

非線形時系列予測のためのロバストな Echo State Network の設計

数理情報学専攻 48186201 秋山 貴則

指導教員 田中 剛平 特任准教授

1 はじめに

非線形ダイナミクスは、天候・経済・音声データなどに現れ、その予測は我々の実生活に広く貢献する技術である。近年、予測手法として高速学習可能な Reservoir Computing (RC) が広く注目を浴びている。特に RC の代表的なモデルである Echo State Network (ESN) は、非線形時系列予測に有効であるが、ハイパーパラメータに性能が強く依存するという問題点を持つ。

本論文では、ハイパーパラメータの変化にロバストな ESN として、多段階学習 ESN と、敵対的学習を用いた ESN の 2 つを提案する。本研究では、これら 2 つの提案手法に関して、非線形時系列予測のベンチマークタスクを用いた実験的解析を通じて、ロバストな ESN の構築に有用である事を確認した。

2 Echo State Network

Reservoir Computing (RC) は、Recurrent Neural Network (RNN) の一種であり、高速学習可能な計算モデルである。図 1 に RNN と RC の比較図を示す。RC は、入力層とリザーブと呼ばれる隠れ層の結合重みを固定し、出力層のみを更新するため、時間方向に遡って勾配を伝播させる必要がなく、計算コストが小さい。特

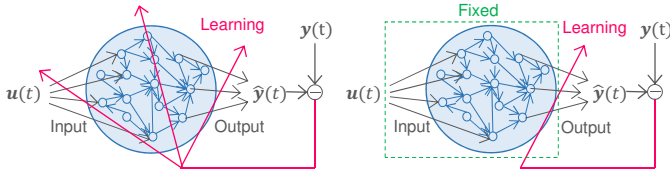


図 1. RNN と RC の比較。

に、Echo State Network (ESN) [1] は、RC の代表的なモデルであり、入力層とリザーブの重みをランダムに決定する。ESN において、入力 $u(t)$ から出力 $y(t)$ の予測を考える。 $u(t)$ は入力層により、高次元空間に写像され、リザーブの状態ベクトルは以下の式で発展する：

$$r(t + \Delta t) = \tanh [Ar(t) + W_{in}u(t)]. \quad (1)$$

出力層は、関数 f_{out} と重み W_{out} を用いて、 $\hat{y}(t) = f_{out}(r(t), W_{out})$ として予測値を出力する。通常、線形変換が用いられる事が多い。これらから、最小二乗誤差規範及び、リッジ回帰によって解析的に最適な出力層重み \hat{W}_{out} を得る。ESN は非線形時系列予測に有効な手

