

再帰型ニューラルネットワークによる信号推定に関して

松原 篤

製作技術課

1 はじめに

信号推定は、信号の時間的な変化を標的となる事象の振る舞いの本質として捉え、その傾向に基づき、事象を推測し定めることである。信号モデルは将来に亘る事象の推定に大きな役割を果たす要素である反面、モデル化が難しい信号、例を挙げると、①規則的あるいは不規則に振幅が大きく変化する信号(信号の消長)、②消滅などにより、その前後で位相の整合性が保証されなくなった信号(信号の不連続性・位相跳躍)のような場合の推定は一筋縄ではいかない。

本報告では、再帰型ニューラルネットワークを活用した信号推定について述べる。再帰型ニューラルネットワーク(Recurrent Neural Network: 以降は”RNN”と記述)¹⁾は、「入力層」、「隠れ層」、「出力層」から構成される標準的なニューラルネットワークに対し、隠れ層の入力に、入力層からの情報だけでなく、当該隠れ層や上位層において事前処理したデータに対する出力(状態)を活用することにより、一連のデータの前後の移り変わり・序列に”意味がある(Semantic)”ことを認識しつつ学習器を生成するネットワークである。事前のデータに対する出力を帰還利用することは、「フィードバック結合」と言われており、RNN の開発経緯には、自らの隠れ層の出力を接続した「Elman Net」²⁾や、出力層の結果を利用した「Jordan Net」³⁾などがある。フィードバック機構を有する RNN では、標準的なニューラルネットワークと異なり、時間的な順序の概念を持ち合わせているため、上記モデル化が難しい信号においても、過去の情報を活かした推定を有効に行うことができる。

報告に際し扱う時系列データとしては、①標準正規分布にしたがう雑音をイメージした乱数データを加えた正弦波信号、② 安静閉眼時に顕著に出現する自発脳波(α 波信号)の 2 種類とし、推定の実現は C++言語(コンパイラ:g++)を用いたプログラミングにより行った。

2 再帰型ニューラルネットワーク

「入力層」、「隠れ層」、「出力層」からなるネットワーク機構において、いま、時系列データ行列 X_{ori} (サイズ: $N_t \times N$)に対し、データ数 Q により N/Q 個のブロックに分割し抽出されたデータ行列を $X^{(p)}$ (サイズ: $N_t \times Q$, ブロック番号: $p \in \{0, \dots, N/Q - 1\}$)とすると、入力層($l = 0$)におけるデータの流れは、

$$\langle \text{入力} \rangle \quad {}^{(0)}U^{(p)} = X^{(p)} \quad \langle \text{出力} \rangle \quad {}^{(0)}Z^{(p)} = f_{(0)}[{}^{(0)}U^{(p)}] \quad \text{where } f_{(0)}: \text{恒等関数} \quad (1)$$

と表される。ただし、 $N_t [= \tau \times \Delta f]$ (Δf : サンプルレート)は時間 τ に対応するデータ点数である。

隠れ層についてはフィードバック機構が形成されるため、第 l 層に対する入出力は次のような形となる。

$$\langle \text{入力} \rangle \quad {}^{(l)}U^{(p)} = {}^{(l-1)}W \quad {}^{(l-1)}Z^{(p)} + {}^{(l-1)}B + \begin{cases} {}^{(l-1)}ElmW \quad {}^{(l-1)}Z^{(p-1)} \\ {}^{(l-1)}JorW \quad {}^{(l-1)}Z^{(p-1)} \end{cases} \quad (2)$$
$$\langle \text{出力} \rangle \quad {}^{(l)}Z^{(p)} = f_{(l)}[{}^{(l)}U^{(p)}] \quad \text{where } f_{(l)}: \text{双曲線正接関数}$$

ネットワークパラメータである ${}^{(l-1)}W$ および ${}^{(l-1)}B$ は、それぞれ隠れ層(第 l 層)の入力に作用する重み係数[サイズ: $UnitNum(l) \times UnitNum(l-1)$]およびバイアス項[サイズ: $UnitNum(l) \times Q$]であり、第 3 項に位置する ${}^{(l-1)}ElmW$ および ${}^{(l-1)}JorW$ は、Elman Net および Jordan Net とした場合のそれぞれの重み係数[サイズはそれぞれ

