

Deep Echo State Networkの各層が持つ特徴量抽出機構 Feature Extraction Mechanism for Each Layer of Deep Echo State Network

千葉工業大学 情報科学部 情報工学科 ○ 神田 佳航, 信川 創

K. Kanda and S. Nobukawa

Department of Computer Science, Chiba Institute of Technology,

Abstract Echo state network (ESN) is an efficient machine learning model as a most typical type of reservoir computing framework. Recently, the research on deep echo state network (deepESN) has been conducted. deepESN consists of an input layer, multiple reservoir layers, and an output layer; this new architecture especially achieves a very high memory capacity (MC). However, the concrete role of each layer in feature extraction has not been revealed. Therefore, deepESN parameter adjustments must be conducted by empirical measurements or grid searches based on a trial-and-error method. To establish a design framework of deepESN, revealing the deepESN parameter related to the feature extraction is crucial. As a method for analyzing the dynamics of neural networks, we have applied multiscale entropy (MSE) analysis to a physiological neural network model and found that complex topological features and multiple neural module structures produce complex temporal scale dependencies. Therefore, we hypothesized that the feature extraction function of each layer in deepESN can be revealed by MSE analysis. In this study to validate this hypothesis, we analyzed the output of each layer using MSE analysis and MC.

1 はじめに

リカレントニューラルネットワーク (RNN) の高速処理を実現した計算モデルの一つとして, Reservoir Computing(RC) が再注目されている [1, 2]. RC で用いるニューラルネットワークは, 出力層のリードアウト部分のシナプス荷重調整のみで学習が可能であるため, 非常に処理速度が速く, 特に時系列処理を行うエッジハードウェアの実装に有利である. そこで本研究で用いる, Echo State Network(ESN) は, RC におけるフレームワークの一種である [3]. ESN は入力層, リザーバー層, 出力層から構成されている. 一般的に世間に普及しているディープラーニングやリカレントニューラルネットワークは, 中間層にあるネットワークのシナプス荷重を backpropagation through[4] や, real-time recurrent learning[5] を用いて逐次的に調整する必要がある. そのため, それらの学習に必要なシナプス荷重の調整を大幅に削減している ESN は, 非常に学習効率が向上する. しかし, 単一の ESN だけでは神経ダイナミクスの時間スケールの多様性が限定されてしまい, 偏った特徴量抽出や次元削除をしてしまう問題がある [6].

こうした問題を解決するアプローチとして, ニューロン内部のダイナミクスによって入力的时间履歴性を持たせるために, リザーバーを構成するニューロンモデルにカオスニューロンを用いるモデルの提案や [7][7], 皮質神経接続の形態の一つであるスモールワールドネット

ワーク性を ESN に導入することで, 広範囲の echo state property(ESP) の領域を実現する手法がある [8]. 中でも Deep Echo State Network(deepESN) は非常に高い記憶容量を持つことがわかっている [9]. deepESN は複数のリザーバー層を用意し階層化を行うことで, 各リザーバーの応答特性の時間スケールの多様化を可能にしている. deepESN の研究では階層化を行うことにより, 多様な時間スケールの表現の可能性を持っていることが示唆されているが, 各層でどのような時間スケールで特徴量抽出がされているのかを言及している文献は無い [9]. そのため, deepESN のパラメータ調整は, 試行錯誤の方法に基づく手動またはグリッドサーチによって解決されるため, deepESN のパラメータ調整に関する特性を明らかにし, 設計指針を立てる必要がある.

そこで各リザーバー層が独自の時間スケールで動くようになっており, それらを統合した全体の deepESN として多様な時間スケールが行われているという仮説を立てた. 本研究ではこの仮説を検証するために, 各層の出力に対して Multiscale Entropy(MSE) 解析や Memory Capacity(MC) で分析し, 結果から各層がどのような時間スケールで特徴量抽出を行っているかをまとめる.

