長・短期依存学習型ニューラルネットワークによる時系列データ予測に関する検討

松原 篤

製作技術課

1　はじめに

人は，目や耳といった感覚器官より獲得した外部刺激を電気信号として伝送媒体である神経を通じて脳内に送り込むことにより，感覚として自覚・認識する．脳内における電気的情報の伝達は神経細胞(ニューロン)間の興奮の伝わりに基づいているが，伝達される情報の確保(記銘)，保持および抽出(想起)については，“記憶”の働きにおける 3 大要素とも言われている．これらの要素は，脳内では海馬および大脳皮質と呼ばれる部位において担われており，前者は一時的な記憶に，換言すれば，“短期記憶(short term memory)”に，後者は学習を反復するなどして永続的な記憶に，総じて言えば，“長期記憶(long term memory)”に，それぞれ関わりを深く持っている．

人の脳内における情報伝達に対する神経細胞の働き・仕組みを模したニューラルネットワークは，情報を，①受容する“入力層”，②制御・調整する“隠れ層”，③供与する“出力層”からなる階層構造を形成しており，ネットワーク上の情報伝達(伝播)において，神経細胞の興奮状態は“活性化関数”とした作用素に，また，隣接する神経細胞を接続するシナプスの結合の強さは“重み”としたネットワークパラメータに，それぞれ置換される．しかしながら，前記入力層，隠れ層および出力層から構成される“標準型”のニューラルネットワークでは，“記憶”に関連した機能を有していない．

一方，現在の情報だけでなく過去の結果を処理に活かす方法には，隠れ層における情報処理の流れ(フロー)において帰還(フィードバック)システムを取り入れた“再帰型ニューラルネットワーク(Recurrent Neural Network: 以降 RNN) ”がある [1)]．さらに，RNN の拡張型として，隠れ層の（神経細胞に見立てた）各ユニットに対する情報の保持や流れを，記憶セル(メモリセル)や専用のゲートを設けることにより，実施・調整することが可能な“長・短期記憶型ニューラルネットワーク(Long Short-Term Memory：以降 LSTM) ”がある [1)]．LSTM では，RNN の特徴である短期的な依存関係の学習(短期記憶)だけでなく，長期的な依存関係(長期記憶)を学習することも可能であり，その根拠は，メモリセルに対し，①入力情報の採用に関する制御を担う「入力ゲート」，②保存された過去情報の取捨選択を行う「忘却ゲート」，③隣接する上位層への情報伝達を制御する「出力ゲート」といった 3 種のゲートの働きによるところがある [2), 3)]．

本報告では，正弦波信号や脳波信号(α波帯域)を対象とした LSTM である長短記憶型ニューラルネットワークによる学習器の予測・推定性能に関し，先に報告の再帰型ニューラルネットワーク(RNN)による方法 [4)]と比較しながら，種々検討したことについて述べる．なお，プログラム実装は，開発環境「Sublime Text 3」下において C++言語[コンパイラ：g++ (MinGW6.3.0)]により行った．

2　再帰型ニューラルネットワーク(RNN)

先述したように，LSTM は RNN を基盤とした拡張型ネットワークシステムであるため，本章では，RNN の構成の骨子について述べる．

時系列データにより構成される行列$X_o \in \mathbb{R}^{N_\tau \times N}$において，データ数$Q$により$N/Q$個のデータセグメントに分割したとき，$p$番目に位置するデータセグメント行列を$X^{(p)} \in \mathbb{R}^{N_\tau \times Q}$(データセグメントインデックス： $p \in \{0, \cdots, N/Q-1\}$)とする．いま，データの流れがネットワークの順伝播にしたがう場合，入力層(便宜上，第 0 層とする)については，次のような入出力関係が成立する．ただし，N_τは時間τに対応するデータ点数を表す[i.e. $N_\tau = \tau f_{sp}$　(f_{sp}: サンプルレート)]．

$$< 入力 > \ {}^{(0)}U^{(p)} = X^{(p)} \quad < 出力 > \ {}^{(0)}Z^{(p)} = f_{(0)}\left[{}^{(0)}U^{(p)}\right] \qquad \text{where } f_{(0)}: id \text{ (identity function)} \tag{1}$$

隠れ層(第l層)については，フィードバック機構が実装されるため，

$$
\begin{aligned}
&< \text{入力} > {}^{(l)}U^{(p)} = {}^{(l-1)}W\ {}^{(l-1)}Z^{(p)} + {}^{(l-1)}B + {}^{(l-1)}W'\ {}^{(l)}Z^{(p-1)} \\
&< \text{出力} > {}^{(l)}Z^{(p)} = f_{(l)}\left[{}^{(l)}U^{(p)}\right] \qquad \text{where} \quad f_{(l)}\text{: Hyperbolic tangent function}
\end{aligned}
\tag{2}
$$

と表される．ネットワークパラメータ行列 ${}^{(l-1)}W \in \mathbb{R}^{N_l \times N_{l-1}}$および${}^{(l-1)}B \in \mathbb{R}^{N_l \times Q}$は，それぞれ重み係数およびバイアス項であり，第 3 項にみられる${}^{(l-1)}W' \in \mathbb{R}^{N_l \times N_l}$は，信号伝達において，処理の序列として 1 つ前$(p-1)$のデータセグメントによる結果${}^{(l)}Z^{(p-1)}$と接続するフィードバック項としての役割を果たす重み係数である．N_{l-1}およびN_lは，それぞれ第$l-1$層および第l層のユニット数である．

L層ネットワークとした場合，出力層(第L層)については，次のように表現される．

$$
\begin{aligned}
&< \text{入力} > {}^{(L)}U^{(p)} = {}^{(L-1)}W\ {}^{(L-1)}Z^{(p)} + {}^{(L-1)}B \\
&< \text{出力} > Y^{(p)} = {}^{(L)}Z^{(p)} = f_{(L)}\left[{}^{(L)}U^{(p)}\right] \quad \text{where}\ \ f_{(L)}\text{: } id \text{ (identity function)}
\end{aligned}
\tag{3}
$$

各パラメータについては，${}^{(L-1)}W \in \mathbb{R}^{N_L \times N_{L-1}}$および${}^{(L-1)}B \in \mathbb{R}^{N_L \times Q}$である．

他方，逆伝播を考えた場合，出力層におけるネットワークパラメータの更新は，次のような手続きをとる．

$$
\begin{aligned}
&D^{(p)} = Y^{(p)} - T^{(p)}, \qquad {}^{(L)}\Delta^{(p)} = D^{(p)} \odot f'_{(L)}\left[{}^{(L)}U^{(p)}\right] \\
&\text{delta}\ {}^{(L-1)}W^{(p)} = {}^{(L)}\Delta^{(p)}\left[{}^{(L-1)}Z^{(p)}\right]^T, \qquad \text{delta}\ {}^{(L-1)}B^{(p)} = {}^{(L)}\Delta^{(p)} \\
&{}^{(L-1)}W^{(p+1)} = {}^{(L-1)}W^{(p)} - \eta^{(p)}\text{delta}\ {}^{(L-1)}W^{(p)}, \quad {}^{(L-1)}B^{(p+1)} = {}^{(L-1)}B^{(p)} - \eta^{(p)}\text{delta}\ {}^{(L-1)}B^{(p)}
\end{aligned}
\tag{4}
$$

$D^{(p)}$は出力層の出力$Y^{(p)}$と教師データ$T^{(p)}$との差であり，デルタ量 ${}^{(L)}\Delta^{(p)}$は入力${}^{(L)}U^{(p)}$に対するネットワークの誤差関数E(後述)の変化量(1 次微分)を表している．また，delta は各ネットワークパラメータの更新量を意味しており，$f'_{(L)}[\cdot]$は活性化関数の導関数，$\eta^{(p)}$は学習率である．

