カオスニューラルネットを用いた強化学習における不応性を有するカオスニューロンの導入

大分大学　○佐藤　克樹，後藤　祐樹, 柴田　克成

Introduction of Chaotic Neurons with Refractoriness in Reinforcement Learning using a Chaotic Neural Network

Oita University　○ Katsuki SATO, Yuki GOTO, Katsunari SHIBATA

Abstract : Aiming for the emergence of 'logical thinking', we have proposed new reinforcement learning using a chaotic neural network, and confirmed that an agent having a chaotic neural network could learn a simple goal reaching task or obstacle avoidance task through reinforcement learning. In this paper, chaotic neurons without refractoriness are introduced in the chaotic neural network and it is shown that the network can learn a simple goal reaching task through reinforcement learning. It is also shown that the network can learn with smaller range of the feedback connection weights than the case without refractoriness. It is observed that by introducing chaotic neurons, the agent behavior becomes more explorative even with the same feedback weight range and that seems to relate deeply to the chaotic property indicated by the Lyapunov exponent.

1. 序論

ロボットの活躍の場は生産工場から我々の生活の場へと広がりロボットにより柔軟な賢さが求められている。本研究室ではシステムの柔軟性や自律性が失われるのを避けるために設計者の介入を極力少なくし、入力のセンサから出力のモータまでのシステムを並列で柔軟に学習できるニューラルネットで構成し、試行錯誤によって自律的学習が可能な強化学習を用いることで、ニューラルネット内に様々な機能を自律的に獲得することを提案してきた [1]。近年では中間層を多数重ねた深層学習 (Deep Learning) の強力な抽象化能力が認識の分野で優れた力を発揮しており [2]、さらには Deep Learning に強化学習を組み合わせた自律的学習によって、人間に TV ゲームで勝った [3] ことなどは、ニューラルネットを強化学習で自律的学習をさせる我々のアプローチの有効性を示唆している。

記憶や予測、思考などの高次機能は時間軸を扱う必要があるため、本研究室ではリカレントニューラルネットを導入し、強化学習を通すことで、簡単なタスクで記憶や予測の機能を獲得できることを確認した [4][5]。しかし思考と呼べるような機能の創発には今日まで至っていない。

思考は外部からの入力がなくとも色々なことを考えたり想像できたりと、自律的に状態を遷移する内部ダイナミクスのように見える。強化学習では探索が必要であるが、この探索も学習者自身の自発的なダイナミクスに由来すると考えることができる。この両者の類似性から、我々は探索の内部ダイナミクスが学習することで思考へと成長するのではないかと考えた。この仮説の実現に向けて、カオスニューラルネットを用いることで内部のカオスダイナミクスによって外部からの乱数付加なしでの探索を可能にし、それを学習させるための新しい強化学習を提案した。そして、この学習方法で、簡単なゴール到達タスクや障害物回避タスクの学習まではできることを示してきた [6][7]。

本研究室では従来中間層フィードバック (FB) 結合部を強くすることでカオスダイナミクスを生成し、学習に用いてきた。一方、各ニューロンに不応性を導入したカオスニューロンを用いて、ニューラルネット内にカオスダイナミクスを生成する方法がある [8]。不応性とは一度発火したニューロンがしばらくの間発火しにくくなるになる性質のことであり、生体ニューロンが実際に持っている性質でもある。

そこで本論文では、不応性を有するカオスニューロンを用いたカオスニューラルネットを用いて、提案した強化学習を行うことができるのかを簡単なゴール到達タスクの学習で確認する。また、その際中間層 FB 部の重み値を変化させながら、不応性を導入していないネットワークとの学習の様子を比較する。

2. カオスニューラルネットを用いた強化学習

強化学習によってエージェントは報酬や罰により動作を自律的に学習することができる。よりよい動作を学習するためには探索を行う必要があり、一般的には外部から乱数を与えて確率的な行動 (探索) を行う。提案している新しい強化学習では乱数付与はせず、ニューラルネット内部のカオスダイナミクスに基づいて探索を行う。また、モータレベルの動作の学習を考えて連続行動を扱える Actor-Critic を用いる。動作生成部である Actor 部をカオスニューラルネットで構成し、状態評価部である Critic 部はカオスの発生しない通常の階層型ニューラルネットを用いる。本論文で使う不応性を有するカオスニューロンを導入したカオスニューラルネットの式を (1)-(5) 式に示す。カオスニューロンモデルは [8] の合原モデルをベースにしているが、ここでは時定数が明確になる表記で表した。時刻 t での中間層 (h) の j 番目ニューロンの内部状態 $u^h_{j,t}$ は (1)-(4) 式で

