

FIT 論文

Hessian-Free 法を用いた Echo State Network の構造探索

濱地 優輝[†] 平川 翼[†] 山下 隆義[†] 藤吉 弘亘[†] 山口 裕[‡] 津田 一郎[†]
 Yuuki Hamaji Tsubasa Hirakawa Takayoshi Yamashita Hironobu Fujiyoshi Yutaka Yamaguti Ichiro Tuda

1. はじめに

脳の特徴の一つである機能分化は、脳の最小単位であるニューロンが、目からの視覚信号や耳からの音声信号などの特定の入力情報に対してのみ反応を示すようになる現象である。機能分化の原理を明らかにすることは、脳機能の理解だけでなく、柔軟で適応性を持つ人工ニューラルネットワークのモデル構造やアルゴリズムの提案にもつながると考えられる。これまでに、機能分化を人工ニューラルネットワークを用いて再現し、その計算原理を明らかにする試みが行われている [1]。機能分化を再現するためには脳を模したモデルが必要となるが、生物学的な構造に基づく人工ニューラルネットワークモデルとしてリザーバコンピュータが提案されている [2][3][4]。リザーバコンピュータは、ニューロン間の接続関係を示す接続構造と、ニューロン間の接続の強さを示す接続重みを持つモデルである。この接続構造と接続重みはランダム接続で、かつ疎な接続構造からなる。このことから、脳内の微小回路と同様のネットワーク構造を持ち、小脳皮質 [5] と新皮質 [6] の計算モデルの一つとされている。先行研究 [1] では、リザーバコンピュータの一種である、Leaky Echo State Network (LeakyESN)[4]を対象とし、遺伝的アルゴリズム [7] を用いて、ネットワーク構造と重みを探索する事で、機能分化を誘発している。誘発の結果、特定の時間または空間パターンに反応するニューロンが発生する事が報告されている [1]。しかし、重みは連続値である為、先行研究による最適な重みの探索には膨大な時間が必要になる。

そこで本研究では、Echo State Network (ESN) [3] を対象とし、Hessian-Free 法と遺伝的アルゴリズムを組み合わせた効率的な探索法を提案する。Hessian-Free 法と従来の遺伝的アルゴリズムの探索を組み合わせる事で、一個体の精度を向上し、学習の安定性の向上が期待できる。

2. Echo State Network

Echo State Network の構造と遺伝的アルゴリズムを用いた学習手法について説明する。

2.1 Echo State Network と機能分化

Echo State Network(ESN) [3]は、図1に示すように入力層、Reservoir 層、出力層の3層で構成され、出力層のみを学習する。入力層と Reservoir 層の重みは0から1の間でランダムに決定する。Reservoir 層の次時刻における出力値 x_{t+1} は、式(1)で表す事ができる。 $f(\cdot)$ は活性化関数であり、ESNでは主に \tanh 関数が使われる。 \tanh 関数は、出力を-1から1の間に変換する活性化関数である。 \tanh 関数は出力値が

毎回正規化されるため Reservoir 層における発散の問題が起らない。出力層 z_t は、時刻を t 、重み行列を W_{out} としたとき式(2)により表される。 $g(\cdot)$ は活性化関数であり、ESNにおいては多クラス分類では Softmax 関数、波形予測のような分類問題では \tanh 関数を用いられる事が多い。Softmax 関数は、出力を0から1の間のベクトルに変換する活性化関数である。Yamaguti ら[1]は、ESNを図1のように Reservoir 層のニューロンを Input Neurons と Output Neurons に分けた拘束条件付き ESN を用いて空間・時間パターンの分類を行っている。Ridge 回帰による出力層の学習後、遺伝的アルゴリズムを用いて Reservoir 層の構造と重みを探索する。探索後の各ニューロンの相互情報量から、ネットワークの機能分化を確認している。

$$h_{t+1} = f(W_{in}x_t + W_{res}h_t) \quad (1)$$

$$z_t = f(W_{out}h_{t+1}) \quad (2)$$

2.2 Ridge 回帰

ESN では一般的には出力層の重みの学習を Ridge 回帰で行う。Ridge 回帰は重回帰分析を行う際の損失関数に対して正則化項を付与したものである。Ridge 回帰を式(3)に示す。 W_{out} は更新する重み行列、 X は Reservoir 層の出力値を学習したい分だけ縦方向に並べた行列、 α はハイパーパラメータ、 I は単位行列、 y は正解ラベルである。Ridge 回帰は最小二乗法の項と正則化の項で構成されている。そして、最小二乗法は、正解データ y_t と誤差が小さくなる重みを求めている。正則化項は過学習を防ぐためにL2ノルムの2乗を用いて重みが大きくなるようにしている。L2ノルムとは、複雑な推定モデルにペナルティを与えることで過学習を抑える手法の一つである。

