

機械学習による時系列データの学習過程の解明

中井拳吾

東京海洋大学 学術研究院

1. はじめに

機械学習は物理学の様々な分野で注目されている。近年、機械学習の一種で、再帰的ニューラルネットワーク構造をもったリザーバーコンピューティング [1, 5] が時系列やリャプノフ指数などの予測において有効であることが報告されている。我々はこの機械学習手法を流体のエネルギー変数に適用し、時間発展の予想等を成功させている [2, 3]。

一方で、単に時系列を予測できるだけではその現象のモデルを構成したと言うには不十分である。そこで、本研究では学習量(データの量)などに注目しつつ学習により得た時間発展モデルがもとの力学系構造をどの程度再現できるかを明らかにする。

本稿は小林 幹氏 (立正大学), 斉木 吉隆氏, 堤 夏輝氏(一橋大学) との共同研究に基づくものである。

2. 機械学習手法

まずリザーバーコンピューティングによる学習方法について説明する。 $\frac{d\phi}{dt} = \mathbf{f}(\phi)$ で表される力学系の変数, $\mathbf{u} = \mathbf{h}_1(\phi) \in \mathbb{R}^M$ と $\mathbf{s} = \mathbf{h}_2(\phi) \in \mathbb{R}^P$ について考える。ただし、時系列データ \mathbf{u} は \mathbf{s} の時系列を予測する時刻にないとする。リザーバーベクトル $\mathbf{r} \in \mathbb{R}^N$ ($N \gg M$) は次の式で支配されている。

$$\mathbf{r}(t + \Delta t) = (1 - \alpha)\mathbf{r}(t) + \alpha \tanh(\mathbf{A}\mathbf{r}(t) + \mathbf{W}_{\text{in}}\mathbf{u}(t)) \quad (1)$$

ただし、 Δt は時間幅を表す。また、行列 \mathbf{A} , \mathbf{W}_{in} は $N \times N$, $N \times M$ 行列である。パラメータ α ($0 < \alpha \leq 1$) は \mathbf{r} の力学系の非線形性を表し、式(1)の Δt に依存して決める。リザーバーベクトルは $(0, 1]$ の一様乱数から定めた $r(-\tau)$ を初期値とする。ただし、 $\tau/\Delta t (\gg 1)$ はトランジェント時間に相当する。

行列 \mathbf{W}_{in} の各列はただ一つのみ $[-\sigma, \sigma]$ の一様乱数から値を与え、それ以外は 0 とする。行列 \mathbf{A} は $D \times N$ 個の非 0 の要素を持つ疎行列とする。 $D \times N$ 個の非 0 成分は $[-1, 1]$ の一様乱数から値を与える。さらに、行列 \mathbf{A} については最大固有値が ρ になるように行列のすべての要素を調整する。

リザーバーベクトル \mathbf{r} は次のようにして得たいベクトルに変換する。

$$\hat{\mathbf{s}}(t) = \mathbf{W}_{\text{out}} \mathbf{r}(t) + \mathbf{c}.$$

これにより $L = T/\Delta t$ ステップのリザーバーベクトル $\{\mathbf{r}(l\Delta t)\}_{l=1}^L$ を得る。

$0 < t \leq T$ (学習時間と呼ぶ) での時系列 $\hat{\mathbf{s}}$ が既知の時系列 \mathbf{s} の近似になるように \mathbf{W}_{out} と \mathbf{c} を決定する。この決定がリザーバーコンピューティングの学習に相当する。これは次の二次形式が最小化するように \mathbf{W}_{out} と \mathbf{c} を決定することに対応する。

$$\sum_{l=1}^L \|(\mathbf{W}_{\text{out}}\mathbf{r}(l\Delta t) + \mathbf{c}) - \mathbf{s}(l\Delta t)\|^2 + \beta[\text{Tr}(\mathbf{W}_{\text{out}}\mathbf{W}_{\text{out}}^T)] \quad (2)$$

ただし \mathbf{q} に対して $\|\mathbf{q}\|^2 = \mathbf{q}^T\mathbf{q}$ とし, $\beta \geq 0$ を含む項は \mathbf{W}_{out} の過剰適合を避けるために導入している. この決定が妥当であれば, $t > T$ (予測時間と呼ぶ) での $\hat{\mathbf{s}}(t)$ は $\mathbf{s}(t)$ を再現できるはずである. 二次形式 (式 (2)) を最小化する解を $\mathbf{W}_{\text{out}}^*$ と \mathbf{c}^* とすると予測時間では

$$\hat{\mathbf{s}}(t) = \mathbf{W}_{\text{out}}^* \mathbf{r}(t) + \mathbf{c}^*$$

として時系列 $\hat{\mathbf{s}}$ を得る. $\mathbf{W}_{\text{out}}^*$ と \mathbf{c}^* は次のように書けることが知られている [参考文献 1].

