「The “echo state” approach to analysing and training recurrent neural networks」のまとめ

市村 剛大

2019 年 6 月 22 日

概要

echo state network とは出力層のみ学習するリカレントニューラルネットワークである

1 Echo states

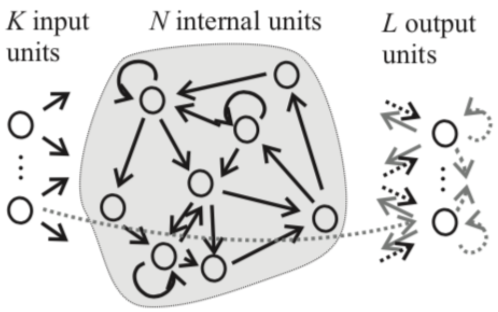

図 1 echo state network の基本構成

図 1 のようなニューラルネットワークを考える。入力数 K、ニューロン数 N、出力数 L であり、ある時間 n での入力ベクトルを $\mathbf{u}(n) = (u_1(n), ..., u_K(n))$、内部状態ベクトルを $\mathbf{x}(n) = (x_1(n), ..., x_N(n))$、出力ベクトルを $\mathbf{y}(n) = (y_1(n), ..., y_L(n))$ とする。また $N \times K$ の入力層荷重行列を $\mathbf{W}^{in}$、$N \times N$ の内部結合荷重行列を $\mathbf{W}$、$L \times (K + L + N)$ の出力層行列を $\mathbf{W}^{out}$、そして $N \times L$ の出力層から内部ニューロンへの戻り値行列を $\mathbf{W}^{back}$ とする。このとき、内部状態は次式のように更新される。

$$\mathbf{x}(n+1) = \mathbf{f}(\mathbf{W}^{in}\mathbf{u}(n+1) + \mathbf{W}\mathbf{x}(n) + \mathbf{W}^{back}\mathbf{y}(n)), \tag{1}$$

ここで $\mathbf{f}$ は活性化関数 (ベクトル) である。また出力ベクトル $\mathbf{y}(n+1)$ は次式のように得られる。

$$\mathbf{y}(n+1) = \mathbf{f}^{out}(\mathbf{W}^{out}(\mathbf{u}(n+1), \mathbf{x}(n+1), \mathbf{y}(n))), \tag{2}$$

$\mathbf{f}^{out}$ は出力層の活性化関数 (ベクトル) である。以降簡単のために入力 $\mathbf{u}(n)$ のシークエンス (入力列) を表 1 のように略記する。また、演算子 T を導入し、入力列 $\bar{\mathbf{u}}^k$ が入力されたのちの内部状態を $\mathbf{x}(x +$

表 1　入力 $\mathbf{u}(n)$ のシークエンス (列) の略記法

$\bar{\mathbf{u}}^{\pm\infty}$	左右無限大の入力列
$\bar{\mathbf{u}}^{+\infty}$	右に無限大に続く入力列
$\bar{\mathbf{u}}^{-\infty}$	左に無限大に続く入力列
$\bar{\mathbf{u}}^{k}$	長さ k の入力列

$h) = T(\mathbf{x}(n), \mathbf{y}(n), \bar{\mathbf{u}}^k)$ のように表すこととする。これは、出力層からのフィードバックがない場合は $\mathbf{x}(x+h) = T(\mathbf{x}(n), \bar{\mathbf{u}}^k)$ となる。このとき、以下のように出力からのフィードバックがない場合の echo state network を定義する。

定義 1　標準コンパクト性条件[*1]が成立し、ネットワークが出力層からのフィードバックを持たない場合を考える。このとき、もしネットワークの状態 $\mathbf{x}(n)$ が左に無限の入力列 $\bar{\mathbf{u}}^{-\infty}$ によって一意に決まるならば、ネットワークは echo state を持つ。さらに正確にいえば、すべての入力ベクトル、ネットワークの状態に対して、$\mathbf{x}(i) = T(\mathbf{x}(i-1), \mathbf{u}(i))$ と $\mathbf{x}'(i) = T(\mathbf{x}'(i-1), \mathbf{u}(i))$ が存在するならば、$\mathbf{x}(n) = \mathbf{x}'(n)$ である。

echo state property は次のように記述することもできる。入力 echo 関数 $\mathbf{E} = (e_1, ..., e_N)$ が存在し、$e_i : U^{-\mathbb{N}} \to \mathbb{R}$ のとき、すべての左に無限の入力列履歴に対して、現在の状態は