求められ、(5) 式のシグモイド関数を通し中間層出力 $o^h_{j,t}$ を求める。ここで N^l は l 層のニューロンの数、$w_{j,i}$ は j 番目ニューロンの i 番目の結合重み値であり FW はフォワード部、FB はフィードバック部である。

$$u^{\xi}_{j,t} = \left(1-\frac{\Delta t}{\tau}\right)u^{\xi}_{j,t-\Delta t} + \frac{\Delta t}{\tau}\sum_{i=1}^{N^{in}} w^{FW}_{j,i}\ o^{in}_{i,t} \quad (1)$$

$$u^{\eta}_{j,t} = \left(1-\frac{\Delta t}{\tau}\right)u^{\eta}_{j,t-\Delta t} + \frac{\Delta t}{\tau}\sum_{i=1}^{N^{h}} w^{FB}_{j,i}\ o^{h}_{i,t-\Delta t} \quad (2)$$

$$u^{\zeta}_{j,t} = \left(1-\frac{\Delta t}{\kappa\tau}\right)u^{\zeta}_{j,t-\Delta t} - \frac{\Delta t}{\kappa\tau}\ \alpha\ o^{h}_{j,t-\Delta t} \quad (3)$$

$$u^{h}_{j,t} = u^{\xi}_{j,t} + u^{\eta}_{j,t} + u^{\zeta}_{j,t} - \theta \quad (4)$$

$$o^{h}_{j,t} = \frac{1}{1+exp(-g\cdot u^{h}_{j,t})} \quad (5)$$

式 (1) はフォワード項、(2) はフィードバック項、式 (3) は不応項をそれぞれ表しており、式 (4) のように、この 3 つの項の和がカオスニューロンの内部状態となる。(5) 式のシグモイド関数により、内部状態から出力を求める。[8] から $\Delta t = 1$、$\tau = 1.25$、$\kappa = 8$ とした。α、θ、g はそれぞれ不応性のスケーリングパラメータ、しきい値、ゲインでありここでは 3, 0, 2 とした。また、出力層及び Critic 部の計算式を載せる。出力層及び Critic 部は静的なニューロンを用いている。

$$u^{l}_{j,t} = \sum_{i=1}^{N^{(l-1)}} w^{l}_{j,i}\cdot o^{l-1}_{i,t} \quad (6)$$

$$o^{l}_{j,t} = \frac{1}{1+exp(-u^{l}_{j,t})} - 0.5 \quad (7)$$

カオスニューラルネットは現在の状態ベクトル $\boldsymbol{S}_t$ を入力として行動ベクトル A_t を、通常のニューラルネットは状態評価値である Critic 出力 V_t をそれぞれ出力する。学習に用いる TD 誤差 $\hat{r}_t$ は以下のように求め、通常の誤差逆伝搬 (BP) 法で学習する。

$$\hat{r_t} = r_{t+\Delta t} + \gamma\cdot V_{t+\Delta t} - V_t \quad (8)$$

$r_{t+\Delta t}$ は実際の動作後に得られた報酬、γ は割引率であり 0.96 とした。TD 誤差を用いて Critic 部の教師信号 T_V は以下のようにして求める。

$$T_V = r_{t+\Delta t} + \gamma\cdot V_{t+\Delta t} \quad (9)$$

カオスニューラルネットは因果トレース [9](ただし、[9] とは使い方が少し異なる) を用いて重み値の更新量 $\Delta w^{l}_{j,i,t}$ を決定する。

$$\Delta w^{l}_{j,i,t} = \eta\cdot c^{l}_{j,i,t}\cdot \hat{r_t} \quad (10)$$

因果トレース $c^{l}_{j,i,t}$ は各ニューロンの各結合部に配置し、ニューロンの出力の増加に寄与した過去の入力 $o^{l-1}_{i,t}$ を保持するようにニューロンの出力の変化 $\Delta o^{l}_{j,t} = o^{l}_{j,t} - o^{l}_{j,t-1}$ を用いて以下の式で計算する。

$$c^{l}_{j,i,t} = (1-|\Delta o^{l}_{j,t}|)c^{l}_{j,i,t-1} + \Delta o^{l}_{j,t} o^{l-1}_{i,t} \quad (11)$$