$$\mathbf{W}_{\text{out}}^* = \delta\mathbf{S}\delta\mathbf{R}^T(\delta\mathbf{R}\delta\mathbf{R}^T + \beta\mathbf{I})^{-1}$$

$$\mathbf{c}^* = -[\mathbf{W}_{\text{out}}^* \bar{\mathbf{r}} - \bar{\mathbf{s}}]$$

ただし, $\bar{\mathbf{r}} = \sum_{l=1}^L \mathbf{r}(l\Delta t) / L$, $\bar{\mathbf{s}} = \sum_{l=1}^L \mathbf{s}(l\Delta t) / L$ とし, \mathbf{I} は $N \times N$ の単位行列, $\delta\mathbf{R}$ は l 行が $\mathbf{r}(l\Delta t) - \bar{\mathbf{r}}$ となるような行列とする. 同様に, $\delta\mathbf{S}$ は l 行が $\mathbf{s}(l\Delta t) - \bar{\mathbf{s}}$ となるような行列とする.

ただし, 予測時間 $t > T$ での時系列データ \mathbf{u} は利用できないため, $\mathbf{u}(t)$ として式 (1) で得た $\hat{\mathbf{s}}(t)$ を用いる:

$$\mathbf{r}(t + \Delta t) = (1 - \alpha)\mathbf{r}(t) + \alpha \tanh(\mathbf{A}\mathbf{r}(t) + \mathbf{W}_{\text{in}}\hat{\mathbf{s}}(t)).$$

3. 設定と結果

ローレンツ方程式 $\frac{dx}{dt} = -10x + 10y$, $\frac{dy}{dt} = -xz + rx - y$, $\frac{dz}{dt} = xy - 8z/3$ について考える.

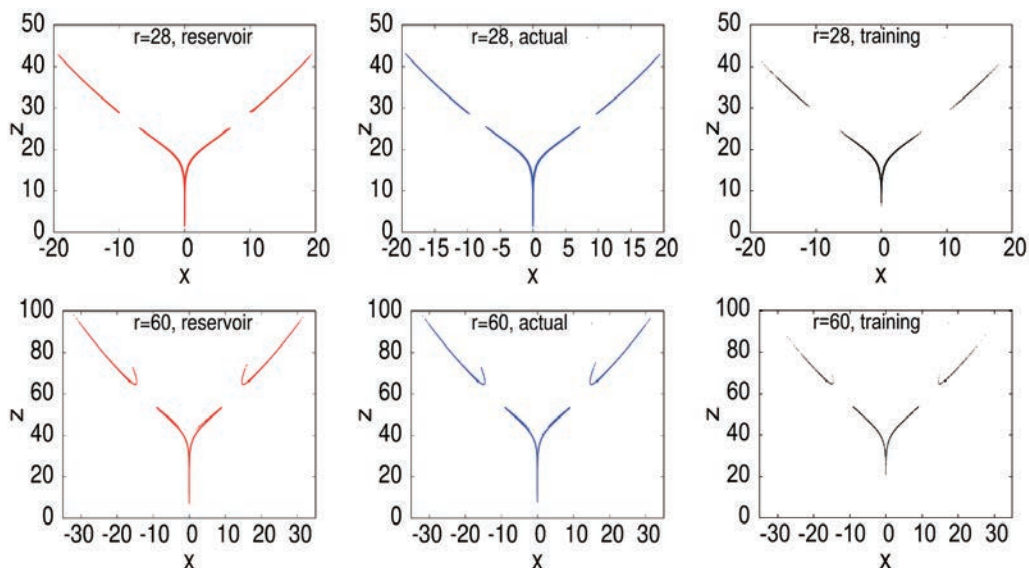
この方程式系はカオス性がある. 本研究ではパラメータ r として 28 (もしくは 60) として得られる時系列を扱う. $r = 60$ 付近では力学系は非双曲的であり, パラメータの微小な変化により力学系の構造が大きく変化しうる一方で, $r = 28$ 付近では力学系は双曲的で構造安定であり, そのような構造変化は起こらないことが知られている. ローレンツ方程式の時間発展により得られた時系列データを $\{\mathbf{u}, \mathbf{s}\}$ としてリザーバーコンピューティングにより学習させて得られた時間発展モデルが, 元の力学系構造をどの程度再現するかを解析する. 特に本誌では会軌道の断面や不変密度分布といった不変集合に注目する.

3.1 ポアンカレ断面

学習により得た機械学習モデルのアトラクタの大きさや概形について, ローレンツ方程式のものとの比較をする. 3次元では比較が困難であるため, 2次元平面に落として考察する. 具体的には解軌道が適当な平面を横切るときにプロットして得られるポアンカレ断面というものを用いて比較をする. 学習により得た機械学習モデルから得られた 10^6 時間の時系列データとローレンツ方程式の時間発展により得られた 10^6 時間の時系列データ, 学習に用いた 5000 時間の時系列データのそれぞれに対して, 適当な平面として $x = y$ となる平面にポアンカレ断面を書いた (第 1 図).