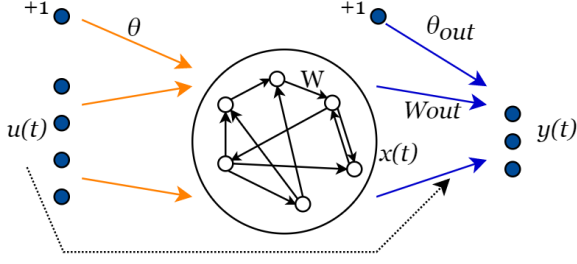


図 1: 単一の Echo State Network の状態遷移

2 手法

2.1 Echo State Network

通常の ESN は、学習の必要がないリザーバー層によって実行される計算によって離散時間動的システムを実装し、出力層で学習される線形リードアウトによって入力時系列データの回帰が行われる。単一の ESN のモデルを図 1 に示す。

図のリザーバーは N_U 次元の入力時系列とニューロン数が N_R のリザーバーユニットによって、式 (2) の状態遷移関数に従って更新される。

$$\mathbf{x}(t) = \tanh(W_{in}\mathbf{u}(t) + \boldsymbol{\theta} + W\mathbf{x}(t-1)) \quad (1)$$

$\mathbf{u}(t) \in \mathbb{R}^{N_U}$ と $\mathbf{x}(t) \in \mathbb{R}^{N_R}$ はそれぞれの時間スケール t での入力と、リザーバーの状態を示す。 $W_{in} \in \mathbb{R}^{N_R \times N_U}$ は、入力からリザーバー層にかけての入力重み行列である。 $\boldsymbol{\theta} \in \mathbb{R}^{N_R}$ はリザーバーへのバイアスの重みベクトルである。ここでリザーバーユニットの入力バイアスは 1 に等しいと仮定する。 $W \in \mathbb{R}^{N_R \times N_R}$ は現在のリザーバー重量行列、 \tanh は要素ごとに適用される双曲線正接活性化関数である。リザーバーパラメータは、Echo State Property (ESP) で指定された制約に従って初期化され、トレーニングは行われない。従って、 W_{in} および $\boldsymbol{\theta}$ の重み値は一様分布から選択される。行列 W の値は、スペクトル半径 α のランダムな行列であり、次のような順に従って決定される。まず、一様分布からランダム行列 W_0 を生成し、次に W_0 と $\rho(W_0)$ に基づいて、スペクトル半径 α を再スケールリングする。

$$W = \alpha \frac{W_0}{\rho(W_0)}. \quad (2)$$

2.2 Deep Echo State Network

本研究で使用する deepESN の状態遷移式についてまとめめる。本研究ではリザーバー層を複数用意し、直列に

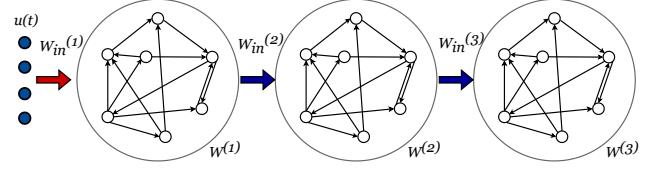


図 2: Deep Echo State Network の状態遷移

接続した図 2 に示すネットワークアーキテクチャを使用している。図 2 は例としてリザーバー層が 3 層の時を示している。リザーバーユニットの次元数を N_R 、リザーバーの状態を $\mathbf{x}(t) \in \mathbb{R}^{N_R}$ とした時、deepESN の各層のリザーバー状態のダイナミクスは、式 (3) で表される。

$$\mathbf{x}^{(l)}(t) = \tanh(W_{in}^{(l)}\mathbf{I}(t) + \boldsymbol{\theta}^{(l)} + W^{(l)}\mathbf{x}^{(l)}(t-1)) \quad (3)$$

ここで、上付き文字 (l) は l 層でのネットワークのパラメータとハイパーパラメータを参照するために使用される。 $W_{in}^{(l)}$ は、 l 層の入力重み行列を示している。 $\boldsymbol{\theta}^{(l)} \in \mathbb{R}^{N_R}$ はバイアスであり、リザーバーへの重みベクトルを示す。 $W^{(l)} \in \mathbb{R}^{N_R \times N_R}$ は各層のリザーバー内部の重み行列を示している。 W_{in} および $\boldsymbol{\theta}$ の重み値は入力スケールリングパラメータ $scale_{in}$ の $[-scale_{in}, scale_{in}]$ 上の一様分布から選択される。特に deepESN は、最初の層 ($l=1$) の時 $W_{in} \in \mathbb{R}^{N_R \times N_U}$ (N_U は入力次元数)、それ以降の層 ($l>1$) の時、 $W_{in} \in \mathbb{R}^{N_R \times N_R}$ となることに注意する。 $\mathbf{I}(t) \in \mathbb{R}^{N_U}$ はタイムステップ t の時の ESN における l 層の構造を示しており場合分けされている (式 (4))。