隠れ層については，フィードバック項を含め 3 種のネットワークパラメータの更新となる．

$$
\begin{aligned}
&{}^{(l)}\Delta^{(p)} = \left(\left[{}^{(l)}W\right]^T {}^{(l+1)}\Delta^{(p)} + \left[{}^{(l-1)}W'\right]^T {}^{(l)}\Delta^{(p+1)}\right) \odot f'_{(l)}\left[{}^{(l)}U^{(p)}\right] \\
&\text{delta}\ {}^{(l-1)}W^{(p)} = \sum_{c}^{C_{max}} {}^{(l)}\Delta^{(p-c)}\left[{}^{(l-1)}Z^{(p-c)}\right]^T, \ \ \text{delta}\ {}^{(l-1)}B^{(p)} = \sum_{c}^{C_{max}} {}^{(l)}\Delta^{(p-c)} \\
&{}^{(l-1)}W^{(p+1)} = {}^{(l-1)}W^{(p)} - \eta^{(p)}\text{delta}\ {}^{(l-1)}W^{(p)}, \quad {}^{(l-1)}B^{(p+1)} = {}^{(l-1)}B^{(p)} - \eta^{(p)}\text{delta}\ {}^{(l-1)}B^{(p)} \\
&\text{delta}\ {}^{(l-1)}W'^{(p)} = \sum_{c}^{C_{max}} {}^{(l)}\Delta^{(p-c)}\left[{}^{(l)}Z^{(p-c-1)}\right]^T, \quad {}^{(l-1)}W'^{(p+1)} = {}^{(l-1)}W'^{(p)} - \eta^{(p)}\text{delta}\ {}^{(l-1)}W'^{(p)}
\end{aligned}
\tag{5}
$$

(5)式に散見される summation (総和)の部分については，例を挙げると，$\sum_c^{Cmax} {}^{(l)}\Delta^{(p-c)}$は，第$l$層$p$番目のデータセグメントに対するデルタ量${}^{(l)}\Delta^{(p)}$から過去$Cmax$前のデータセグメントに対するデルタ量${}^{(l)}\Delta^{(p-Cmax)}$まで，合計$Cmax+1$個のデルタ量を保持し加算していることを意味している．

3 長・短期依存学習型ニューラルネットワーク(LSTM)

LSTM の構成については，RNN において隠れ層を構成する各ユニットを，先述のメモリセルおよび入力，忘却，出力の各ゲートに置換したものとなる．したがって，入力層および出力層の仕組みについては RNN に準拠しているため，本章では，隠れ層の構成にのみ焦点を当て，解説する．

ネットワークの順伝播の流れに際し，LSTM の隠れ層(第l層)における前層(第$l-1$層とし，入力層，隠れ層，いずれの場合でも可)からの情報の扱いに関する最初の手続きは，(2)式に示した RNN の隠れ層の入出力系に準拠する．

$$
\begin{gathered}
{}^{(l)}U^{(p)} = {}^{(l-1)}W\ {}^{(l-1)}Z^{(p)} + {}^{(l-1)}B + {}^{(l-1)}W'\ {}^{(l)}Z^{(p-1)} \qquad \rightarrow \qquad {}^{(l)}g\left[{}^{(l)}U^{(p)}\right] \\
\text{where} \quad {}^{(l)}g\text{: Sigmoid function}
\end{gathered}
\tag{6}
$$

ただし，${}^{(l-1)}W \in \mathbb{R}^{N_l \times N_{l-1}}$，${}^{(l-1)}B \in \mathbb{R}^{N_l \times Q}$，${}^{(l-1)}W' \in \mathbb{R}^{N_l \times N_l}$である．

次に，入力および忘却の各ゲート[$\alpha \in \{in, fgt\}$]およびメモリセルの状態${}^{(l)}S^{(p)}$について，処理の流れは

$$
\begin{aligned}
&{}^{(l)}U_\alpha^{(p)} = {}^{(l-1)}W_\alpha\ {}^{(l-1)}Z^{(p)} + {}^{(l-1)}B_\alpha + {}^{(l-1)}W'_\alpha\ {}^{(l)}Z^{(p-1)} + {}^{(l)}peepW_\alpha\ {}^{(l)}S^{(p-1)} \\
&{}^{(l)}Z_\alpha^{(p)} = {}^{(l)}f_\alpha\left[{}^{(l)}U_\alpha^{(p)}\right] \qquad\qquad \text{where} \quad {}^{(l)}f_\alpha\text{: Hyperbolic tangent function}
\end{aligned}
\tag{7}
$$

$$
{}^{(l)}S^{(p)} = {}^{(l)}Z_{in}^{(p)} \odot {}^{(l)}g\left[{}^{(l)}U^{(p)}\right] + {}^{(l)}Z_{fgt}^{(p)} \odot {}^{(l)}S^{(p-1)} \tag{8}
$$

となる．${}^{(l)}peepW_\alpha \in \mathbb{R}^{N_l \times N_l}$は各々のゲートに対する覗き穴結合に関わる重み係数である．覗き穴結合(peephole connections) [5]はメモリセルの状態を直接ゲートの制御に活かすよう作用する接続形態であり，入力および忘却の各ゲートに際しては，メモリセルの過去の情報${}^{(l)}S^{(p-1)} \in \mathbb{R}^{N_l \times Q}$を，あるタイミングで，それぞれ“入力に活用させる”および“忘却させる”働きがある．

最後に，出力ゲートによる作用を考慮すると，メモリセルからの出力${}^{(l)}Z^{(p)}$は，次のようになる．

$$
\begin{aligned}
&{}^{(l)}U_{out}^{(p)} = {}^{(l-1)}W_{out}\ {}^{(l-1)}Z^{(p)} + {}^{(l-1)}B_{out} + {}^{(l-1)}W'_{out}\ {}^{(l)}Z^{(p-1)} + {}^{(l)}peepW_{out}\ {}^{(l)}S^{(p)} \\
&{}^{(l)}Z_{out}^{(p)} = {}^{(l)}f_{out}\left[{}^{(l)}U_{out}^{(p)}\right] \qquad\qquad \text{where} \quad {}^{(l)}f_{out}\text{: Hyperbolic tangent function}
\end{aligned}
\tag{9}
$$

$$
{}^{(l)}Z^{(p)} = {}^{(l)}Z_{out}^{(p)} \odot {}^{(l)}h\left[{}^{(l)}S^{(p)}\right] \qquad \text{where} \quad {}^{(l)}h\text{: Sigmoid function} \tag{10}
$$

出力ゲートに関わる覗き穴結合は，現在のメモリセルの情報${}^{(l)}S^{(p)} \in \mathbb{R}^{N_l \times Q}$を，任意のタイミングで上位層に伝達させる役割を担っている．以上，LSTM の隠れ層における順伝播の情報の流れを簡略図で表すと，図 1 のようになる．

他方，逆伝播について，入力，忘却，出力の 3 種のゲートに対するデルタ量${}^{(l)}\Delta_\alpha^{(p)} \in \mathbb{R}^{N_l \times Q}$ [$\alpha \in \{in, fgt, out\}$]，