$UnitNum(l) \times UnitNum(l)$ および $UnitNum(l) \times UnitNum(l+1)$]であるが、併記した双方のネットワークは、通常、目的に応じて使い分けられる。また、出力 Z に関して、添え字 $p-1$ は、1つ前(時間的には1つ過去)のブロックのデータを利用したことを意味し、したがって ${}^{(l-1)}Z^{(p-1)}$ および ${}^{(l+1)}Z^{(p-1)}$ は、前層および次層における1つ前のブロックデータに対する出力を表している。なお、 $UnitNum(l)$ は第 l 層のユニット数を表す。

L 層ネットワークとした場合、出力層($l=L$)におけるデータの流れは、以下のようになる。

$$\langle \text{入力} \rangle {}^{(L)}U^{(p)} = {}^{(L-1)}W {}^{(L-1)}Z^{(p)} + {}^{(L-1)}B \quad \langle \text{出力} \rangle Y^{(p)} = {}^{(L)}Z^{(p)} = f_{(L)}[{}^{(L)}U^{(p)}] \quad (f_{(L)}: \text{恒等関数}) \quad (3)$$

${}^{(L-1)}W$ および ${}^{(L-1)}B$ のサイズは、それぞれ $UnitNum(L) \times UnitNum(L-1)$ および $UnitNum(L) \times Q$ 。

逆伝播に際し、ネットワークパラメータの更新は、データのブロック単位、換言すればバッチ処理ごとに行われるが、出力層のネットワークパラメータに関しては、 p 番目のブロック(バッチ)データ対し、

$$D^{(p)} = Y^{(p)} - Tar^{(p)}, \quad {}^{(L)}\Delta^{(p)} = (D^{(p)} + \alpha) \odot f'_{(L)}[{}^{(L)}U^{(p)}], \quad \alpha = [{}^{(L-2)}JorW]^T {}^{(L-1)}\Delta^{(p+1)}$$

$$\text{delta } {}^{(L-1)}W^{(p)} = {}^{(L)}\Delta^{(p)} [{}^{(L-1)}Z^{(p)}]^T, \quad \text{delta } {}^{(L-1)}B^{(p)} = {}^{(L)}\Delta^{(p)} \quad (4)$$

$${}^{(L-1)}W^{(p+1)} = {}^{(L-1)}W^{(p)} - \eta^{(p)} \text{delta } {}^{(L-1)}W^{(p)}, \quad {}^{(L-1)}B^{(p+1)} = {}^{(L-1)}B^{(p)} - \eta^{(p)} \text{delta } {}^{(L-1)}B^{(p)}$$

という手続きを踏む。 $D^{(p)}$ は出力層の出力 $Y^{(p)}$ と教師データ $Tar^{(p)}$ との差であり、デルタ量 ${}^{(L)}\Delta^{(p)}$ に付加する α 項はJordan Net 由来のものである。 $f'_{(L)}[\cdot]$ は活性化関数の導関数であり、 $\eta^{(p)}$ は学習率。また、パラメータの更新量 $\text{delta } {}^{(L-1)}W^{(p)}$ および $\text{delta } {}^{(L-1)}B^{(p)}$ の扱いに関し実際には、(4)式中央の結果を直接同式下段に活用するのではなく、Adam やモメンタムなど、学習効率を高める作用がある演算処理を経由して行うことが多い。