η は学習係数である。本論文ではカオス性を維持するために中間層 FB 部の学習は行わない。

3. シミュレーション

本論文では不応性を有するカオスニューラルネットの学習が可能か確認するために、Fig.1 のように中心座標を (0,0) とした 10 × 10 のフィールド内にゴールを設置したゴール到達タスクを行う。また、比較対象のネットワークとして (4) 式から不応性の項 $u^{\zeta}_{j,t}$ を除いたものを用いた。

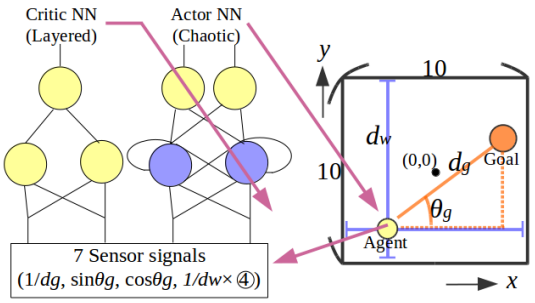


Fig. 1: ゴール到達タスクとネットワークシステム

シミュレーションに用いたパラメータを Table 1 に示す。

Table 1: シミュレーションに用いたパラメータ

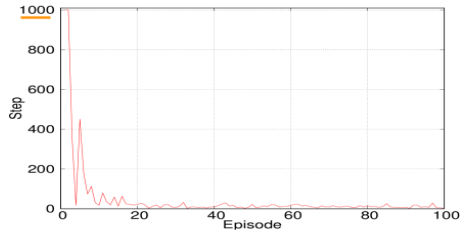
名前		Actor	Critic
中間層ニューロンの数		100	30
シグモイド関数のゲイン	出力	1	
	中間	2	1
学習係数 η	出力	0.3	10
	中間 (FW)	0.3	10
	中間 (FB)	0	-
重み値の値域	中間 (FW)	[-1,1]	
	中間 (FB)	[-1,1]	-
	Other	[-1,1]	

Fig. 1 に示した 7 個の入力情報 (エージェントとゴールとの相対距離、相対角度、各壁との相対距離) をそれぞれ最大値が 1 となるように正規化してネットワークの入力とする。2 つの Actor 出力はそれぞれ x 方向、y 方向の移動量を表し、移動可能範囲が半径 0.5 の円になるように移動方向は変えずに大きさを調整した値にしたがってエージェントが移動する。

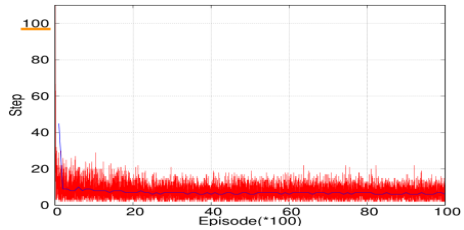
半径 1.0 のゴールと半径 0.5 のエージェントの初期位置は毎試行ランダムに決定する。エージェントはカオスニューラルネットの Actor 出力で動き、ゴールに到達すると 0.4 の報酬を獲得する。また、壁に衝突すると-0.01 の罰が与えられる。エージェントがゴールに到達する、またはステップ上限である 1,000 ステップ経過するまでを 1 試行とし 10,000 試行の学習を行った。

Fig. 2 に学習曲線として、学習初期の様子がわかるために (a)100 試行目までを拡大したものと (b) 全体の学習の様子がわかるために縦軸を拡大した 10,000 試行までの 2 つ示す。学習曲線の赤い線は各試行のエージェントがスタートからゴールするまでのステップ数を表し、青い線は 100 試行毎の平均値である。また、(c) に学習終了 (10,000 試行) 後にゴールの位置を (0, 0) の位置に固定し、エージェントのスタート位置をずらして設置した 8 パターン分の軌道と (d) その時の Critic(状態評価) 値の変化を Fig.2 に示す。

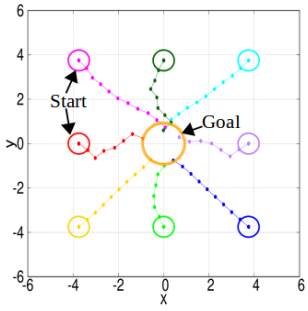
試行回数が増えるとゴールまでのステップ数が下がっている。学習後のエージェントの軌道はゴールに向かっている様子がわかり、その時の Critic 値はエージェントがゴールに近づくと高くなっている。上記のことから、不応性を有するカオスニューロンを用いても新しい強化学習で簡単なタスクの学習ができることがわかった。