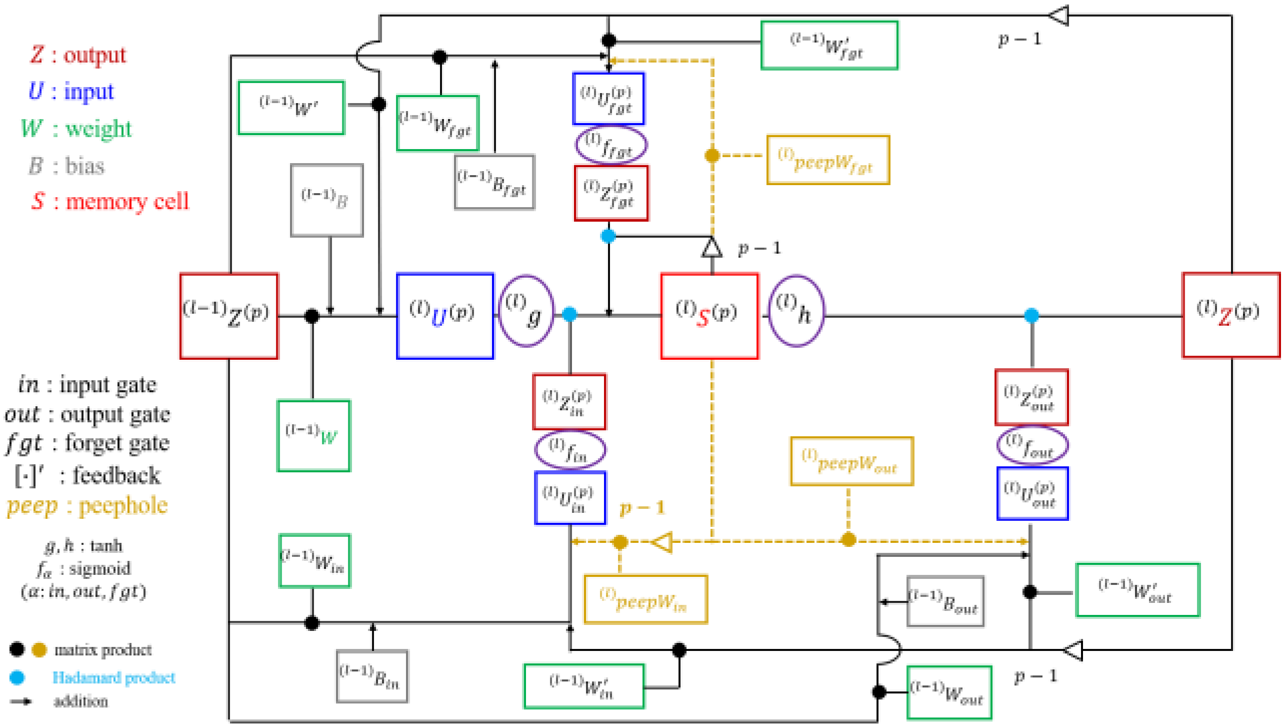


図 1.LSTM 構成(順伝播)

および活性化関数g, h，メモリセルSに関わるデルタ量${}^{(l)}\Delta_{\beta}^{(p)} \in \mathbb{R}^{N_l \times Q}$　$[\beta \in \{g, h, s\}]$は，ネットワークパラメータの更新手続きの序列にしたがい，それぞれ次のように表される．

$$
\begin{aligned}
&{}^{(l)}\Delta_{out}^{(p)} = {}^{(l)}f'_{out}\left[{}^{(l)}U_{out}^{(p)}\right] \odot {}^{(l)}h\left[{}^{(l)}S^{(p)}\right] \odot \left(\left[{}^{(l)}W\right]^{T\,(l+1)}\Delta^{(p)} + \left[{}^{(l-1)}W'\right]^{T\,(l)}\Delta^{(p+1)}\right) \\
&{}^{(l)}\Delta_{h}^{(p)} = {}^{(l)}f_{out}\left[{}^{(l)}U_{out}^{(p)}\right] \odot {}^{(l)}h'\left[{}^{(l)}S^{(p)}\right] \odot \left(\left[{}^{(l)}W\right]^{T\,(l+1)}\Delta^{(p)} + \left[{}^{(l-1)}W'\right]^{T\,(l)}\Delta^{(p+1)}\right) \\
&{}^{(l)}\Delta_{s}^{(p)} = {}^{(l)}\Delta_{h}^{(p)} + \\
&\qquad {}^{(l)}peepW_{in}\,{}^{(l)}\Delta_{in}^{(p+1)} + {}^{(l)}peepW_{fgt}\,{}^{(l)}\Delta_{fgt}^{(p+1)} + {}^{(l)}peepW_{out}\,{}^{(l)}\Delta_{out}^{(p)} + {}^{(l)}\Delta_{s}^{(p+1)} \odot {}^{(l)}Z_{fgt}^{(p+1)} \\
&{}^{(l)}\Delta_{g}^{(p)} = {}^{(l)}\Delta_{s}^{(p)} \odot {}^{(l)}f_{in}\left[{}^{(l)}U_{in}^{(p)}\right] \odot {}^{(l)}g'\left[{}^{(l)}U^{(p)}\right] \\
&{}^{(l)}\Delta_{in}^{(p)} = {}^{(l)}\Delta_{s}^{(p)} \odot {}^{(l)}f'_{in}\left[{}^{(l)}U_{in}^{(p)}\right] \odot {}^{(l)}g\left[{}^{(l)}U^{(p)}\right] \\
&{}^{(l)}\Delta_{fgt}^{(p)} = {}^{(l)}\Delta_{s}^{(p)} \odot {}^{(l)}f'_{fgt}\left[{}^{(l)}U_{fgt}^{(p)}\right] \odot {}^{(l)}S^{(p-1)}
\end{aligned}
\tag{11}
$$

このとき，メモリセルSに関わるネットワークパラメータの更新については

$$
\begin{aligned}
\text{delta}\ {}^{(l-1)}W^{(p)} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{g}^{(p-c)}\left[{}^{(l-1)}Z^{(p-c)}\right]^{T}, &\quad {}^{(l-1)}W^{(p+1)} &= {}^{(l-1)}W^{(p)} - \eta^{(p)}\text{delta}\ {}^{(l-1)}W^{(p)} \\
\text{delta}\ {}^{(l-1)}B^{(p)} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{g}^{(p-c)}, &\quad {}^{(l-1)}B^{(p+1)} &= {}^{(l-1)}B^{(p)} - \eta^{(p)}\text{delta}\ {}^{(l-1)}B^{(p)} \\
\text{delta}\ {}^{(l-1)}W'^{(p)} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{g}^{(p-c)}\left[{}^{(l)}Z^{(p-c-1)}\right]^{T}, &\quad {}^{(l-1)}W'^{(p+1)} &= {}^{(l-1)}W'^{(p)} - \eta^{(p)}\text{delta}\ {}^{(l-1)}W'^{(p)}
\end{aligned}
\tag{12}
$$

となり，また，各ゲート$[\alpha \in \{in, fgt, out\}]$に対するネットワークパラメータの更新は，

$$
\begin{aligned}
\text{delta}\ {}^{(l-1)}W_{\alpha}^{(p)} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{\alpha}^{(p-c)}\left[{}^{(l-1)}Z^{(p-c)}\right]^{T}, &\quad {}^{(l-1)}W_{\alpha}^{(p+1)} &= {}^{(l-1)}W_{\alpha}^{(p)} - \eta_{\alpha}^{(p)}\text{delta}\ {}^{(l-1)}W_{\alpha}^{(p)} \\
\text{delta}\ {}^{(l-1)}B_{\alpha}^{(p)} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{\alpha}^{(p-c)}, &\quad {}^{(l-1)}B_{\alpha}^{(p+1)} &= {}^{(l-1)}B_{\alpha}^{(p)} - \eta_{\alpha}^{(p)}\text{delta}\ {}^{(l-1)}B_{\alpha}^{(p)} \\
\text{delta}\ {}^{(l-1)}W'^{(p)}_{\alpha} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{\alpha}^{(p-c)}\left[{}^{(l)}Z^{(p-c-1)}\right]^{T}, &\quad {}^{(l-1)}W'^{(p+1)}_{\alpha} &= {}^{(l-1)}W'^{(p)}_{\alpha} - \eta_{\alpha}^{(p)}\text{delta}\ {}^{(l-1)}W'^{(p)}_{\alpha} \\
\text{delta}\ {}^{(l)}peepW_{\alpha}^{(p)} &= \sum_{c}^{C_{max}} {}^{(l)}\Delta_{\alpha}^{(p-c)}\left[{}^{(l)}S^{(p-c-\gamma)}\right]^{T}, &\quad {}^{(l)}peepW_{\alpha}^{(p+1)} &= {}^{(l)}peepW_{\alpha}^{(p)} - \eta_{\alpha}^{(p)}\text{delta}\ {}^{(l)}peepW_{\alpha}^{(p)}
\end{aligned}
\tag{13}
$$

$$
\text{where}\ \ \gamma = \begin{cases} 0 & \alpha = out \\ 1 & \alpha \in \{in, fgt\} \end{cases}
$$

という手続きにしたがう．ただし，各ゲートへの覗き穴結合に関する重み係数の更新量delta ${}^{(l)}peepW_{\alpha}^{(p)}$　$\alpha \in \{in, fgt, out\}$は，それぞれ対角行列となっている．なお，活性化関数として導入している双曲線正接関数$g(x)[= h(x)] = \tanh(x)$，およびシグモイド関数$f_{\alpha}(x) = 1/(1 + e^{-\nu x})$の導関数は，それぞれ$g'(x) = 1 - g^2(x)$，および$f'_{\alpha}(x) = \nu f_{\alpha}(x)\{1 - f_{\alpha}(x)\}$である．