隠れ層のパラメータの更新については、次のような形となる。

$${}^{(l)}\Delta^{(p)} = ([{}^{(l)}W]^T {}^{(l+1)}\Delta^{(p)} + \beta) \odot f'_{(l)}[{}^{(l)}U^{(p)}], \quad \beta = [{}^{(l-1)}ElmW]^T {}^{(l)}\Delta^{(p+1)}$$

$$\text{delta } {}^{(l-1)}W^{(p)} = \sum_c^{c_{max}} {}^{(l)}\Delta^{(p-c)} [{}^{(l-1)}Z^{(p-c)}]^T, \quad \text{delta } {}^{(l-1)}B^{(p)} = \sum_c^{c_{max}} {}^{(l)}\Delta^{(p-c)}$$

$$\text{delta } {}^{(l-1)}ElmW^{(p)} = \sum_c^{c_{max}} {}^{(l)}\Delta^{(p-c)} [{}^{(l)}Z^{(p-c-1)}]^T, \quad \text{delta } {}^{(l-1)}JorW^{(p)} = \sum_c^{c_{max}} {}^{(l)}\Delta^{(p-c)} [{}^{(l+1)}Z^{(p-c-1)}]^T \quad (5)$$

$${}^{(l-1)}W^{(p+1)} = {}^{(l-1)}W^{(p)} - \eta^{(p)} \text{delta } {}^{(l-1)}W^{(p)}, \quad {}^{(l-1)}B^{(p+1)} = {}^{(l-1)}B^{(p)} - \eta^{(p)} \text{delta } {}^{(l-1)}B^{(p)}$$

$${}^{(l-1)}ElmW^{(p+1)} = {}^{(l-1)}ElmW^{(p)} - \eta^{(p)} \text{delta } {}^{(l-1)}ElmW^{(p)}$$

$${}^{(l-1)}JorW^{(p+1)} = {}^{(l-1)}JorW^{(p)} - \eta^{(p)} \text{delta } {}^{(l-1)}JorW^{(p)}$$

ただし、 β 項はElman Net 由来のものである。パラメータの更新量のサマーシオンにおいて、添え字 $p-c$ は現在のブロックデータ $X^{(p)}$ から過去 c ブロック分のデータ $X^{(p-c)}$ までを考慮することを意味しており、簡素な例を挙げると、 $\sum_c^d {}^{(l)}\Delta^{(p-c)}$ は、第 l 層において現在のブロックデータ $X^{(p)}$ から過去 d のブロックデータ $X^{(p-d)}$ までの $(d+1)$ 個分のデルタ量の和を算出する。なお、実際の更新量の演算に関しては、前記同様、Adam やモメンタムなどの処理を経由する。

学習のための評価指標として、以下に示す誤差関数 E がよく採用される。上記 $D^{(p)} = [d_0^{(p)} \dots d_q^{(p)} \dots d_{Q-1}^{(p)}]$ に対し、

$$E(\text{epoch}) = \frac{1}{2} \sum_{p=0}^{N-1} \left[\sum_{q=0}^{Q-1} \{d_q^{(p)}\}^T d_q^{(p)} \right] \Rightarrow \min_{\text{epoch}} E(\text{epoch}) \quad (6)$$

(epoch : 学習回数)

学習の進展方向は、誤差関数 E が単調減少し最小化する向きとなり、学習は適度な設定条件にしたがい、打ち切られる。

3 解析

3.1 データ構成

学習および推論における信号推定のイメージを、それぞれ図 1 に示す。推定の骨子は、現時刻より過去ウィンドウ長 τ 分の時系列データを活用して次時刻の状態を確定することである。ただし、テストデータによる推論は、学習と異なり、その都度確定した推

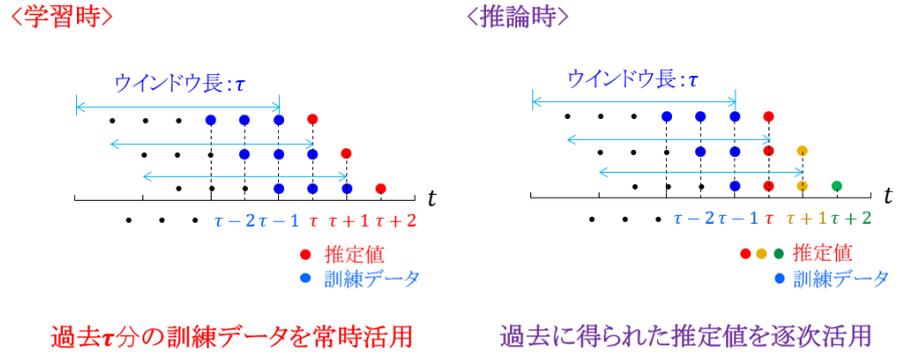


図 1. 推定イメージ

定値を、次の推定に活用することになっている。これを踏まえると、時刻 $t \geq 2\tau$ における推定は、ウィンドウ長 τ 分すべて、過去に確定された推定値を活用して行うことになる。