法である [2] が、出力層のスケーリング、リザーブの隣接行列 A のスペクトル半径 ρ などのハイパーパラメータに強く性能が依存する、という課題がある。

3 多段階学習 Echo State Network

3.1 機構

提案法の機構を図 2 に示す。多段階学習 ESN は、入

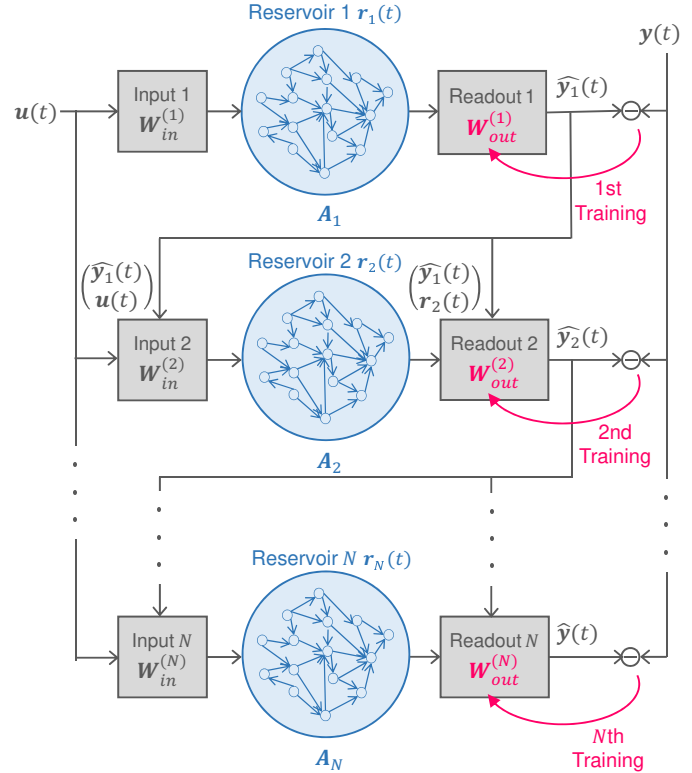


図 2. N 段階学習 ESN の機構。

力層、リザーブ、出力層を有する予測器 N 個から構成され、 N 個の出力層が段階的に学習される。各予測器の出力が次の予測器の入力層と出力層の両方に入力されるのが特徴である。まず、各予測器の入力層重み、リザーブの隣接行列をランダムに決定する。次に、各予測器の出力層を段階的に学習する。各予測器の学習は、通常の ESN と同様に解析的に計算できる。最終的に、 N 個目の予測器の出力を、全体の予測信号として利用する。本提案手法は、予測誤差が予測器を経る毎に減少するモデルになっている。

3.2 実験的評価

Lorenz 方程式と、NARMA10 の予測を通じて提案法の有効性を評価する。合計リザーブサイズが同一という条件の下で、2 段階学習 ESN と、通常の ESN の予測精度を比較した。評価指標は、Lorenz 方程式では

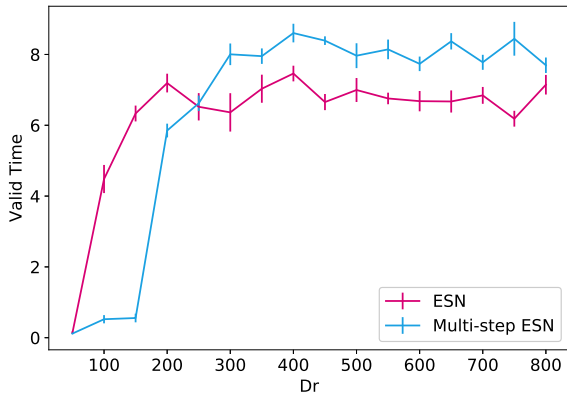


図 3. Lorenz 方程式の予測精度の比較.

Valid Time [4], NARMA10 では RMSE を用いる. 図 3 に Lorenz 方程式の結果を示す. 結果から, 提案法は通常の ESN を上回る精度で予測が可能である事が分かる. また, 本研究では, ESN と多段階学習 ESN のリアプノフ指数を計算し, 提案法がスペクトル半径の変化に対してロバストに力学系への適応性能を発揮する事が示された (図 4).

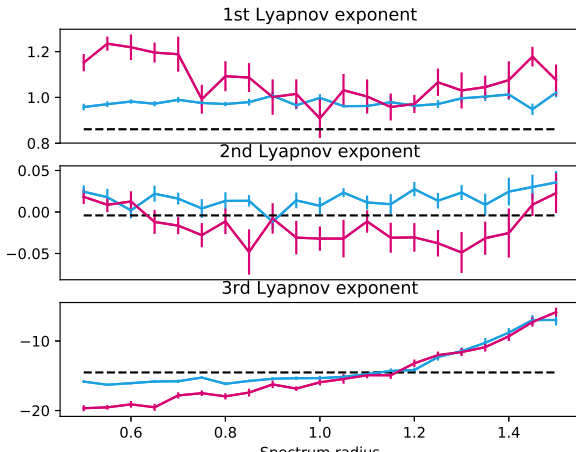


図 4. Lorenz 方程式の予測におけるリアプノフ指数の比較.

4 敵対的誤差を用いた Echo State Network

4.1 機構

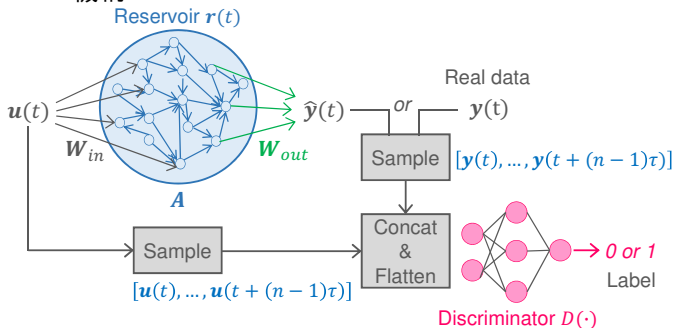


図 5. 敵対的学習を用いた ESN の概要図.