まず, 学習に用いた 5000 時間の時系列データにより得たポアンカレ断面はローレンツ方程式の 10^6 時間の時系列データにより得られたポアンカレ断面に比べて小さいことが見て取れる. 他方で, 機械学習モデルから得られたポアンカレ断面はローレンツ方程式の 10^6 時間の時系列

データにより得られたポアンカレ断面と似ていることがわかる。このことから、ローレンツ方程式の 10^6 時間の時系列データにより得られたポアンカレ断面に比べて学習に用いた 5000 時間の時系列データにより得たポアンカレ断面は小さいにもかかわらず、機械学習モデルから得られたポアンカレ断面はローレンツ方程式の断面を十分に再現していることがわかる。



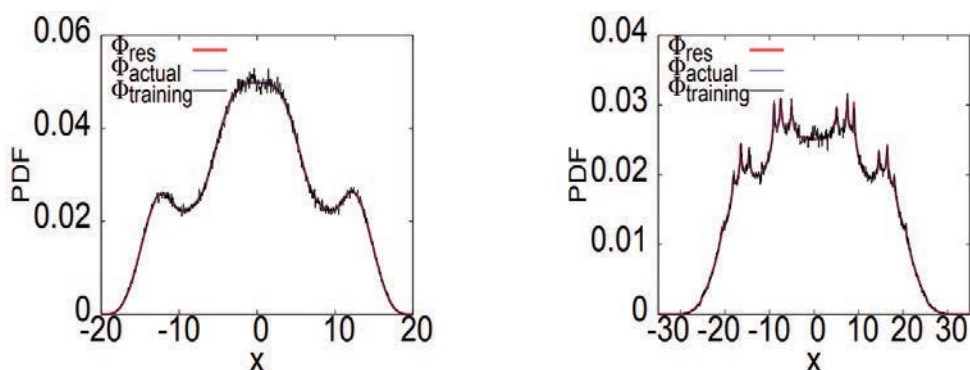
第 1 図: ポアンカレ断面の様子 ($r = 28$ (上段) $r = 60$ (下段))。

機械学習モデルの時間発展によって得られた時系列データと比較のためローレンツ方程式の時間発展によって得られた時系列データ, この機械学習モデルを作るために用いた時系列データのそれぞれに対して, $|x - y| < 0.05$ となったときにプロットした。それぞれの時間長は 10^6 , 10^6 , 5000 である。

3.2 密度分布

機械学習モデルから得られた 10^6 時間の時系列とローレンツ方程式の時間発展により得られた 10^6 時間の時系列, 学習に用いた 5000 時間の時系列に対して, x 変数の時系列データの滞在時間密度分布を計算した (第 2 図)。

学習に用いた 5000 時間の時系列データにより得た密度分布はローレンツ方程式の 10^6 時間の時系列データにより得られた密度分布に比べて激しく振動していることが見て取れる。他方で, 機械学習モデルから得られた密度分布とローレンツ方程式の 10^6 時間の時系列データにより得られた密度分布の差はかなり小さいことが見て取れる。このことから, 学習に用いた時系列データから得られた密度分布は時間が短い, つまりデータ量が少ないために激しく振動しているにもかかわらず, 機械学習モデルから得られた密度分布はローレンツ方程式の密度分布を十分に再現していることがわかる。特に, $r = 60$ の密度分布は特異な構造があるが, この構造も機械学習モデルは再現していることが見て取れる。



第2図: ローレンツ方程式の滞在時間分布 ($r = 28$ (左図) $r = 60$ (右図))。

機械学習モデルの時間発展によって得られた x 変数の密度分布 (Φ^{res})と比較のためローレンツ方程式の時間発展によって得られた x 変数の密度分布 (Φ^{actual})を描いた。また、この機械学習モデルを作るために用いたトレーニングデータの x 変数の密度分布 (Φ^{training}) も描いた。

4. まとめ

決定論的な力学系の時系列から機械学習力学系モデルを構築し、それが時系列のどのような構造を再現しうるかを明らかにした。本研究では、一見すると学習には不十分に思えるデータ量の学習データから構築した機械学習モデルであってもポアンカレ断面や密度分布といった不変密度分布が高精度に再現できることが明らかになった。なお本結果は、論文 [4] としてまとめて査読付き国際誌に投稿した。

参考文献

- [1] H. Jaeger, and H. Haas, *Science*, 304, pp. 78-80, (2004).
- [2] K. Nakai, and Y. Saiki, *Physical Review E* 98, 023111:1-6, (2018).
- [3] K. Nakai, and Y. Saiki, *Discrete and Continuous Dynamical Systems Series S* 14, 3, (2021).
- [4] M. Kobayashi, K. Nakai, Y. Saiki, and N. Tsutsumi, *Dynamical system analysis of a data-driven model constructed by reservoir computing*, *Physical Review E* 104, 044215:1-7, (2021).
- [5] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, *Chaos*, 27, pp. 041102. pp. 391-403, (2017).