以上を考慮すると実際は、データ行列 X_{ori} のサイズは $N_\tau \times (N - N_\tau + 1)$ 、分割したブロック数は $(N - N_\tau + 1)/Q$ となる($p \in \{0, \dots, (N - N_\tau + 1)/Q - 1\}$)。また、バッチ処理とした場合、1 ブロックのデータに対し、 Q 個分の推定値が一度に得られることになる。参考のため、 p 番目のブロックデータ $X^{(p)}$ (サイズ: $N_\tau \times Q$)および得られる推定値 $\hat{Y}^{(p)}$ は次のように表現される。

$$X^{(p)} = \begin{bmatrix} x_{pQ}^{(p)} & \cdots & x_{pQ+(Q-1)}^{(p)} \\ \vdots & \ddots & \vdots \\ x_{pQ+(N_\tau-1)}^{(p)} & \cdots & x_{pQ+(N_\tau-1)+(Q-1)}^{(p)} \end{bmatrix} \quad (7)$$

$$\hat{Y}^{(p)} = [\hat{y}_{pQ+N_\tau}^{(p)} \cdots \hat{y}_{pQ+N_\tau+(Q-1)}^{(p)}]$$

3.2 設定条件

表 1. ネットワーク運用のための諸条件

	正弦波信号	脳波信号
訓練データ	10Hz 正弦波+標準正規分布にしたがう雑音	後頭部安静閉眼時データ(複数部位)
教師データ	10Hz 正弦波	α 波信号(フィルタ出力)
テストデータ	10Hz位相差信号($\pi/4, \pi/2$) + 上記雑音	後頭部安静閉眼時データ
サンプルレート Δf (Hz)	200	200
全データ長(時間&点数)	$0.1s \times 3cycles (N = 60)$	$1s (N = 200)$
誤差関数	(6)式参照	$E_{all}(epoch) = \sum_{ch=0}^{CH-1} E_{ch}(epoch)$

ネットワークの運用のための条件を表 1 に示す。正弦波信号の場合のテストデータには、訓練データに対し位相差をもつものを 2 種類用意した。雑音の大きさについては、実効値レベルとして正弦波信号の7%程度に統一している。誤差関数においては、原則、(6)式にしたがうものであるが、第 1 サメーションの上限は、 $(N - N_\tau + 1)/Q - 1$ となる。

安静閉眼時に α 波が大きく観測される後頭部の 2 つの部位:左後頭部 O1, 右後頭部 O2 から取得される信号は同相である。このため、脳波信号による学習においては、2 チャネル入力に対応した学習器となる。ただし、各観測部位から取得される信号には、標的の α 波以外の脳波や雑音等が含まれるため、実際の訓練、テストおよび教師データには、観測信号に対しフィルタ処理を施したものを採用している。さらに、脳波信号の場合の誤差関数 E_{all} については、観測部位ごとに算出された誤差関数 E_{ch} の和を評価値とした。

入力層のユニット数はデータウィンドウ長に対応したデータ点数 N_τ と等価であるとし、ブロック単位で考えた場合(バッチ処理)の入力数は $N_\tau \times Q$ となる。出力層のユニット数は 1 個とするため、ブロック単位での個数は $1 \times Q$ 。また、訓

練データ数は $N - N_\tau + 1$ となるが、例を挙げると、 $N_\tau = 6$ の場合(正弦波信号に対し 1/4 周期分に相当)、正弦波信号および脳波信号では、それぞれ 55 個および 195 個となる。隠れ層のユニット数は入力層のユニット数の 8 掛けとした。