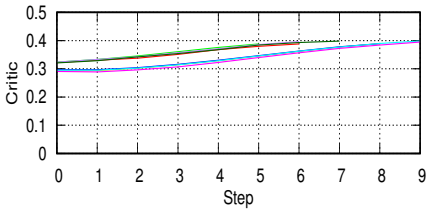
(a) 学習曲線 (100 試行まで、時間軸を拡大)



(b) 学習曲線 (縦軸を拡大)



(c) 学習後の軌道 (8 パターン)



(d) 学習後の Critic 値の変化 (8 パターン)

Fig. 2: 不応性を有するカオスニューロンを導入したカオスニューラルネットを用いた時の学習の様子
(FB 部の重み値-1〜1)
(a,b) 学習曲線 (ゴールまでのステップ数)
(c) 学習後の軌道と (d) その時の Critic 値の変化

次に不応性を有するカオスニューロンで構成されたニューラルネットと不応性がない (式 (3) の不応項がない) カオスニューラルネットの学習成功率を比較する。2 つのネットワークの FB 部の重み値の範囲 $[-w_{max}^{FB} : w_{max}^{FB}]$ の w_{max}^{FB} を 0.3 から 2 までの 9 通り変化させ、それぞれ乱数系列を変えて 20 回学習させて学習成功回数を比較した。その結果を Fig. 3 に示す。学習終了後の 8 パターンの軌道全てが 15 ステップ以内ゴールしていれば成功とした。

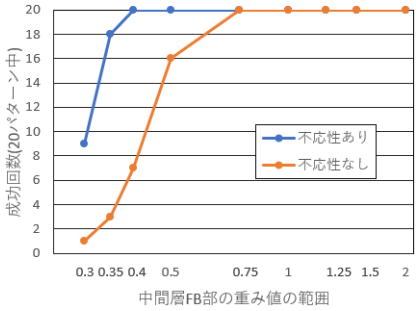


Fig. 3: 中間層 FB 部の重み値の範囲 w_{max}^{FB} による学習成功率の不応性の、有無による比較

不応性の有無にかかわらず FB 部の重み値を小さくしていくと学習ができなくなるが、不応性なしの方が早く落ちることがわかる。その原因を探るため、学習初期の探索の様子を Fig. 4、Fig. 5 に示す。

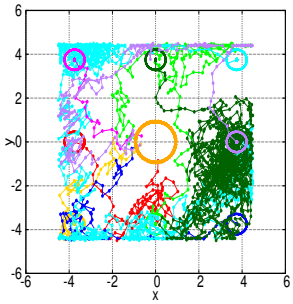


Fig. 4: 学習前の探索の様子
(不応性あり、FB 部の重み値-1〜1)

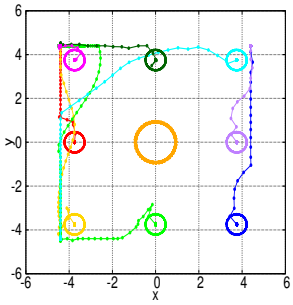


Fig. 5: 学習前の探索の様子
(不応性なし、FB 部の重み値-1〜1)

不応性あり (Fig. 4) ではフィールド全体で探索を行っているのに対して、不応性なし (Fig. 5) では壁にぶつかり続け、うまく探索ができていないことがわかる。ただし、学習を進めると壁衝突時の罰によって壁から抜け出して探索するようになり、最終的にはゴールに辿り着く。しかし、FB 部の重み値をさらに小さくしていくと、壁から罰を受けてもフィールド内の探索ができなくなる。不応性がある場合でも FB 部の重み値をさらに小さくしていくと同様に探索ができなくなる。Fig. 3 より、同じ重み値でも学習成功率に差があるのは不応性ありの方がカオス性が強く探索ができているからだと考えられる。また、FB 部の重み値の範囲が小さくなるとカオス性が弱まり、うまく探索できなくなると考えられる。