$$\mathbf{I}^{(l)}(t) = \begin{cases} \mathbf{u}(t) & \text{if } l=1 \\ \mathbf{x}^{(l-1)}(t) & \text{if } l>1. \end{cases} \quad (4)$$

deepESN での出力計算に関しては、標準の ESN のように全てのリザーバーユニットの出力を組み合わせる。リザーバーの階層構造を考慮し、リザーバー数 N_L を用いると、各タイムステップ t での deepESN の出力は式 (5) のように計算できる。

$$\mathbf{y}(t) = W_{out}[\mathbf{x}^{(1)}(t)\mathbf{x}^{(2)}(t)\dots\mathbf{x}^{(L)}(t)] + \boldsymbol{\theta}_{out} \quad (5)$$

ここで出力 $\mathbf{y}(t) \in \mathbb{R}^{N_Y}$ はベクトルであり、 $W_{out} \in \mathbb{R}^{N_Y \times N_L N_R}$ (N_Y は出力の次元数) はリザーバーから read-out にかけての重み行列となる。 $\boldsymbol{\theta}_{out} \in \mathbb{R}^{N_Y}$ はバイアスである。

2.3 評価指標

2.3.1 Memory Capacity

リザーバーの性能を評価するにあたり、MC タスクで提案されるアプローチを用いることで有効性を検討できる

[10]. このタスクは、リザバーのアクティブ化に基づいて入力の変延バージョンをどれだけ上手く呼び出すことができるかを評価することにより、RC ネットワークの短期記憶容量の測定値を示すことができる。MC は τ_m ステップ前の入力と出力の決定係数で定義される。ここで入力信号 $u(t)$ は $[-0.8, 0.8]$ の一様分布から抽出したランダム値とし、教師信号 $y_d(t)$ は $u(t - \tau_m)$ で与えられる。MC は、式 (7) で計算する。

$$MC = \sum_{\tau_m=1}^T MC_{\tau_m} \quad (6)$$

$$MC_{\tau_m} = \frac{\text{cov}^2(u_{\tau_m}, y)}{\sigma^2(u_{\tau_m})\sigma^2(y)} \quad (7)$$

ここで、 $\text{cov}(u_{\tau_m}, y)$ は教師信号と ESN の出力の共分散である。 $\sigma^2(u_{\tau_m})$ と $\sigma^2(y)$ はそれぞれ教師信号、出力の分散を示す。本研究では入力信号は 6000 の時系列データが含まれており、そのうちの 5000 ステップをトレーニングに、残りの 1000 ステップは教師データとして使用した。今回の MC タスクの設定は事前に行われている以下の研究をもとにパラメータ各種の設定を行っている [11]。特に、10 個の完全に接続されたリザバーユニットと 10 層のリザバー層と入力信号のスケール、層間のスペクトル半径を $\rho = 0.9$ を備えた deepESN アーキテクチャを実装した。

2.3.2 Multiscale Entropy 解析

MSE 解析とは、時系列の複雑さを複数の時間スケールで粗視化して定量化する手法である。各層の出力時系列 $x^{N_L}(t)$ を時間スケールファクター (τ) を用いて、式 (8) のようにオーバーラップしないウィンドウで粗視化している。粗視化された時系列は、 τ が 1 の場合、元の時系列と同じであり、 τ が大きいほど、時間スケールが長い [12]。

$$x_j^\tau = \left(\frac{1}{\tau}\right) \sum_{i=(j-1)\tau+1}^{j\tau} y_i, (1 \leq j \leq \frac{t}{\tau}) \quad (8)$$

粗視化された各層の出力時系列の複雑さは、サンプルエントロピー (Sample Entropy: SampEn) によって定量化される。SampEn は、データセット (t) の中で、許容範囲 (r) 内で互いに類似している m 個の連続したデータセットの次の $(m+1)$ 個の連続したデータセットの数を自然対数で表したもので、 t は時系列の長さである。式