$$W_{out} = (X^T X + \alpha I)^{-1} X^T y \quad (3)$$

2.3 遺伝的アルゴリズム

遺伝的アルゴリズムは、生物の進化を模倣した特定の計算問題に依存しない最適化アルゴリズムである。遺伝的アルゴリズムでは、対象とする問題の解を個体、個体内の変数を遺伝子とする。この個体を複数用意し、評価関数を元に適応度の高い個体を選択する。この選択した個体を元に交叉、突然変異などの操作が行われ、新しい個体が生成される。この一連の操作を世代とし、十分な世代数を行うことで問題に適した解を求める。Yamaguti ら[1]は、遺伝的アルゴリズムを用いる事で Reservoir 層の接続構造と重みの探

[†] 中部大学 Chubu University

[‡] 福岡工業大学 Fukuoka Institute of Technology

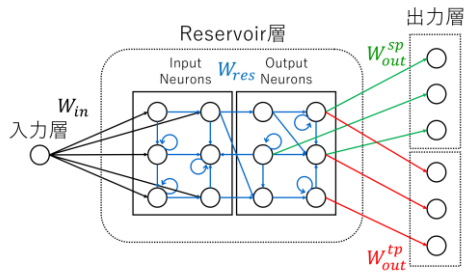


図 1 拘束条件付き ESN のネットワーク構造

索を同時に行った。しかし、全ニューロンの重みの組み合わせは膨大なため、遺伝的アルゴリズムによる最適な構造と重みの両者を十分に探索することは困難である。

3. 提案手法

本研究では、ニューロンの重み最適化と ESN の構造探索を分担し、段階的に行う手法を提案する。まず、全ニューロンの重みを Hessian-Free 法で簡単に学習し、その後、遺伝的アルゴリズムによる Reservoir 層の構造探索を行う。この提案手法では、Hessian-Free 法を用いる事で連続値である重みを効率的に探索する事ができるようになる。また、これを繰り返すことで、構造の探索と重みの探索が並行して探索され、それにより、遺伝的アルゴリズムによる構造探索では、重みに適した構造を探索することが可能となる。また、構造と重みの探索後、探索によって得られた Reservoir 層の接続情報と重みの探索空間を用いて全ニューロンの重みを Hessian-Free 法で十分に再学習する事で重みの最適化も可能となる。ここでの ESN の構造とはニューロン間の接続の有無であり、接続がない場合重みは 0 となり学習も行わない。接続がある場合は重みの探索を行う。

3.1 Hessian-Free 法を用いた ESN の学習

Hessian-Free 法は、再帰型ニューラルネットワークの学習における勾配消失・爆発問題を解決しつつ、長期的なデータの特徴抽出が可能な学習法である [8]。その名の通り、二階偏微分行列であるヘッセ行列を使用せずに、ニュートン法と同様の計算を行うのが特徴である。Hessian-Free 法はニューロンの重みの更新値が閾値以下の時は重みの探索を共役勾配法で行う。共役勾配法を用いた場合の重みの更新式を式 (4) に示す。

$$W_{k+1} = W_k - T_k M_k \quad (4)$$

ここで、 W_k は現在の重みの集合であり、 W_{k+1} は更新後の重みの集合である。式 (4) の T_k は接ベクトル、 M_k は共役勾配である。共役勾配法は深層学習に良く使われる勾配降下法のように接ベクトル T_k をそのまま計算に使用するのではなく共役勾配 M_k を計算し積を求める事で直線探索を可能にした手法である。一方で、更新値が閾値以上の時は重みの探索をガウスニュートン法で行う。ガウスニュートン法を用いた場合の重みの更新式を式 (5) に示す。ガウスニュートン法はヘッセ行列の近似を用いたニュートン法である。

$$W_{k+1} = W_k - (\nabla F_k^T \nabla F_k)^{-1} \nabla F_k \nabla F_k \quad (5)$$

Algorithm 1 提案手法のアルゴリズム

(個体数 N 最大世代数 G)

Require: 「現世代」, 「次世代」集合を用意

現世代に N 個体をランダムに生成

for 最大世代数 G 回繰り返す **do**

全ての個体で Hessian-Free 法を用いて
評価関数により、現世代の各個体の適応度を計算
全ての層の重み接続を学習(step1)