各々のデルタ量に寄与する学習のための評価指標としては，出力層の出力$Y^{(p)}$と教師データ$T^{(p)}$との差$D^{(p)}$を，$D^{(p)} = \left[\boldsymbol{d}_0^{(p)} \cdots \boldsymbol{d}_q^{(p)} \cdots \boldsymbol{d}_{Q-1}^{(p)}\right]$と表記した場合，次に示す二乗誤差関数$E$が採用される(RNN，LSTM 共通)．

$$
E(epo) = \frac{1}{2}\sum_{p=0}^{\frac{N}{Q}-1}\left[\sum_{q=0}^{Q-1}\left\{\boldsymbol{d}_q^{(p)}\right\}^{T}\boldsymbol{d}_q^{(p)}\right] \quad \Rightarrow \quad \min_{epo} E(epo)
\tag{14}
$$

（epo：学習回数）

4 解析

4.1 データの準備

学習では，専用に用意した任意の時間幅のデータに対し，各時刻より過去ウィンドウ長分のデータを抽出したものを入力層に適用することにより，次時刻の状態（推定値）が出力層より順次得られるものとする．一方，推論では，入力層へのデータの適用は学習時に準拠するものの，得られた（次時刻の）推定値を，当該時刻のデータとしてその都度，入力層へのデータに取り入れるものとする．

換言すれば，「入力層への入力 ⟹ 推定値(出力層からの出力)」の対応関係は，学習では，

$$ {}^{k}\boldsymbol{u} = \begin{bmatrix} u_{k-N_\tau+1} \\ \vdots \\ u_k \end{bmatrix} \in \mathbb{R}^{N_\tau} \quad \Longrightarrow \quad z_{k+1} $$

のように，N_τ次元ベクトル ${}^{k}\boldsymbol{u}$の入力に対し，スカラー量とした推定値z_{k+1}が得られることになる．kおよびN_τは，それぞれ時刻に対応するサンプル番号(インデックス)，および前述のウィンドウ長(時間)に対するデータ点数である．

推論では，

$$ {}^{k}\boldsymbol{u} = \begin{bmatrix} u_{k-N_\tau+1} \\ \vdots \\ u_k \end{bmatrix} \in \mathbb{R}^{N_\tau} \quad \Longrightarrow \quad z_{k+1} $$

$$ {}^{k+1}\boldsymbol{u} = \begin{bmatrix} u_{k-N_\tau+2} \\ \vdots \\ u_k \\ u_{k+1} \end{bmatrix} \in \mathbb{R}^{N_\tau} \quad \Longrightarrow \quad z_{k+2} $$

$$ \vdots $$

$$ {}^{k+N_\tau}\boldsymbol{u} = \begin{bmatrix} u_{k+1} \\ u_{k+2} \\ \vdots \\ u_{k+N_\tau} \end{bmatrix} \in \mathbb{R}^{N_\tau} \quad \Longrightarrow \quad z_{k+N_\tau+1} $$

のように，推定値の確定が順次進行するにしたがい，サンプル番号$k+N_\tau$に相当する推定値を取り入れた際，時間的にそれ以降の入力層への入力データ${}^{m}\boldsymbol{u}$　$(m \geq k+N_\tau)$は，常にN_τ個の直近の推定値により構成されることになる．

以上を踏まえ，実際の推定では，バッチ処理を考慮した場合，p番目に位置するデータセグメント行列$X^{(p)}$および対応する推定値$\hat{Y}^{(p)}$は，それぞれ次のように表現される．

$$ X^{(p)} = \begin{bmatrix} x_{pQ}^{(p)} & \cdots & x_{pQ+(Q-1)}^{(p)} \\ \vdots & \ddots & \vdots \\ x_{pQ+(N_\tau-1)}^{(p)} & \cdots & x_{pQ+(N_\tau-1)+(Q-1)}^{(p)} \end{bmatrix}, \qquad \hat{Y}^{(p)} = \begin{bmatrix} \hat{y}_{pQ+N_\tau}^{(p)} & \cdots & \hat{y}_{pQ+N_\tau+(Q-1)}^{(p)} \end{bmatrix} \tag{15} $$

元のデータ行列$X_o \in \mathbb{R}^{N_\tau \times (N-N_\tau+1)}$に対し，データセグメント数は$(N-N_\tau+1)/Q$；$p \in \{0, \cdots, (N-N_\tau+1)/Q-1\}$.

4.2 各種パラメータ設定

ネットワークの運用に際し，正弦波信号および脳波信号の推定に関する諸条件をそれぞれ表 1 および表 2 に示す．正弦波信号については，訓練データ$y_i(t) = A_i \sin(\omega t + \theta_i)$ $i \in \{1,2,3\}$に対しテストデータ$y_j(t) = A_j \sin(\omega t + \theta_j')$ $j \in \{1,\cdots,5\}$とした場合，位相の異なるデータに対する学習器の適応性を検証するため，表 1 の条件にしたがい，推論用テストデータとして，(訓練データと)位相が同じデータ 1 種(θ_i)，位相の異なるデータ 4 種($\theta_j' \neq \theta_i$)を準備した [i.e. $\theta_2 = \pi/2$, の場合，$\theta_j' = \{0,\ \pi/4,\ 3\pi/4,\ \pi\}$　$j \in \{1,2,4,5\}$]．訓練データとテストデータの位相差は$\theta = \theta_j' - \theta_i$とする．また，“位相の異なるデータを複数同時に推定する”ことに関する学習器の能否を確認するため，訓練データの組み合わせとして [$(\theta_1, \theta_2) = (0,\ \pi/2)$, $(\theta_1, \theta_3) = (0,\ \pi)$, $(\theta_1,\ \theta_2,\ \theta_3) = (0,\ \pi/2,\ \pi)$]の 3 通りを考えた．

以降，学習として，1 種類の訓練データによる推定を「単体推定」と，また位相の異なる訓練データを複数活用した推定を「複数同時推定」と，それぞれ呼称する．

信号に付け加える雑音については，正規分布にしたがうものとし，分散値としてその大きさは，訓練データ用では信

表 1. 正弦波推定に関わる諸条件

	単体推定	複数同時推定
訓練データ	10Hz 正弦波 + 雑音(標準正規分布) (位相$\theta_i = \{0,\ \pi/2,\ \pi\}\ i \in \{1,2,3\}$)	10Hz 正弦波+ 雑音(標準正規分布) 組み合わせ:$(0,\ \pi/2)$, $(0,\ \pi)$, $(0,\ \pi/2,\ \pi)$
教師データ	10Hz 正弦波 (位相$\theta_i = \{0,\ \pi/2,\ \pi\}\ i \in \{1,2,3\}$)	10Hz 正弦波 3 通り 組み合わせ:$(0,\ \pi/2)$, $(0,\ \pi)$, $(0,\ \pi/2,\ \pi)$
テストデータ	10Hz正弦波 5 種(位相$\theta'_j = \{0, \pi/4,\ \pi/2,\ 3\pi/4,\ \pi\}\ j \in \{1, \cdots, 5\}$) + 上記雑音の 2 倍	
サンプルレートf_{sp} (Hz)	200	
全データ長(時間&点数)	$0.1\text{s} \times 5\text{cycles}\ (N = 100)$	
層数(隠れ層+出力層)	2～Max10	
(入力層ユニット数, バッチ数)	(21, 20) 1 周期相当	
誤差関数	(14)式参照	$E_{all}(epo) = \sum_{com=0}^{N_{cmb}-1} E_{com}(epo)$

号の実効値レベルの10%程度に, また, テストデータ用では同レベルの 20%程度(訓練データ用の 2 倍相当)に, それぞれ設定した. 入力層のユニット数N_τは信号 1 周期分程度のデータ点数とし, 1 回のバッチ処理に利用するデータ数(バッチ数)Qも同程度とした[${}^{(0)}U^{(p)} = X^{(p)} \in \mathbb{R}^{N_\tau \times Q}$:(1)式参照].

誤差関数に関し, 単体推定の場合は, 原則, (14)式にしたがうものとするが, 第 1 サメーションの上限は, $(N - N_\tau + 1)/Q - 1$である. 複数同時推定の場合は, データの組み合わせの数をN_{cmb}とすると, 訓練データ毎に算出された誤差E_{com}の総和E_{all}を評価値としている. $N - N_\tau + 1$は入力層が実際に受容する訓練データ数である.