(9) は SampEn の導出式である。

$$S_E(m, r) = -\log \frac{U_{m+1}(r)}{U_m(r)} \quad (9)$$

MSE 解析は、粗視化によって複数の時間スケールで時系列データの複雑さを定量化する。

3 結果

3.1 入力信号スケールによる Memory Capacity の違い

deepESN の各リザバー層に入る入力信号はスケールが行われており、スケールの値によって MC の性能が変わる。図 3 の (a)(b) は横軸に、MC で設定された時間スケール τ_m をとり、縦軸に各 τ_m ステップごとの MC の性能を示す。実線と誤差棒は、10 回の試行の平均と、標準誤差を示している。(a)(b) はそれぞれ、 $scale_{in} = 0.1, 1.0$ となっている。図 3 の結果より、 $scale_{in} = 0.1$ では、 $\tau_m = 40$ まで MC の最大値が維持され、そのあと $\tau_m = 100$ になるまで緩やかに MC の値が減少した。一方 $scale_{in} = 1.0$ では $\tau_m = 8$ まで、MC の最大値を示し、その後 $\tau_m = 11$ までに減少した。

3.2 マルチスケールエントロピー解析を行った read out 層における各層の出力

以下の図 4 は、各層に対応したリードアウト層の状態を MSE で解析を行った結果であり、青線とオレンジ線はそれぞれ $scale_{in} = 0.1, 1.0$ の時の結果を示している。実線と誤差棒は MSE 解析における各スケールファクターごとの各層の平均とその標準誤差がプロットされており、縦軸に SampEn の値をとり、横軸はリザバーの各層を示している。平均と標準誤差は 10 回試行した値を使用している。まず浅い層に注目してみると、早い時間スケールを表す $\tau = 1$ では $scale_{in} = 0.1$ と $scale_{in} = 1.0$ はどちらも SampEn ≈ 2 とし、 $\tau = 20$ では、 $scale_{in} = 0.1$ と $scale_{in} = 1.0$ は SampEn ≈ 0.6 を示した。次に深い層である $layer = 6 - 10$ に注目すると、 $\tau = 1$ では $scale_{in} = 1.0$ が $scale_{in} = 0.1$ より高い SampEn の値を示したが、一方 $\tau = 20$ では、 $scale_{in} = 0.1$ が $scale_{in} = 1.0$ よりも高い SampEn を示した。

4 考察

本研究では、deepESN の全体の性能の評価と、それに基づく deepESN のアーキテクチャによる分析を行った。まず入力信号のスケールによる MC タスクの性能評価の結果は、図 3 の結果は各層に供給される

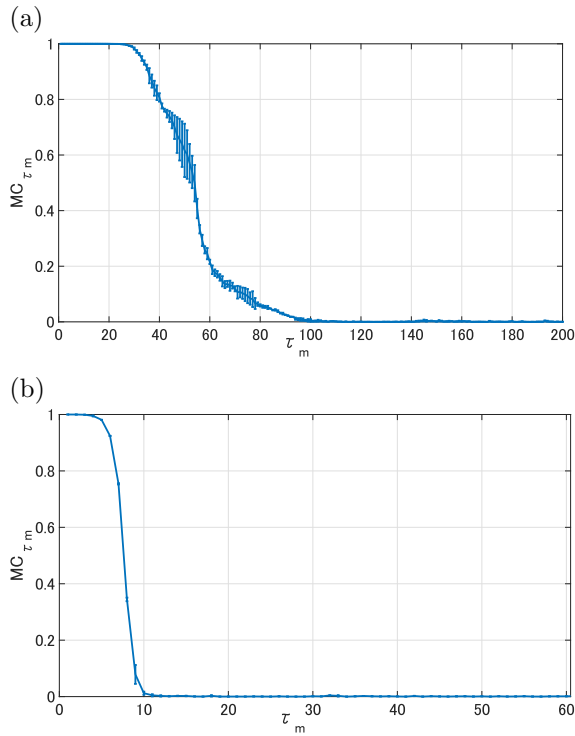


図 3: (a)(b) は $scale_{in} = 0.1, 1.0$ の時の MC スコアを示しており、縦軸に MC の値をとり、横軸は時間スケール τ_m を示す。実線と誤差棒は 10 回試行の平均、標準誤差を示す。

入力信号強度の調整を行うことで、deepESN 全体の記憶容量に大きな差が生まれた。この結果より、deepESN にも入力信号強度の調整による性能の差が生まれることが確認できるため、ハイパーパラメータである入力信号のスケールリングが与える、性能への影響が重要であると考えられる。

入力信号の調整による各層への影響を見た、リードアウト層での MSE 解析の結果の図 4 から、deepESN の浅い層は早い時間スケールでの特徴量抽出をおこなっており、深い層は早い時間スケールから遅い時間スケールにおいて一様に特徴量抽出を行っていることが明らかになった。