適応度の高い個体を生存させる

if ある確率に基づき動作を行う **then**

生存個体を 2 つ選択し交叉を行う
ある確率以上であれば個体を 1 つ選択し
突然変異を行う

end if

次世代の個体数 N 個になるまで

交叉と突然変異を繰り返す

(step2)

次世代の内容を現世代に移す (step3)

end for

現世代の中で生存した個体を「解」として出力

ここで、式 (5) の F_k は誤差の値、 ∇F_k は勾配ベクトルである。Hessian-Free 法をこのように共役勾配法とガウスニュートン法を状況によって切り替えながら最適値を求める手法である。

3.2 提案手法による構造の探索

Reservoir 層のニューロンの構造と重みの探索を並行して行うために、遺伝的アルゴリズムによる構造探索と Hessian-Free 法を用いた重みの探索を並行して行う。図 2 に提案手法による構造探索と重みの探索の流れを示す。まず、図 2 の step 1 にて、全ニューロンの重みを対象として Hessian-Free 法で構造探索の負担にならない程度に学習する。提案手法では、最高精度個体を再学習によって重みの最適化を行う事ができるので、遺伝的アルゴリズムの構造探索と並行して行う重みの探索時に重みを最適化する必要はない。図 2 における step 1 のモデルの紫色の矢印は学習の対象となる重みを示す。図 2 の step 2 では、遺伝的アルゴリズムを用いて一世代分、Reservoir 層のニューロンの接続構造を探索する。図 2 の step 2 でのオレンジ色の矢印は探索の対象となるモデルの接続構造を示す。図 2 の step 3 で、遺伝的アルゴリズムによって生成される個体の精度が収束するまで、step 1 と step 2 を繰り返す。これにより重みの探索と構造の探索を並行して行う事が可能となる。一連のアルゴリズムの流れを Algorithm 1 に示す。

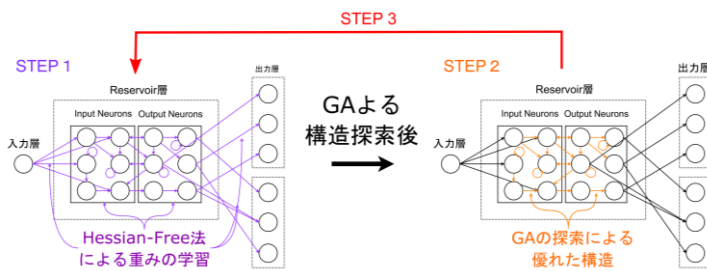


図2 ESNの構造探索の流れ

3.3 Hessian-Free 法による重みの再学習

構造探索後、探索によって得られた Reservoir 層の接続情報と重みの探索空間を用いて全ニューロンの重みを対象として Hessian-Free 法で再学習する。これにより提案手法では、探索によって得られた接続構造に対して、最適化した重みを獲得することが可能となる。従来手法では、構造探索と重みの最適化を同時に行っていた為、構造の変化と重みの変化が同時に起きてしまい、獲得した重みが構造の変化によって意味をなさなくなる場合が考えられていた。しかし、提案手法では、構造探索時でもこの再学習によって重みが最適化されるので、重みの探索時間を短縮しつつ、再学習によって重みの探索を最適化する事が可能となる。

4. 実験概要

Reservoir 層の構造と重みの探索について、従来手法と提案手法の比較実験を行う。実験には入力層 16 個、Reservoir 層に 32 個、出力層の 6 個のニューロンを持つ拘束条件付き ESN を用いて、タスク設定の分類を行う。提案手法における Hessian-Free 法の学習回数は 20 Epoch である。遺伝的アルゴリズムのパラメータは、個体数を 200、淘汰する数を 180、突然変異率を 25 パーセント、世代数を従来手法で 500 世代、提案手法で 5 世代とする。遺伝的アルゴリズムによる探索後、提案手法と従来手法で Reservoir 層のニューロンと正解ラベルから相互情報量を求め機能分化を確認する。

4.1 タスク設定

視覚パターンと聴覚パターンの同時入力を表すために、空間パターンと時間パターンの積を入力と見なし学習データを作成した。学習データの作成式を式(6)に示す

$$u(k, l) = ak\cos(2\pi flt) \quad (6)$$

ここで、空間情報と時間情報が変化する入力パターン $u(k, l)$ において変化する空間情報のパターンを k とした、 a_k の値は 1 か -1 が f_l に依存する周期で変化する、これは k の分割数によりどの程度変化するか決定される。この空間情報 k は分割数 2, 4, 8 の 3 パターンの中で変化する。同様に、時間情報を l とした、 f_l は 14, 18, 116 が決定される。この l が変化することで、時間情報のパターンが決定される。この数式によって作成されるデータセットを図 3 に示す。