一方, 脳波信号においては, 表 2 のように, α波(8-13Hz)が強く出現する左右の後頭部に配置した電極(左 O1 部, 右 O2 部)から獲得される各々の信号に対し, 低周波および高周波成分(アーチファクト)をフィルタ処理(遮断周波数: $f_1 = 3\text{Hz}, f_2 = 30\text{Hz}$)により除去したものを訓練データとし, また, 教師データには, α波帯域の成分のみを抽出(フィルタ処理: $f_1 = 5\text{Hz}, f_2 = 17\text{Hz}$)したものを採用した. 誤差関数についても, 電極ごとに得られるデータにより算出された誤差E_{ch}の総和E_{all}とした.

表 2. 脳波信号推定に関わる諸条件

訓練データ	後頭部安静閉眼時データ(複数部位: $CH = 2$; フィルタ処理)
教師データ	α波信号(フィルタ処理)
テストデータ	後頭部安静閉眼時データ(複数部位: $CH = 2$;フィルタ処理)
サンプルレートf_{sp} (Hz)	200
全データ長(時間&点数)	1sec or 2sec ($N = 200$ or 400)
層数(隠れ層+出力層)	2～3
(入力層ユニット数, バッチ数)	(21, 20) or (41,40)
誤差関数	$E_{all}(epo) = \sum_{ch=0}^{CH-1} E_{ch}(epo)$

(5), (12)および(13)式のネットワークパラメータの更新量に関わるC_{max}については, 過去いくつかのバッチ処理による算出量を活用する際, 次のような場合分けを採用している[$\alpha \in \{in, fgt, out\}$].

$$C_{max} = \begin{cases} p \\ p-1 \end{cases} \quad \left(0 \le p \le \frac{N_\tau - 1}{Q}\right) \quad \begin{cases} \text{delta}\ {}^{(l-1)}W^{(p)}, \text{delta}\ {}^{(l-1)}B^{(p)}, \text{delta}\ {}^{(l-1)}W_\alpha^{(p)}, \text{delta}\ {}^{(l-1)}B_\alpha^{(p)}, \text{delta}\ {}^{(l)}peepW_{out}^{(p)} \\ \text{delta}\ {}^{(l-1)}W'^{(p)}, \text{delta}\ {}^{(l-1)}W'^{(p)}_\alpha, \text{delta}\ {}^{(l)}peepW_{in\ \text{or}\ fgt}^{(p)} \end{cases}$$

$$C_{max} = \frac{N_\tau - 1}{Q} \quad \left(\frac{N_\tau - 1}{Q} < p < \frac{N - N_\tau + 1}{Q}\right) \qquad \text{common}$$

4.3 結果および検討

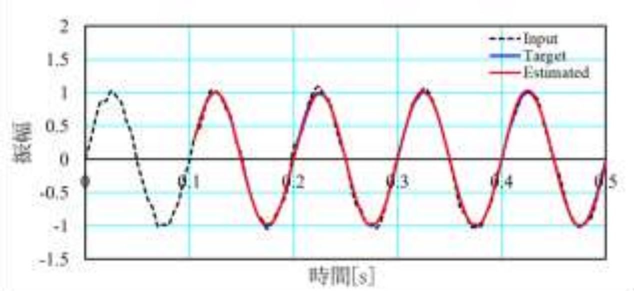
方法論の検証については, 上述の表 1 および表 2 に基づき, 種々実施したが, 本節では, 代表的な条件によるものを事例として示すことにする.

4.3.1 正弦波信号

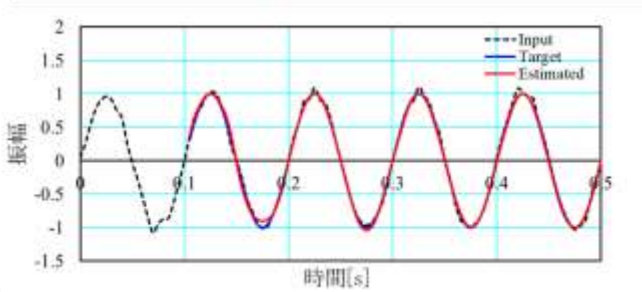
表 1 の諸条件にしたがい, 単体推定を行った際の事例として, 訓練データ($\theta_1 = 0$)による学習時およびテストデータ($\theta'_3 = \pi/2$)による推論時の結果を, RNN, LSTM の 2 つの学習法を比較する形で, それぞれ図 2 および図 3 に示す.

図には，推定波形(Estimated)と教師データ(Target)の類似度を意味する指標として算出した 2 種類の評価値[相関係数Co/二乗誤差Er]を併記している．

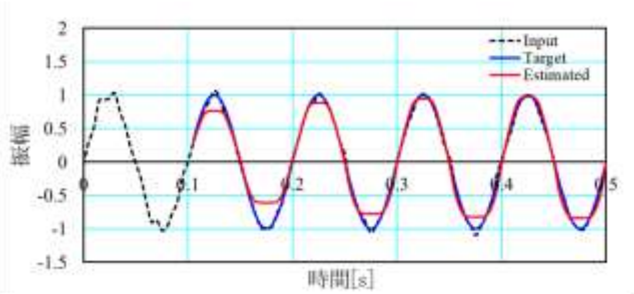
学習時においては，図 2 のように，2 層ネットワーク(隠れ層数 1)では，RNN, LSTM 双方の方法とも推定結果に遜色が見られないが，8 層(隠れ層数 7)のような深層ネットワークでは LSTM の方が教師データとのマッチングがうまくとれており，優れている様子がわかる．一方，推論時においては，2 層ネットワークの場合，学習時と同じ位相のテストデータについては，RNN, LSTM 共に，学習時と同水準の推定性能を発揮していたことを確認しているが，図 3 に示すように，位相の異なるデータに対しては，方法論の違いにより優劣がみられるものの，推定は学習時に比べ劣化する



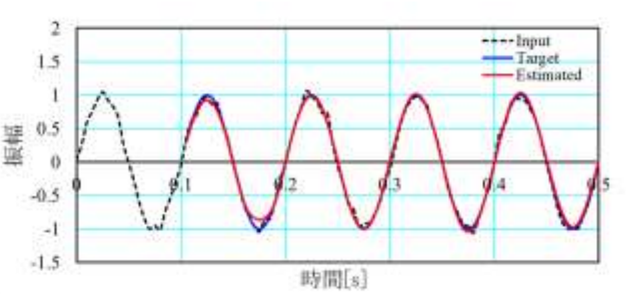
(a) 2 層［0.999/0.0286］



(b) 2 層［0.999/0.0773］

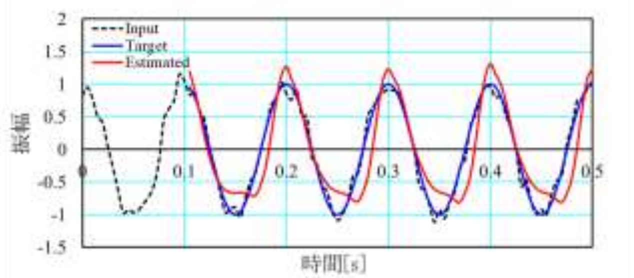


(c) 8 層［0.988/1.09］

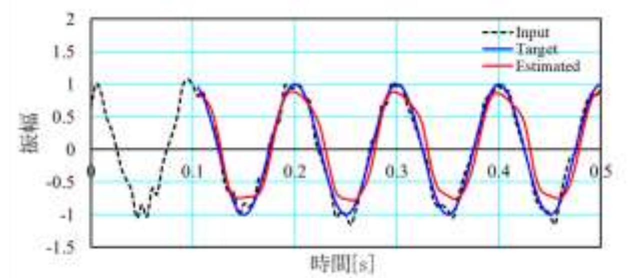


(d) 8 層［0.998/0.116］

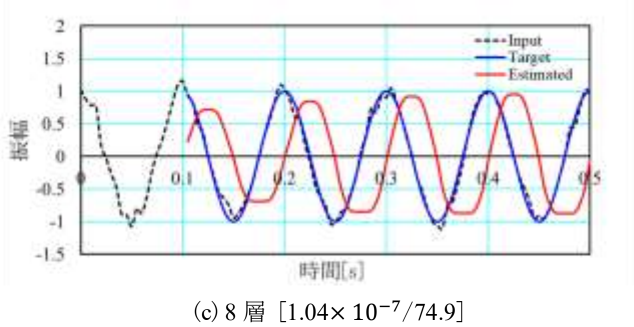
図 2．正弦波推定(学習時$\theta_1 = 0$; 左:RNN,右:LSTM)



