}\frac{1}{T}\sum_{t=1}^{T} ln\frac{d_t^{(p)}}{d_{t-1}^{(p)}} \tag{6}$$

ここでエージェントを Fig.3 (b) の 17 パターンと、障害物を x= -5, 0, 5, y=1 の 3 パターンの計 51（$= P$）パターンで計算した。Fig.3(a-1) ではゴールまでのステップ数が他の 2 ケースより多く、(b-1) ではエージェントが不自然に行動し、障害物に衝突し続けている。(a-1) の擬似リアプノフ指数は約 1.4 であり、他の 2 ケースと比べて大きい。このことから $|\mathbf{w}^{FB}|$ が大きいとカオス性が強すぎて、学習によってそれを抑えられないことが考えられる。スケールが 0.7 である (b-2) では、エージェントが滑らかに動いてゴールしていることがわかる。さらにスケールが小さい 0.5 の (a-3) では、ゴールするまでに衝突することが多い試行の頻度が高い。

エージェントの初期位置による行動の分布を Fig.4 に示す。(a) は障害物を左、右に避けてゴールした初期位置をそれぞれ青、赤でプロットしている。(b) はゴールするまでに障害物に衝突した回数が 0 回、1 ～ 5 回、6 ～ 10 回、10 回以上のときの初期位置をそれぞれ青、緑、紫、赤でプロットしている。(a-1) は 2 領域の境界が障害物正面に現れているように見えるが、(b-1) から衝突しやすいことがわかる。(b-2) よりスケールが 0.7 の時は衝突する初期位置が最も少なくなる。

4 結論

今回 ChNN を用いた強化学習を障害物回避タスクに適用し、リカレント結合重み値のスケールを変化させて、

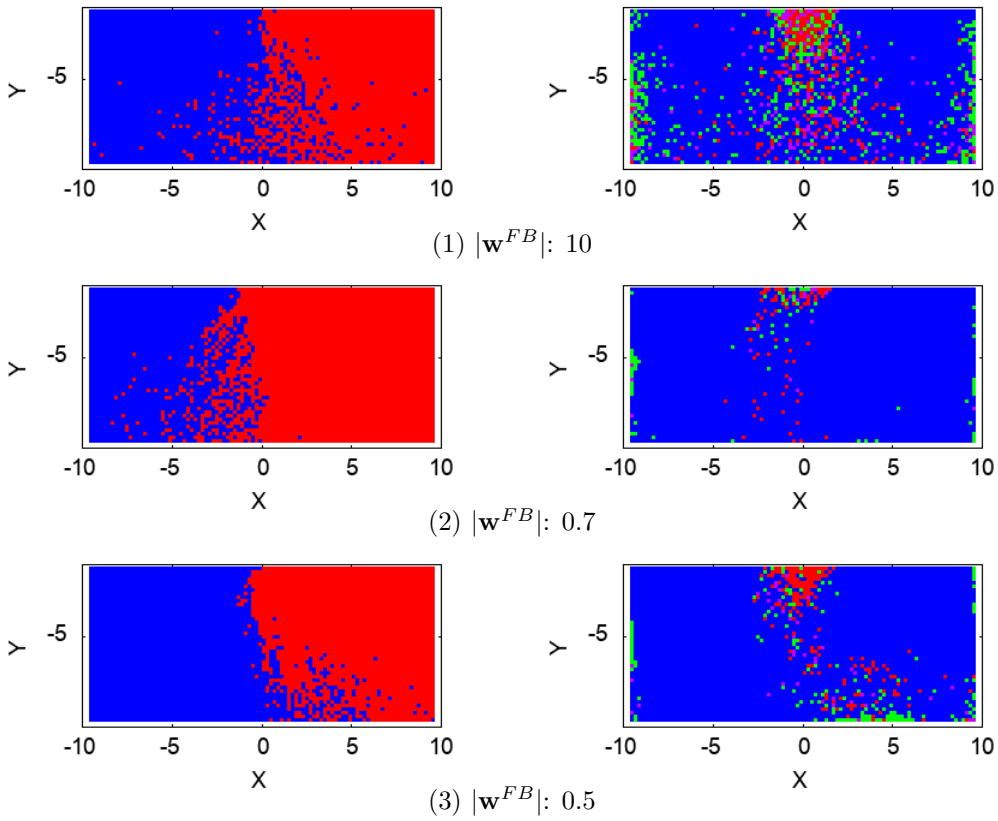


Fig. 4　$|\mathbf{w}^{FB}|$ 3 ケースにおけるエージェントの初期位置毎の行動の違い

学習パフォーマンスがどのように変化するか観察した。スケールが大きいほど、カオス性が強くエージェントが不自然に動き障害物に衝突する頻度が増え、小さくするとエージェントは滑らかに動くようになるが、小さくしすぎるとエージェントは十分に探索しなくなり、最後には学習できなくなる。リカレント結合重み値のスケールを適切な値にすることは重要である。一方、同じリカレント NN であるリザバネットで複雑なダイナミクスが簡単に学習できる FORCE learning での、リカレント結合重み値のスケールが大きいとカオスを抑えることができず、小さいとカオス性が失われる結果 [10] と類似しており、ChNN のカオス性が学習パフォーマンスに深く関係していることを示唆している。しかし今回一番良い結果（$|\mathbf{w}^{FB}| = 0.7$）でさえ、エージェントは障害物に何回か衝突していたので、更なる改善が必要である。

謝辞

本論文は JSPS 科研費 (15K00360) の補助を受けた。

参考文献

1) 柴田克成: 強化学習とニューラルネットによる知能創発, 計測と制御, Vol. 48, No. 1, pp.106-111 (2009)

2) 柴田克成, 後藤祐樹: 深層学習が示唆する end-to-end 強化学習に基づく機能創発アプローチの重要性と思考の創発に向けたカオスニューラルネットを用いた新しい強化学習, 認知科学, Vol. 24, No. 1, pp.96-117 (2017)

3) V. Mnih, et al.: Playing Atari with Deep Reinforcement Learning, NIPS Deep Learning Workshop 2013 (2013)

4) D. Silver, et al.: Mastering the game of Go with deep neural networks and tree search, Nature 529, pp.484-489 (2016)

5) K. Shibata, H. Utsunomiya: Discovery of Pattern Meaning ..., Proc. of IJCNN 2011, pp. 1445-1452 (2011)

6) K. Shibata, K. Goto: Emergence of Flexible Prediction-Based Discrete Decision ..., Proc. of ICDL-Epirob 2013, ID 15 (2013)

7) K. Kaneko, I. Tsuda: Chaotic itinerancy, Chaos, 13(3), pp.926-936 (2003)

8) K. Shibata, Y. Sakashita: Reinforcement Learning with Internal-Dynamics-based Exploration Using a Chaotic Neural Network, Proc. of IJCNN 2015, #15231 (2015)

9) K. Shibata, Y. Goto: New Reinforcement Learning Using a Chaotic Neural Network for Emergence of "Thinking" – "Exploration" Grows into "Thinking" through Learning –, arXiv:1705.05551(RLDM17)

10) D. C. Sussillo.: Learning in Chaotic Recurrent Neural Networks, Ph.D.Thesis, Columbia University (2009)