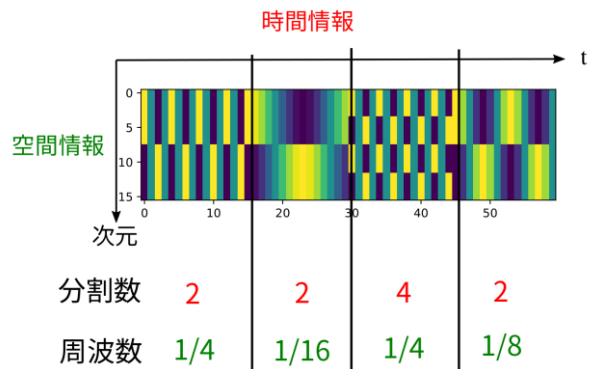


図3 入力されるデータセットのパターン例

4.2 相互情報量を用いた機能分化の傾向調査

Yamaguti ら[1]は、相互情報量を組み合わせることで機能分化を表現した。情報 X を Input Neurons と Output Neurons、情報 Y を時間情報と空間情報の正解ラベルと定義し、4 パターンの相互情報量を求めた。Input Neurons と Output Neurons の出力は最大値から最小値の間で 8 段階に離散化し、頻度分布を作成し、算出した。また、正解ラベルはラベル数で離散化し、頻度分布を作成し計算した。これらの定義を元に、 X を Input Neurons の出力を n_{in} 、 Y を空間情報の正解ラベル I_{sp} とした時の 1 パターンを式(7)に示す。

$$I(n_{in}; I_{sp}) = H(n_{in}) + H(I_{sp}) - H(n_{in}, I_{sp}) \quad (7)$$

求めた相互情報量の内、空間情報を x 軸と時間情報を y 軸とすることで、ニューロンがどちらの情報に寄与しているかが判別できる。よって、Input Neurons と Output Neurons が機能分化しているかを定性的に評価できる。

4.3 実験結果

図 4 と図 5 に従来手法と提案手法の精度の推移(移動平均)を示す。図 4 と図 5 の移動平均値のグラフから、提案手法で従来手法と比べて早い世代から高い精度の個体を生成出来ている。また、従来手法は時間情報における精度の向上が見られないが、提案手法では空間情報の精度の向上と共に時間情報の精度が向上している事が確認できる。さらに図 4 と図 5 の個体全体の分布から、提案手法は 1 世代における精度の分布幅が小さく、安定して Reservoir 層の重みと構造探索が可能である。表 1 に従来手法と提案手法の最高精度を示す。提案手法では、探索時における最高精度と再学習によって得られた最高精度を分けて表記した。再学習では提案手法の探索によって得られた最高精度個体の構造と重みの探索方向を用いて 200 Epoch 分の学習を行った。表 1 より時間情報と空間情報の両方で探索時の時点で提案手法の精度が高い。これにより、Hessian-Free 法を用いることで重みの探索が接続構造に対して効率的に行われ、遺伝的アルゴリズムによる構造探索も効果的な探索が可能である。

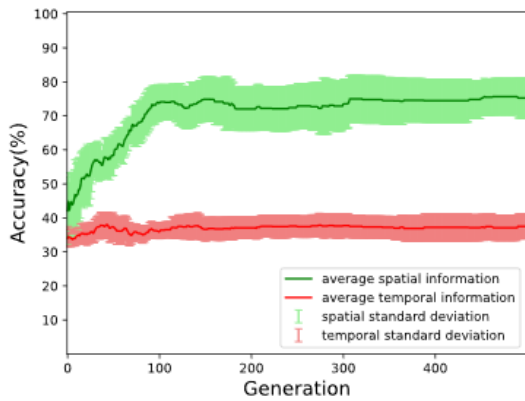


図4 従来手法の精度の推移

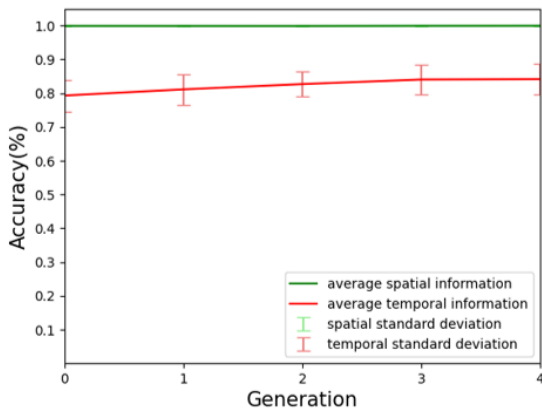


図5 提案手法の精度の推移

また、図6に再学習したモデルの精度の推移を示す。図6と表1の再学習の最高精度の値から、再学習後は提案手法の構造探索時点よりも時間情報の精度が向上し、過学習が起きない汎用性の高い構造探索が出来ている事が分かる。このことから提案手法の再学習は重みの探索に効果的である事が分かる。