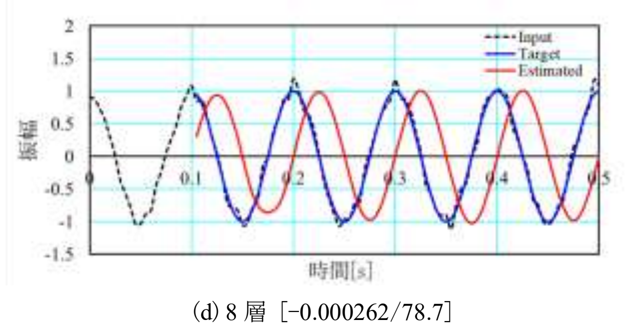
(a) 2 層［0.901/7.91］



(b) 2 層［0.970/2.74］



(c) 8 層［1.04×10^{-7}/74.9］



(d) 8 層［−0.000262/78.7］

図 3．正弦波推定(推論時:$\theta_3' = \pi/2$; 左:RNN,右:LSTM)

ことがわかる．さらに，8 層の場合では，学習法にかかわらず，推定波形が教師(or 入力)データと大きく食い違う様子が伺えるが，このときの相関係数がゼロに近い値であることから，互いの信号が直交関係(双方の信号の位相差$\theta = \theta_3' - \theta_1 = \pi/2$)にあるという裏付けにもなっている．

いま，LSTM に着目し，図 4 に，5 種のテストデータに対する推論時のネットワークの階層別評価値(相関係数)の結果を示す．図より，比較的良好な推定は，① すべてのテストデータに対しては，唯一 2 層の場合であること，② 3 層以上のネットワークでは，訓練データと同じ位相のデータであること，がわかる．さらに，5 層以上では，推定波形が図 3(d)の 8 層のように教師データと大きく異なる様子を確認しているが，相関係数が図のような値を示した根拠としては，推定波形の平均値が教師データのものと同様，ゼロに近いため，双方の信号の標準偏差が共に「ノルム」と，また，共分散が「内積」と等価・同義になり，延いては相関係数が互いの波形（ベクトル）のなす角，換言すると「マッチング角」の余弦値に収束する傾向にあるところに帰着する．例を挙げると，訓練データとの位相差$\theta = \pi/4$のテストデータの場合は，相関係数$Co = 0.706$であるため，$Co \simeq \cos(\pi/4) = 1/\sqrt{2}$，また，位相差$\theta = \pi$のデータ場合は$Co = -0.999 \simeq \cos(\pi) = -1$．このようなことから，ネットワークの構成が階層を深めるほど，訓練データに準拠しないデータへの対応が厳しくなり(過学習の強化)，延いては，ネットワークが入力の状況にかかわらず，訓練データに準拠したデータ(同相)を出力する傾向にあることを示唆している．

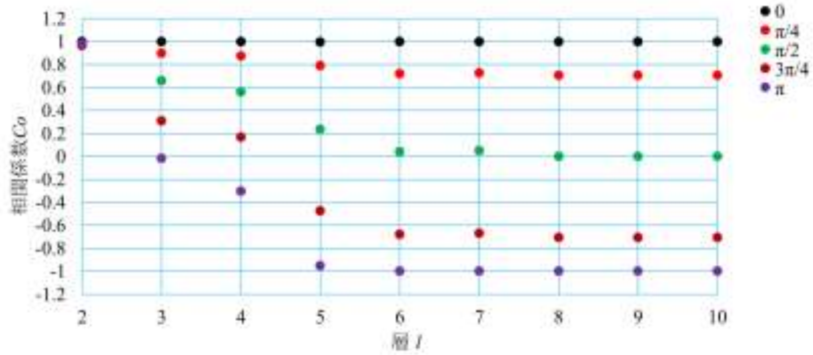
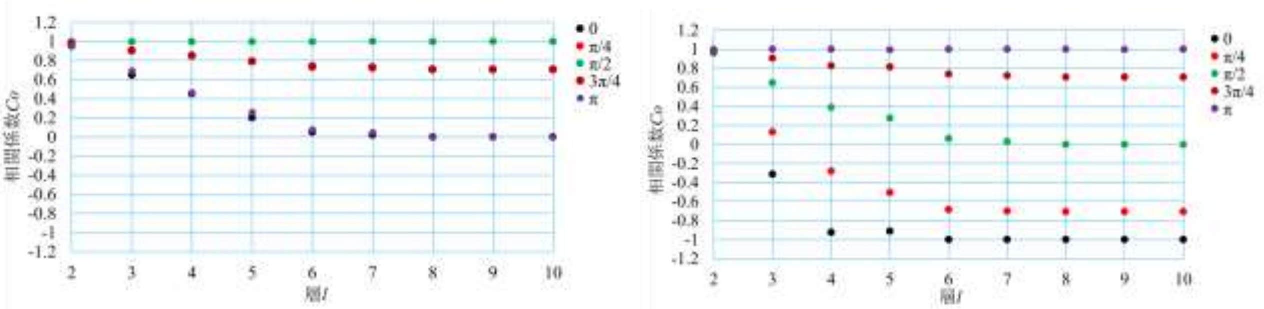


図 4. ネットワークの階層別評価結果(相関係数)

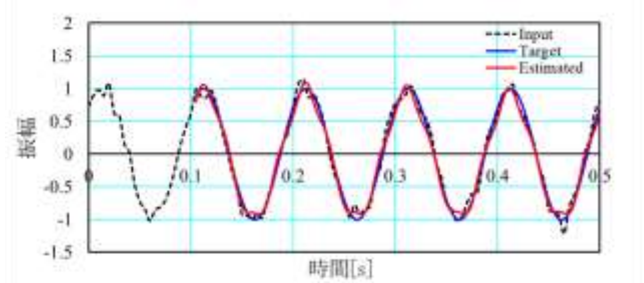


(a) $\theta_2 = \pi/2$　　(b) $\theta_3 = \pi$

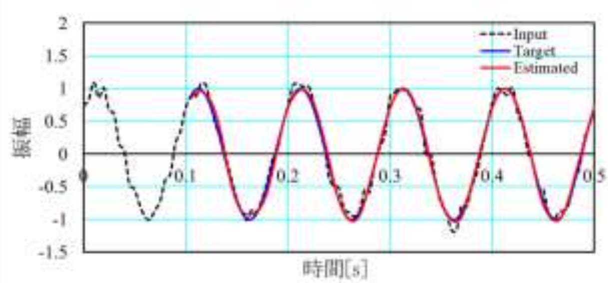
図 5. 正弦波推定($\theta_i \ i \in \{2,3\}$)システムに対する推論時の評価結果(相関係数)

図 5 は訓練データの位相θ_iを$\theta_2 = \pi/2$および$\theta_3 = \pi$とした場合の学習器に対する推論結果(LSTM)であるが，図 4 の正弦波信号(θ_1 =0)のときと同様，良好な推定は 2 層ネットワークであるか，データが訓練データと同じ位相であるかに限られるが，6 層以上では，実際，相関係数が訓練データとテストデータの(訓練データからみた)相対的な位相差に対応した値として収束する様子も理解できる．例として，同図(a)においては，訓練データ$\theta_2 = \pi/2$，テストデータ$\theta_4' = 3\pi/4$の場合は双方の信号の相対的な位相差$\theta = \theta_4' - \theta_2 = \pi/4$ であるため，$Co \simeq \cos(\pi/4) = 1/\sqrt{2}$（図●印）．同じく図(b)においては，訓練データ$\theta_2 = \pi$，テストデータ$\theta_1' = \pi/4$の場合は$\theta = \theta_1' - \theta_2 = \pi/4 - \pi = -3\pi/4$．よって，相関係数は$Co \simeq \cos(-3\pi/4) = -1/\sqrt{2}$（図●印）．