(5)式に示した隠れ層に関わるネットワークパラメータの更新量に関するサメーションの上限 C_{max} については、過去いくつか固定数分のブロックデータを活用する際、固定数に満たない場合を配慮し、

$$C_{max} = \begin{cases} p & (0 \leq p \leq \frac{N_\tau - 1}{Q}) \\ \frac{N_\tau - 1}{Q} & (\frac{N_\tau - 1}{Q} < p < \frac{N - N_\tau + 1}{Q}) \end{cases} \begin{cases} \text{delta}^{(l-1)}W^{(p)}, & \text{delta}^{(l-1)}B^{(p)} \\ \text{delta}^{(l-1)}ElmW^{(p)}, & \text{delta}^{(0)}JorW^{(p)} \\ \text{common} \end{cases}$$

としている。

ネットワークパラメータの更新に際しては Adam を適用し、学習の打ち切りに関する条件としては、適度な上限に達した場合(e.g. $epoch = 700 \sim 800$)と、直前の学習による誤差関数との差がない場合(収束条件: $\varepsilon = 1.0 \times 10^{-12}$)のいずれかの成立を採用した。

3.3 結果および検討

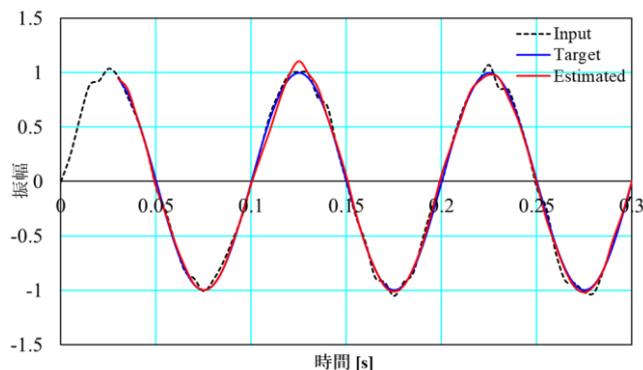
本節では、ネットワーク構成を Elman 型(隠れ層帰還系)として行っている。

3.3.1 正弦波信号

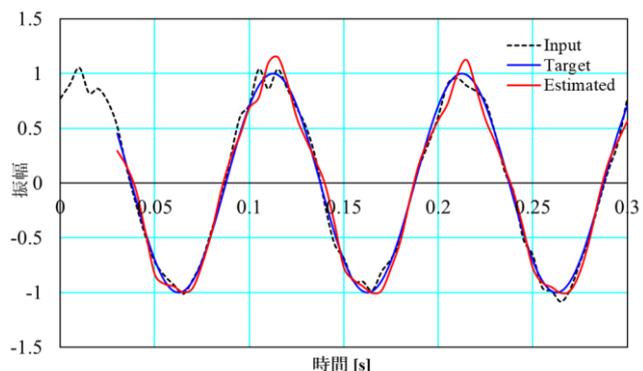
2 層(隠れ層は 1 つ)の RNN 構成において、データパラメータ $(N_\tau, Q) = (6, 5)$ に対する訓練データおよびテストデータ(位相差 $\pi/4, \pi/2$)の推定結果を図 2 に示す。図では、雑音入り正弦波入力信号、教師用正弦波信号(Target)および推定波形を併記すると共に、推定波形の評価のため、教師信号との誤差や相関度を、[相関係数 / 誤差の二乗和]という形で算出・表記している。

図より、テストデータでは、訓練データの場合に比べ、時系列において振幅の変化量の極性が変わる極値近傍など、教師信号との食い違いが所々散見されるものの、周期性が充分保たれており(相関性の保証)、位相差を有する信号においても、良質な推定が行える様子が伺える(再現性の保証)。