5 おわりに

本研究では、deepESN の設計指針を得るために MC や MSE 解析を用いて、ハイパーパラメータである信号強度の調整による比較を行った。今回の結果から、deepESN の特性が浅い層と深い層とで異なるダイナミクスの特徴量抽出を行っていることが分かった。本研究は deepESN の特性を理解することにつながったため、今後はこれら

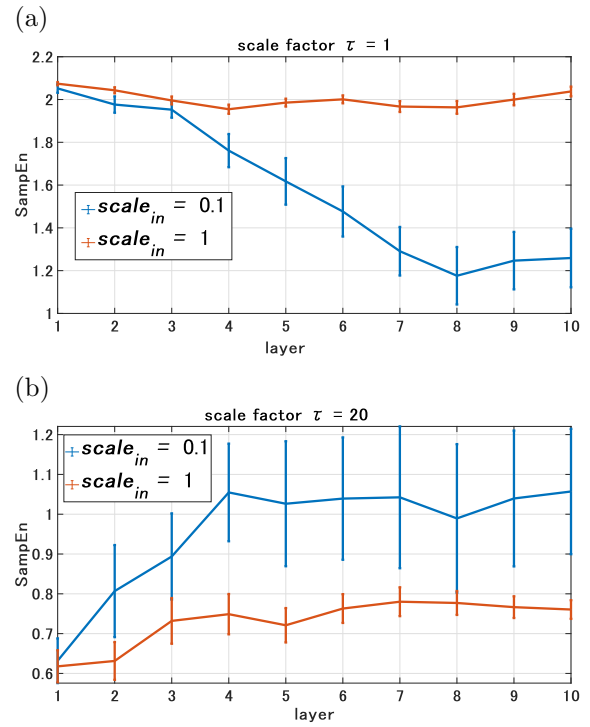


図 4: (a)(b) はそれぞれスケールファクター $\tau = 1, 20$ の時の MSE 解析の結果を示しており、縦軸に SampEn の値、横軸にリードアウト層における各層の結果を表している。青線とオレンジ線は入力信号スケールリング $scale_{in} = 0.1, 1.0$ を示しており、実線と誤差棒は 10 回試行の平均、標準誤差を示す。

の結果をもとに、他の様々な予測・分類タスクを用いて deepESN の各層の機能を評価していく。

参考文献

- [1] Herbert Jaeger. The “echo state” approach to analysing and training recurrent neural networks with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, Vol. 148, No. 34, p. 13, 2001.
- [2] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, Vol. 3, No. 3, pp. 127–149, 2009.
- [3] Herbert Jaeger. Echo state network. *scholarpedia*, Vol. 2, No. 9, p. 2330, 2007.

- [4] Jürgen Schmidhuber. Deep learning. *Scholarpedia*, Vol. 10, No. 11, p. 32832, 2015.
- [5] Jacob Menick, Erich Elsen, Utku Evci, Simon Osindero, Karen Simonyan, and Alex Graves. Practical real time recurrent learning with a sparse approximation. In *International Conference on Learning Representations*, 2020.
- [6] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Comparison between deepesns and gated rnns on multivariate time-series prediction. *arXiv preprint arXiv:1812.11527*, 2018.
- [7] Yudai Ebato, Sou Nobukawa, and Haruhiko Nishimura. Effect of neural decay factors on prediction performance in chaotic echo state networks. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1888–1893. IEEE, 2021.
- [8] Yuji Kawai, Jihoon Park, and Minoru Asada. A small-world topology enhances the echo state property and signal propagation in reservoir computing. *Neural Networks*, Vol. 112, pp. 15–23, 2019.
- [9] Claudio Gallicchio, Alessio Micheli, and Luca Pedrelli. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, Vol. 268, pp. 87–99, 2017.
- [10] Herbert Jaeger, et al. *Short term memory in echo state networks*, Vol. 5. GMD-Forschungszentrum Informationstechnik, 2001.
- [11] Benjamin Schrauwen, Marion Wardermann, David Verstraeten, Jochen J Steil, and Dirk Stroobandt. Improving reservoirs using intrinsic plasticity. *Neurocomputing*, Vol. 71, No. 7-9, pp. 1159–1171, 2008.
- [12] Madalena Costa, Ary L Goldberger, and C-K Peng. Multiscale entropy analysis of biological signals. *Physical review E*, Vol. 71, No. 2, p. 021906, 2005.