提案法の機構を図 5 に示す. 本提案手法は, Generative Adversarial Network (GAN) [3] の枠組みを ESN に適用したモデルであり, 幅広い RC モデルに適用可能である. ESN の出力は, 時間遅れ座標系に変換された後, Discriminator に入力される. Discriminator は,

入力データが実データか ESN が生成したデータであるかを判別する. 全体の損失関数は従来の二乗誤差と敵対的誤差の重み付き和として定義され, ESN の出力層と Discriminator の重みが交互に最適化される. 詳細なアルゴリズムは Algorithm1 を参考にされたい.

Algorithm 1 ミニバッチ確率勾配法を用いた, 敵対的学習を用いた ESN の学習アルゴリズム. 本研究では, $k_G = 2$ とした.

Require: 時間遅れ座標系におけるパラメータ n, τ , 敵対的誤差の重み ω_D , バッチサイズ m , Discriminator の 1 学習イテレーションに対する Generator の学習イテレーション数 k_G .

- 1: ESN の出力層の重み \mathbf{W}_{out} は事前に従来の二乗誤差規範を用いて学習しておく.
- 2: **for** 学習イテレーション数 **do**
- 3: $-T \leq t \leq 0$ から m サンプル $t^{(1)}, \dots, t^{(m)}$ 取る.
- 4: Discriminator のパラメータを確率勾配に従い Loss が降下するように更新する:

$$\nabla_{\theta_D} \frac{1}{m} \sum_{i=1}^m [-\log(1 - D(\mathbf{u}, \hat{\mathbf{y}}, t^{(i)})) - \log D(\mathbf{u}, \mathbf{y}, t^{(i)})]. \quad (5.6)$$

- 5: $E_{L_{ADV}}, E_{L_{SE}}$ を計算する.
- 6: **for** k_G 回 **do**
- 7: $-T \leq t \leq 0$ から m サンプル $t^{(1)}, \dots, t^{(m)}$ 取る.
- 8: ESN の出力層の重み \mathbf{W}_{out} を確率勾配に従い Loss が降下するように更新する:

$$\nabla_{\mathbf{W}_{out}} \left(\frac{\omega_D}{m} \sum_{i=1}^m [-\log D(\mathbf{u}, \hat{\mathbf{y}}, t^{(i)})] + (1 - \omega_D) \frac{E_{L_{ADV}}}{E_{L_{SE}}} L_G^S \right). \quad (5.7)$$

- 9: **end for**
- 10: **end for**

4.2 実験的評価

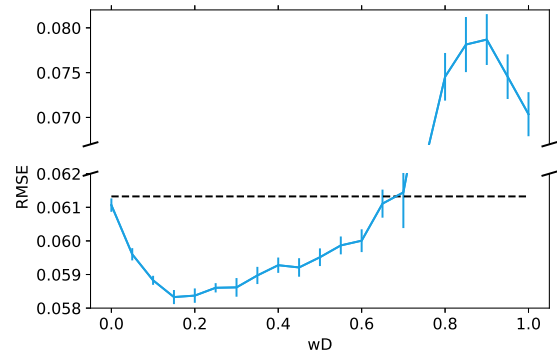


図 6. 敵対的誤差の重み ω_D と NARMA10 の予測精度の関係.

本研究では, リザーバーの隣接行列のスペクトル半径が適切でない ($\rho = 0.4$) 場合に, NARMA10 の予測精度が改善されるかを実験した. 図 6 の結果から, 敵対的誤差の導入により予測精度が改善される事が分かる.

5 結論

本研究では, 多段階学習 ESN と敵対的学習を用いた ESN を提案し, ロバストな ESN の構築に有用であることを実験的に示した. 実データへの応用が今後の展望として考えられる.

参考文献

- [1] H. Jaeger, *GMD Report 148*, 2001.
- [2] H. Jaeger and H. Haas, *Science*, Vol.304, No. 5667, pp. 78–80, 2004.
- [3] I. Goodfellow et al., *Advances in Neural Information Processing Systems 27*, pp. 2672–2680, 2014.
- [4] J. Pathak et al., *Chaos: An Interdisciplinary Journal of Nonlinear Science*, Vol. 28, No. 041101, 2018.