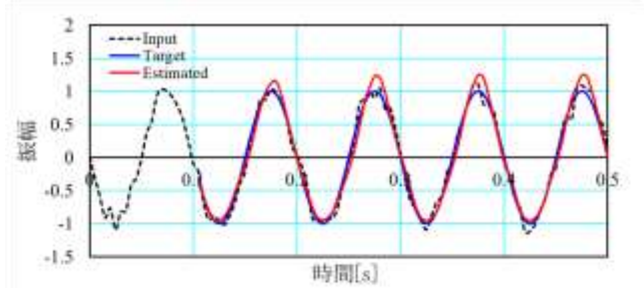
次に，複数同時推定については，訓練データとして位相の異なる波形データの組み合わせを代表し，$(\theta_1, \theta_2) = (0, \pi/2)$および$(\theta_1, \theta_2, \theta_3) = (0, \pi/2, \pi)$の 2 通りについて述べる．$(\theta_1, \theta_2) = (0, \pi/2)$に対する推論時の結果例を図 6 に示す．なお，学習時，および推論時の訓練データと同じ位相のテストデータについては，単体推定と同様，



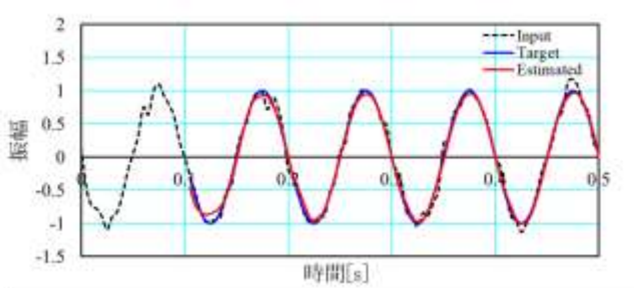
(a) RNN [0.994/0.518]



(b) LSTM [0.998/0.169]

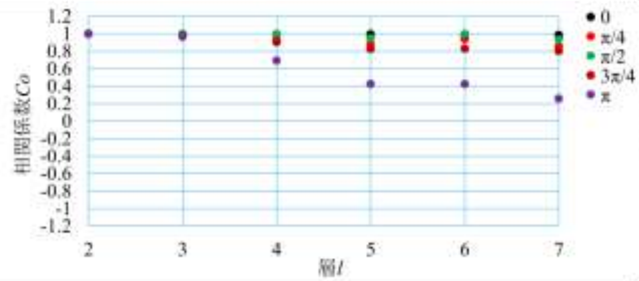


(c) RNN [0.990/0.944]

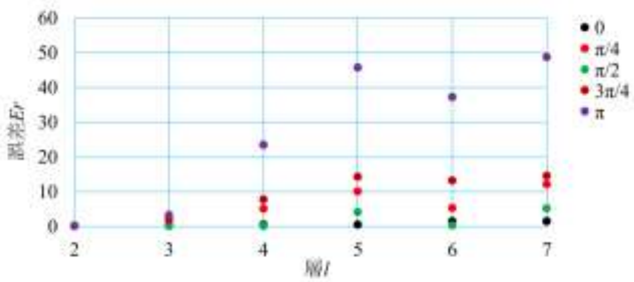


(d) LSTM [0.998/0.198]

図 6. 2 種(0, $\pi/2$)同時推定(上: $\theta_2' = \pi/4$, 下: $\theta_5' = \pi$)



(a) 相関係数Co



(b) 二乗誤差Er

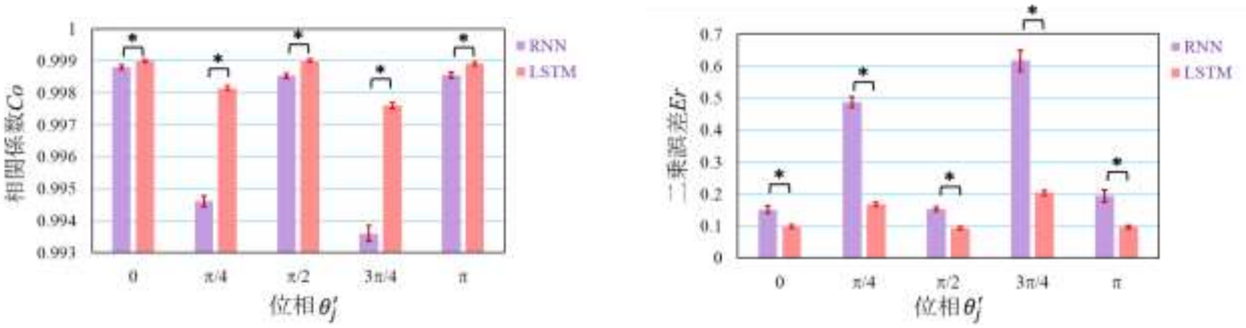
図 7. 2 種(0, $\pi/2$)同時推定に対する評価(LSTM)

良質な推定が充分に行えることを確認している. 図より, 本推定では, 訓練データが有する位相情報と異なるテストデータ($\pi/4$およびπ)において, 情報の補完作用の効果も想定されるが, 単体推定と比較して, RNN では推定性能の改善が, また, LSTM では更なる性能向上が, それぞれなされた様子が見て取れる. 加えて, LSTM では, 図 7 より, 良質な推定は 2〜3 層構造において保証されるものと推察される.

他方, 組み合わせ$(\theta_1, \theta_2, \theta_3) = (0, \pi/2, \pi)$の場合は, 波形の推定性能, 相関レベルおよび LSTM の優位性などについては, $(\theta_1, \theta_2) = (0, \pi/2)$の場合と同水準の傾向にあったが, 二乗誤差$Er$の値は 図 7(b)と異なり, 各層のネットワーク上において$Er < 10$を満足するものであった. いま, 本組み合わせによる推定に対し, RNN と LSTM の評価値(相関係数および二乗誤差)に対する統計学的な有意の有無を調査するため, 2 層ネットワーク構成において, 表 3 に示す条件の下, t 検定を行った. この際, 正弦波信号に加える雑音を, 実装上, default_random_engine クラスとした雑音疑似乱数生成エンジンの起動を通じ 100 回生成することにより, 方法論ごとに学習器を生成する回数N_rを 100 とし

表 3. t 検定条件

学習器生成数N_r	RNN, LSTM 共に 100
有意水準	0.05
検定の種類	両側
帰無仮説	各学習法(RNN, LSTM)に対しそれらの評価値(相関係数 or 二乗誤差)に差がない
統計量, 自由度の算出	Welch 法



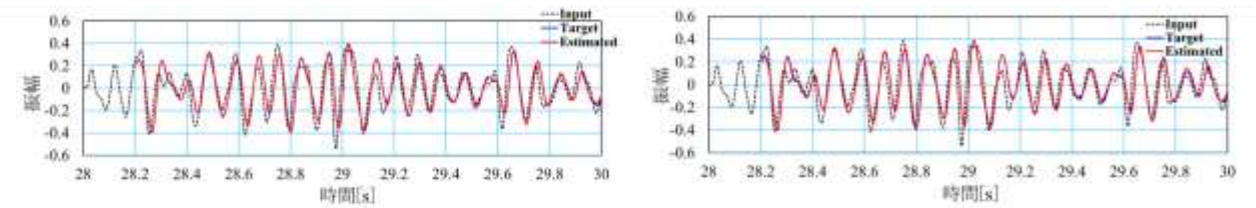
(a) 相関係数Co　　　　(b) 二乗誤差Er

図 8. 3 種$(0,\ \pi/2\ ,\pi)$同時推定に対する t 検定

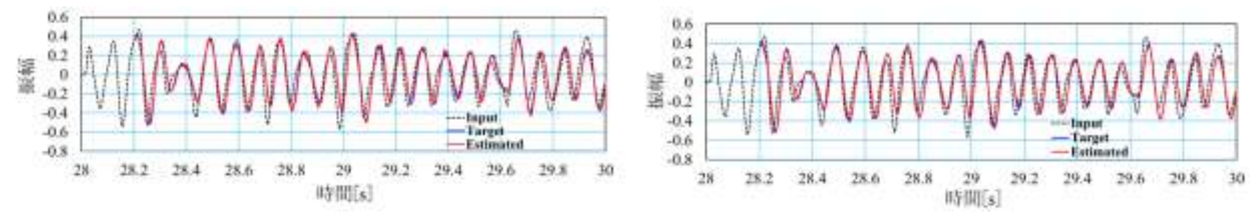
た. 以上に基づいた検定結果を, 図 8 に示す.