$$\mathbf{x}(n) = \mathbf{E}(..., \mathbf{u}(n-1), \mathbf{u}(n)) \tag{3}$$

のように表せる。

定義 2　(a)$\forall i < n$ に対して $T(\mathbf{x}(i), \mathbf{u}(i+1)) = \mathbf{x}(i+1)$ とき、左に無限の状態列 $\bar{\mathbf{x}}^{-\infty}$ は左に無限の入力列 $\bar{\mathbf{u}}^{-\infty}$ と互換性がある (compatible) と呼ぶ。(b) 同様に、$\forall i$ に対して $T(\mathbf{x}(i), \mathbf{u}(i+1)) = \mathbf{x}(i+1)$ とき、左右に無限の状態列 $\bar{\mathbf{x}}^{\infty}$ は左右に無限の入力列 $\bar{\mathbf{u}}^{\infty}$ と互換性がある。(c) もし状態列 $..., \mathbf{x}(n-1), \mathbf{x}(n)$ が存在するならば、$T(\mathbf{x}(i), \mathbf{u}(i+1)) = \mathbf{x}(i+1)$ かつ $\mathbf{x} = \mathbf{x}(n)$ であるとき、ネットワーク状態 $\mathbf{x} \in A$ は入力列 $\bar{\mathbf{u}}^{-\infty}$ と終端互換性がある (end-compatible) と呼ぶ。(d) もし $\mathbf{x}(n-h), ..., \mathbf{x}(n) \in A^{h+1}$ が存在するならば、$T(\mathbf{x}(i), \mathbf{u}(i+1)) = \mathbf{x}(i+1)$ かつ $\mathbf{x} = \mathbf{x}(n)$ であるとき、ネットワーク状態 $\mathbf{x} \in A$ は入力列 $\bar{\mathbf{u}}^{h}$ と終端互換性がある (end-compatible) と呼ぶ。

定義 3　標準コンパクト条件が成立し、出力層からのフィードバックがないとする。

1. すべての右側に無限な入力列 $\bar{\mathbf{u}}^{+\infty}$ に対して、null 列 $(\delta_h)_{h\geq 0}$ が存在する場合を考える。すべての状態 $\mathbf{x}, \mathbf{x}' \in A$、すべての $h \geq 0$、すべての有限入力列 $\bar{\mathbf{u}}^h = \mathbf{u}(n), ..., \mathbf{u}(n+h)$ について $d(T(\mathbf{x}, \bar{\mathbf{u}}^h), T(\mathbf{x}', \bar{\mathbf{u}}^h)) < \delta_h$ となるとき、これを状態契約 (state contracting) であるという。ここで d は $\mathbb{R}^N$ 上のユークリッド距離である。
2. すべての左側に無限な入力列 $\bar{\mathbf{u}}^{-\infty}$ に対して、null 列 $(\delta_h)_{h\geq 0}$ が存在する場合を考える。すべての状態 $\mathbf{x}, \mathbf{x}' \in A$、すべての $h \geq 0$、すべての有限入力列 $\bar{\mathbf{u}}^h = \mathbf{u}(n-h), ..., \mathbf{u}(n)$ について $d(T(\mathbf{x}, \bar{\mathbf{u}}^h), T(\mathbf{x}', \bar{\mathbf{u}}^h)) < \delta_h$ となるとき、これを状態忘却 (state forgetting) であるという。
3. すべての左側に無限な入力列 $\bar{\mathbf{u}}^{-\infty}$ に対して、null 列 $(\delta_h)_{h\geq 0}$ が存在する場合を考える。すべての状態 $\mathbf{x}, \mathbf{x}' \in A$、すべての $h \geq 0$、すべての有限入力列 $\bar{\mathbf{u}}^h = \mathbf{u}(n-h), ..., \mathbf{u}(n)$、すべての $\bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}^h, \bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}^h$

[*1] (i) 入力をコンパクト集合 U から得ること、(ii) ネットワーク状態がコンパクトセット A に属すこと。

形式の左に無限の入力列、すべての $\bar{\mathbf{w}}^{-\infty}\bar{\mathbf{u}}^h$ と終端互換性がある $\mathbf{x}$ と $\bar{\mathbf{v}}^{-\infty}\bar{\mathbf{u}}^h$ と終端互換性がある $\mathbf{x}'$ について $d(\mathbf{x}, \mathbf{x}) < \delta_h$ となるとき、これを入力忘却 (input forgetting) であるという。

命題 1　標準コンパクト条件が成立し、出力層からのフィードバックがない場合を考える。T が状態、入力のなかで連続であると仮定する。このとき状態契約、状態忘却、入力忘却の特性は、すべてネットワークが echo states を持っているということに等しい。

命題 2　標準コンパクト条件が成立し、出力層からのフィードバックがない場合を考える。T が状態、入力のなかで連続である、echo states であるとする。このとき、すべての左に無限の入力列 $\bar{\mathbf{u}}^{-\infty}$ について、すべての $\epsilon > 0$ に対して、$d(\mathbf{u}(k), \mathbf{u}'(k)) < \delta$(ただし k は $-h \geq k \geq 0$ である任意の k) を満たすようなすべての入力列 $\bar{\mathbf{u}}'^{-\infty}$ について $d(\mathbf{E}(\bar{\mathbf{u}}^{-\infty}), \mathbf{E}(\bar{\mathbf{u}}'^{-\infty})) < \epsilon$ となるような $\delta > 0$ と $h > 0$ が存在する。

命題 3　活性化関数ユニット $f_i = \tanh$ である場合を考える。(a) 荷重行列 W が $\sigma_{max} = \Lambda < 1$ を満たすとする。ここで σ_{max} は最大の特異値である。このとき、すべての状態 $\mathbf{x}, \mathbf{x}' \in [-1, 1]^N$ について $d(T(\mathbf{x}, \mathbf{u}), T(\mathbf{x}', \mathbf{u})) < \Lambda d(\mathbf{x}, \mathbf{x}')$ となる。(b) 荷重行列がスペクトル半径 $|\lambda_{max}| > 1$ であるとする。ここで λ_{max} は W の最大の固有値である。このときネットワークは漸近的不安定 null 状態 (asymptotically unstable null state) を持つ。これは、このネットワークが任意の 0 を含む入力セット U、許容される状態セット $A = [-1, 1]^N$ に対して、echo states を持たないということを意味する。