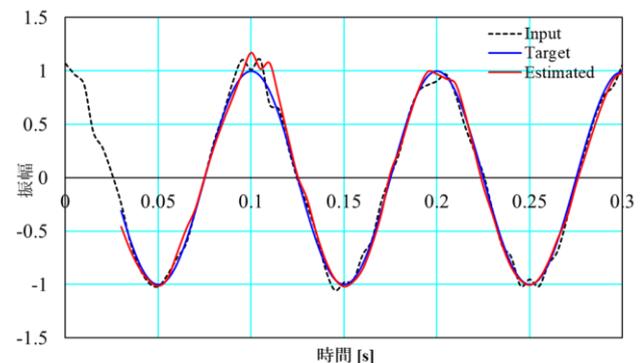
ウィンドウ長に対応したデータ点 N_τ に対する推定波形の相関性および再現性を表す評価値を図 3 に 2 軸グラフとして示す。ただし、 $Q=5$ とした場合であり、3 層のネットワークは 2 つの隠れ層を有したものである。図より、相関性、再現性、いずれの側面から、3 層ネットワークは 2 層の場合よりやや劣る傾向にあるように思われる。詳細は割愛するが、1 ブロックのデータ数 Q を変数として諸々検討した結果をあわせると、ネットワークの構成や入力データの位相



(a) 訓練データ [0.998/0.0616]



(b) テストデータ(位相差 $\pi/4$) [0.993/0.343]



(c) テストデータ(位相差 $\pi/2$) [0.995/0.278]

[*/*] = [Correlation Coefficient/Sum of Squared Errors]

図 2. 正弦波信号の推定波形

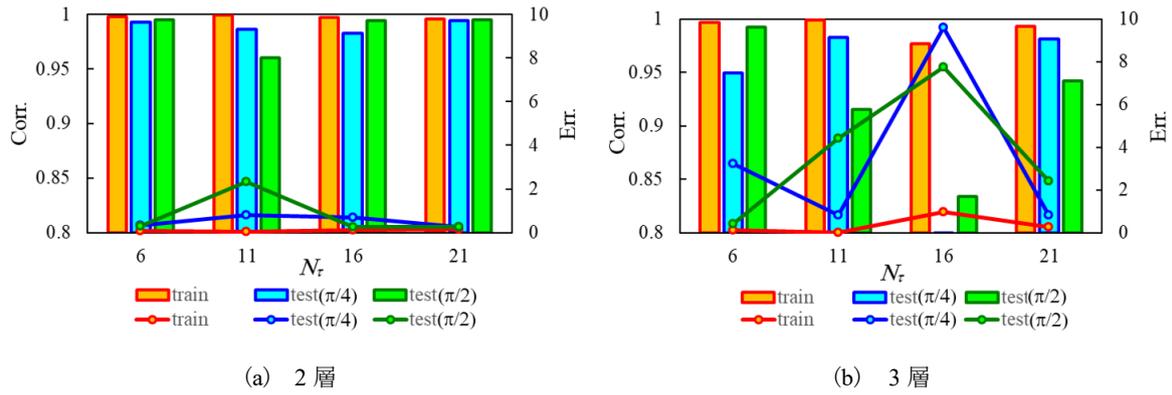


図 3. データ点 N_τ に対する推定波形の評価

状態の程度に依存することは無論のことながら、相関性および再現性を満足する条件のひとつに、『 $N_\tau < T$ (信号の周期), かつ $N_\tau \approx Q$ 』の成立があるように思われる. 特に条件の後者については, (7)式において, 入力層のユニットへのデータ $\mathbf{x}_{N_\tau}^{(p)} = [x_{pQ}^{(p)}, \dots, x_{pQ+(N_\tau-1)}^{(p)}]^T$ と1ブロック Q 内の各先頭データを意味する $\mathbf{x}_Q^{(p)} = [x_{pQ}^{(p)}, \dots, x_{pQ+(Q-1)}^{(p)}]^T$ が, 同じ時間幅で, 同値・同相(位相差なし)となることを指しており, このような情報量のバランスの良さも学習器の汎化性能を高める要素になるものと推察される.

3.3.2 脳波信号

2層ネットワークに際し, 2つの観測部位のデータに対する訓練データおよびテストデータの推定結果 $[(N_\tau, Q) = (6, 5)]$ を図4に示す. ネットワークに対する入力信号および教師信号(Target)はそれぞれアーチファクトを除去した(遮断周波数: 3Hz, 30Hz), および α 帯域の信号のみを抽出した(遮断周波数: 5Hz, 17Hz)第4次バターズ型バンドパスフィルタ出力波形とした. また, $[*/*]$ は, 正弦波信号の場合と同様, [相関係数/誤差の二乗和]を示す. 図より, 各部位共に, テストデータによる推論では, 訓練データによる学習に比べ, 比較的高い周波数のリップル(ripple)現象が見られているものの, 比較的高い相関と良好な再現がなされることがわかる.