そこで、FB 部の重み値の大小および不応性の有無とカオス性との関係を見るために、これらを変化させながらカオス性の指標であるリアプノフ指数を観察した。リアプノフ指数とは、カオス性を調べる指標であり、微小な摂動の時間的な広がりを表す。リアプノフ指数が正となればカオス性があることを示す。本論文では入力を 0 にし中間層同士の FB 結合のみで 50 ステップネットワークを回した後に中間層ニューロンの内部状態に 0.001 の大きさに正規化された乱数を加え更に $T = 50$ ステップ回し、乱数を加えなかった場合と加えた場合の中間層ニューロンの内部状態を用いてリアプノフ指数 λ を求める。これを Fig. 3 で用いた乱数系列 $P = 20$ 個を用いて式 (12) のように計算し、平均と標準偏差を求める。

$$\lambda = \frac{1}{T \cdot P} \sum_{p=1}^{P} \sum_{t=1}^{T} \ln \frac{d_{p,t+\Delta t}}{d_{p,t}} \tag{12}$$

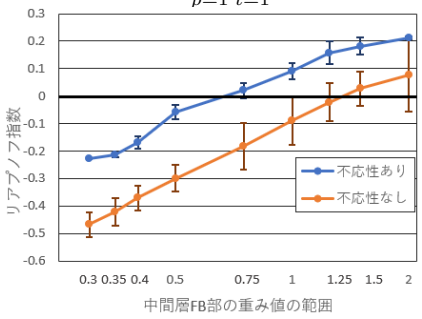


Fig. 6: 中間層 FB 部の重み値の範囲による
リアプノフ指数の、不応性の有無による比較

Fig. 6 より、全体的な傾向として不応性の有無にかかわらず、重み値を小さくしていくとカオス性が小さくなっていることがわかり、さらに、不応性なしの方が不応性ありに比べてカオス性を早く失っていることがわかる。以上より、カオスダイナミクスによって探索を行う強化学習では、カオス性が探索に大きく関わっていると言え、探索をうまくすることで学習成功率が上がると考えらる。不応性を導入することで同じ FB 部の重み値でもカオス性が上がり、FB 部の重み値が小さくなってもカオス性を維持できき探索をうまく行うことができたため、学習できたと考えられる。

4. 結論

不応性を有するカオスニューロンを導入したカオスニューラルネットを用いて強化学習によって簡単なゴール到達タスクを学習させることができた。不応性を除いたカオスニューラルネットと比較したところ、FB 部の重み値の範囲を小さくしていった場合、不応性ありの方が学習性能の下がりが遅いことを示した。また、リアプノフ指数を観察したところ、FB 部の重み値を小さくするとカオス性が小さくなるが、不応性ありの方が全体的にカオス性が大きいことも示した。今後の課題として、カオス性を揃えた上での学習性能の違いを調べる必要がある。さらに、カオスニューロンのパラメータによってどのようにカオス性が変化していくのかも調べる必要がある。

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：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, No1, pp.96-117 (2017)

[2] Y. Bengio : Learning deep architectures for ai. *Foundations and Trend in Machine Learning*, Vol.2, No.1, pp.1-127 (2009)

[3] V. Mnih, K. Kavukcuoghu, etc : Playing Atari with Deep Reinforcement Learning, *DeepMind Technologies* (2015)

[4] K. Shibata and H. Utsunomiya : Discovery of Pattern Meaning from Delayed Rewards by Reinforcement Learning with a Recurrent Neural Network, Proc. of IJCNN. 2011, pp. 1445-1452,(2011)

[5] K. Shibata and K. Goto : Emergence of Flexible Prediction-Based Discrete Decision Making and Continuous Motion Generation through Actor-Q-Learning, Proc. of Int'l Conf.ID 15 (2013)

[6] K. Shibata and Y. Sakashita : Reinforcement Learning with Internal-Dynamics-based Exploration Using a Chaotic Neural Network, Proc. of Int'l Joint Conf. (2015)

[7] Y. Goto and K. Shibata : Emergence of Higher Exploration in Reinforcement Learning Using a Chaotic Neural Network, Proc. of ICONIP, pp. 40-48, (2016)

[8] G. Matsumoto, K. Aihara, M. Ichikawa and A. Tasaki : Periodic and Nonperiodic Responses of Membrane Potentials in Squad Giant Axons During Sinusoidal Current Stimulation, Journal of Theoretical Neurobiology, Vol.3, No.1, pp.1-14 (1984)

[9] 柴田克成：因果トレース-並列かつ主観的時間スケールの導入による過去の処理の効率的学習, NC2013-115, 電子情報通信学会技術報告, pp157-162 (2014)