表1 従来手法と提案手法の最高精度比較[%]

分類種類	従来手法	提案手法	再学習後
時間情報	85.6	100.0	100.0
空間情報	55.3	91.2	97.7

4.4 機能分化の傾向調査

図7(a)に従来手法、図7(b)に提案手法の時間と空間の相互情報量の結果を示す。図7(a)から、従来手法は空間、時間情報量の両者ともに高くなるように進化していることが分かる。一方、提案手法では空間、時間情報のどちらかが高くなるように進化している。以上より、提案手法でも、従来手法と分布形式が違う形ではあるが、同様に或いはそれ以上に明確にニューロンの機能分化を誘発することが可能である事が言える。

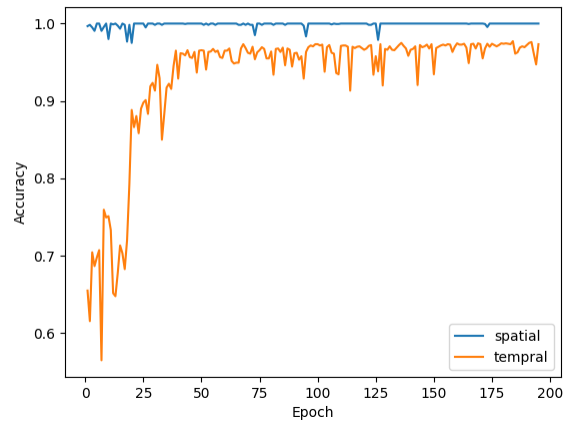
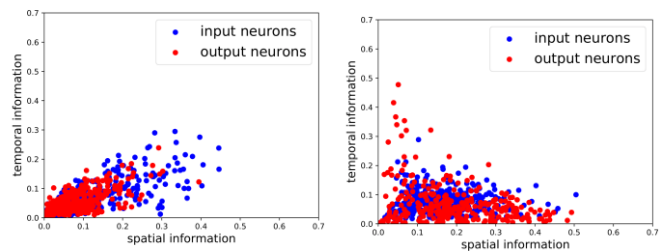


図6 提案手法におけるESNの再学習の精度推移



(a) 従来手法

(b) 提案手法

図7 機能分化の分析

5. おわりに

本研究ではEcho State Networkの効率的な構造探索手法を提案した。提案手法によって、効率的かつ安定的に機能分化の誘発が可能となった。今後の予定としては、探索した構造がどのような意味を持つように進化しているのかの分析や、探索後のReservoir層のニューロンが時間と空間パターンの情報をどのように処理しているのかについての分析、考察などが挙げられる。

参考文献

- [1] Ichiro Tsuda, Yutaka Yamaguti, "Functional differentiations in evolutionary reservoir computing networks", in Chaos: An Interdisciplinary Journal of Nonlinear Science, Vol.31, No.1(2021).
- [2] Maass W, Natschläger T, "Real-time computing without stable states: A new framework for neural computation based on perturbations", Neural computation, Vol.14, No.11(2002).
- [3] Jaeger Herbert, Haas Harald, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication", Science, Vol.304, No.5667(2004).
- [4] Lun Shu-Xian, Yao Xian-Shuang, "A novel model of leaky integrator echo state network for time-series prediction", Neurocomputing, Science, Vol.159, pp.58-66(2015).
- [5] Yamazaki Tadashi, Tanaka Shigeru, "A spiking network model for passage-of-time representation in the cerebellum", European Journal of Neuroscience, Vol.26, No.8(2007).
- [6] Enel Pierre, Procyk, "Reservoir computing properties of neural dynamics in prefrontal cortex", PLoS computational biology, Vol.12, No.6(2016).
- [7] Golberg, David E, "Genetic algorithms in search, optimization, and machine learning", Addison wesley, Vol.1989, No.102(1989).
- [8] Martens James, Sutskever, "Learning recurrent neural networks with hessian-free optimization", ICML(2011).