また, 各部位の脳波推定波形に対する評価を図5に示すが, テストデータによる推論では, 相関度および再現度が小さくなる条件が多少あるものの, 観測部位間において極端な差異もなく, 正弦波信号の場合と同様, 総じて良質な

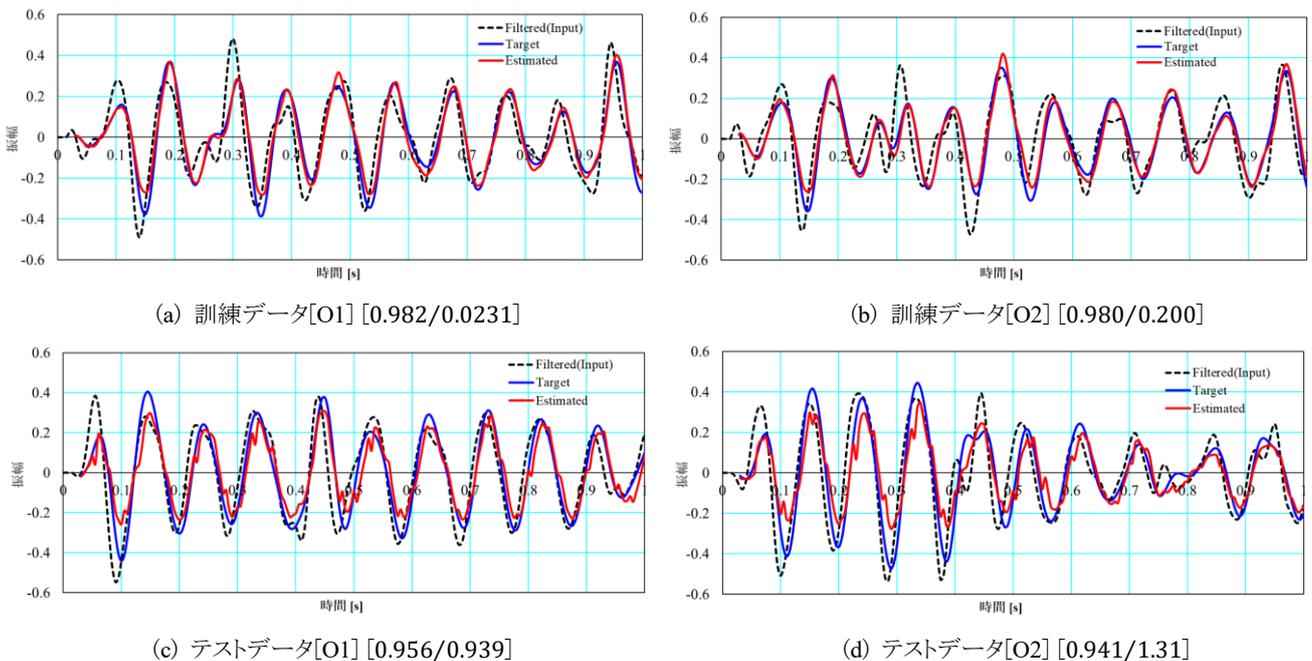
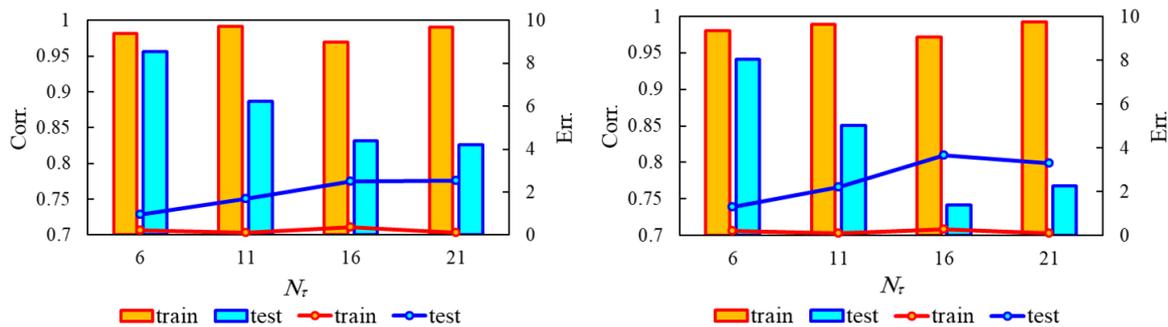


図 4. 脳波信号の推定波形



(a) 左後頭部 O1

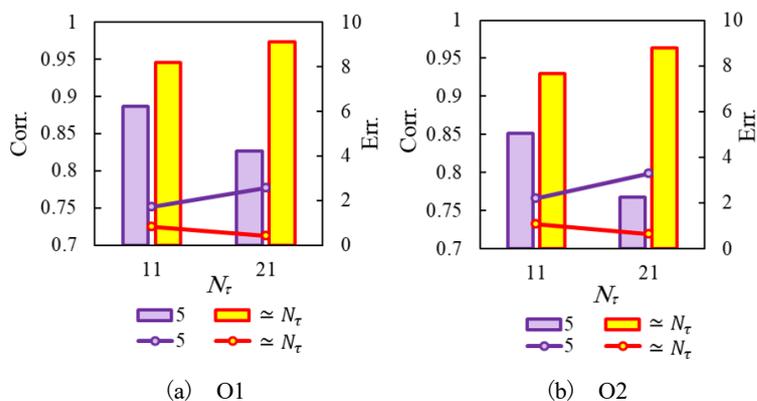
(b) 右後頭部 O2

図 5. データ点 N_t に対する脳波推定波形の評価 (2 層ネットワーク)

推定が行える様子が見て取れる。

加えて図 6 に、テストデータに対し $Q \approx N_t$ [具体的には, $(N_t, Q) = (11, 10), (21, 20)$] のときの評価状況を示すが、 $Q = 5$ の場合に比べ $Q \approx N_t$ の条件の方が良好であると言える。したがって、脳波信号においても、正弦波信号の場合と同様、条件「 $Q \approx N_t$ 」が学習器の汎化性能を高める条件のひとつであることと想定される。

なお、3 層ネットワークについても、上述の 2 層の場合と同程度のレベルの結果が得られることを確認している。



(a) O1

(b) O2

図 6. 推論時の (N_t, Q) に対する評価

4 おわりに

本報告で扱った脳波信号は、①その振る舞いにゆらぎがある、②個人差が顕著に現れる、③同一人物においても日々大きく変化する、といったことが一般的であるため、先述したように信号の消長や、消滅などによる位相の不連続性が観測される可能性が高く、それゆえモデル化が実現しづらい信号である。同じくして、現実を考慮すれば安易な条件設定ではあったものの、扱った正弦波信号は、学習器の雑音や位相の異なる信号への適合という点において、一定の有益な知見をもたらす効果的な実験材料であったものと思われる。

最後に、実装した RNN 解析プログラムでは、Elman Net と Jordan Net を切り替えることができ、また必要に応じて双方を組み合わせたことができるよう、構成している。ただし、後者においては、3 層以上 ($L > 2$) の場合、第 $(L - 2)$ 層以下の隠れ層に対する接続元は、実際には出力層ではなく、1 つ上位の層(次の隠れ層)となるよう、改良した形にしている(前記式の Jordan Net 項は改良表記)。なお、Elman Net はこれまで述べてきた「時系列データ」や、また「言語処理」などに、一方で、Jordan Net は「運動制御」等に、それぞれ性能を発揮するネットワークと言われている。

参考文献

- 1) 岡谷貴之, 深層学習 改訂第 2 版, 講談社, pp.125-132 (2022)
- 2) Elman. J. L., Finding Structure in Time, Cognitive Science, 14, pp.179-211 (1990)
- 3) Jordan. M. I., Serial Order: A Parallel Distributed Processing Approach (ICS Report 8604), San Diego: Institute for Cognitive Science, pp.149-191 (1986)