図より, 訓練データと異なる位相のテストデータ$\theta_2' = \pi/4$および$\theta_4' = 3\pi/4$については, 多少, 相関係数が小さく, 誤差も大きいものの, いずれの評価値も, すべてのテストデータにおいて, 帰無仮説が棄却され, LSTM が RNN との間に有意差(図のアスタリスク「*」表記)があることから, 優れた性能を有していることがわかる. ただし, 図のエラーバーは標準誤差$SE = \sqrt{v_a/N_r}$に基づくものであり, v_aは不偏分散($a \in$ {RNN,LSTM})である. なお, 本検定の詳細については,「附録 1」を参照されたし.

4.3.2 脳波信号



(a) RNN(O1) [0.997/0.0475]　　　　(b) LSTM(O1) [0.997/0.0572]

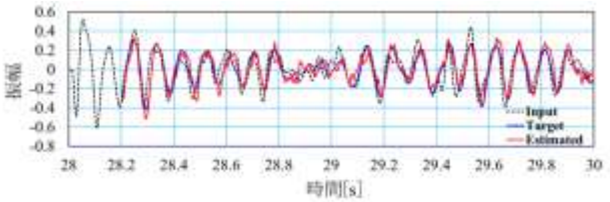


(c) RNN(O2) [0.995/0.156]　　　　(d) LSTM(O2) [0.998/0.0858]

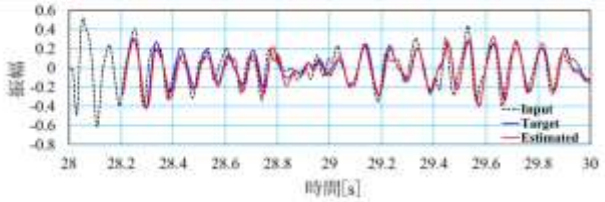
図 9. 脳波信号推定(学習時)

2 層ネットワーク構成において, 表 2 の条件$[(N_\tau,\ Q) = (41{,}40)]$に基づき, 2 つの後頭部に配置した電極(O1, O2)から得られた脳波信号に対する学習時の推定結果(複数同時推定：2sec 分)を図 9 に示すが, RNN, LSTM の双方共, 同水準で良質な結果を得た. 一方, 図 10 および図 11 に, 生成した各学習器に対する推論結果(テストデータ 2 種：同じく 2sec 分)を示すが, RNN の場合は, 学習時と異なり, 鋸歯状をなしている箇所が推定波形に多く散見されるのに対し, LSTM の方は形状が平滑であり, かつ良好な推定が行えていることから, RNN より優れている様子が伺える. 参考のため, 3 層のネットワーク構成においては, 双方の方法共に, 学習時は, 2 層ネットワークのときと同水準の結果であり, 推論時も 2 層ネットワークに比べ, 評価値として全体的に数%の信頼性を失う程度で LSTM の優位は不変であった.

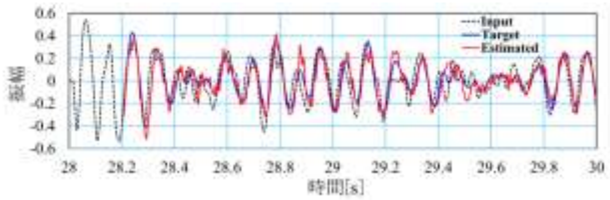
以上を踏まえ, 大きな振幅変化や位相跳躍などが生じやすく, 安定性が常時保証されるとは限らない脳波信号においても, LSTM により, 比較的良好・良質な推定が行えることが判明した.



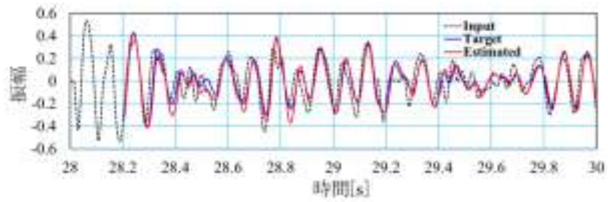
(a) RNN(O1) [0.927/1.33]



(b) LSTM(O1) [0.978/0.517]

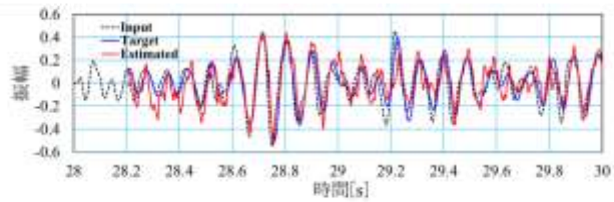


(c) RNN(O2) [0.934/1.22]

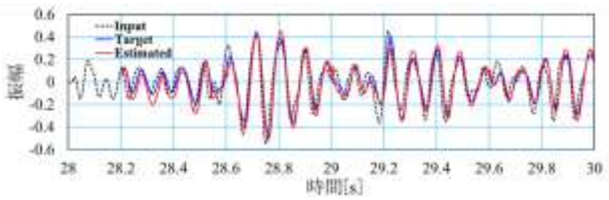


(d) LSTM(O2) [0.976/0.554]

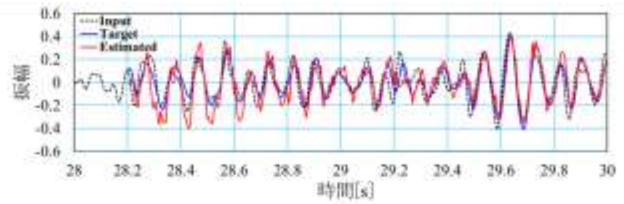
図 10. 脳波信号推定(推論時　テストデータその 1)



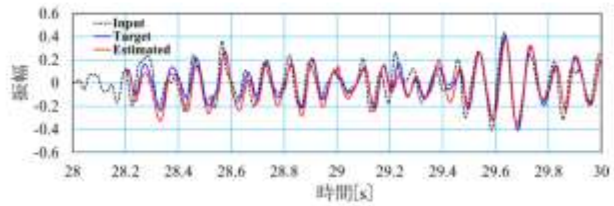
(a) RNN(O1) [0.845/3.20]



(b) LSTM(O1) [0.979/0.697]



(c) RNN(O2) [0.884/2.30]



(d) LSTM(O2) [0.961/0.780]

図 11. 脳波信号推定(推論時　テストデータその 2)

5　おわりに

本信号推定システムにおける学習の打ち切りについては，単体推定の場合，(14)式に示す誤差関数Eが学習ごとの推移として単調減少となるため，適度な条件(直前の値と比較して変化がない)を満足するとき，実施される．一方，複数同時推定の場合は，位相の異なる各訓練データをバッチ処理ごとに交互に入れ替え学習を進めるため，誤差関数の総和E_{all}は単体推定の場合の単調減少と異なり，上下に変動するケースが多い．それゆえ，打ち切る条件には，単体推定に準拠した条件$[E_{all}(epo-1)-E_{all}(epo)]^2 < conv(=1.0\times10^{-4})$の他に，誤差が然るべき値まで小さくなることを見越した$E_{all}(epo) < \varepsilon\ \ (=1.0\times10^{-3})$ を追加している．また，議論の根幹にはしていないが，ネットワークパラメータである重み係数についても，可視化を通じて簡略に触れているので，附録 B を参考にされたい．

最後に，本システムは，隠れ層のユニット数を分割利用することにより，ネットワーク構成として LSTM，RNN 双方の方法論を組み合わせたハイブリッド(hybrid)型システムとしても可動するよう，設計している．実際，ユニット数を割合として互いに 50%として分け合った状態で正弦波や脳波信号の推定を行ってみたが，精度および優位についての大小，あるいは優劣の関係性は RNN< hybrid <LSTM という予め想定しやすい傾向を示すに留まった．今後は，ハイブリッ

ド型が最適システムとなる対象信号を見出だし，検討・議論を深めていきたい．

参考文献

1) 瀧雅人，これならわかる深層学習　入門，講談社，pp.177-196 (2017)
2) S. Hochreiter，J. Schmidhuber：Long Short-Term Memory，Neural Computation，9，pp.1735-1780 (1997)
3) F. A. Gers，J. Schmidhuber and F. Cummins：Learning to Forget：Continual Prediction with LSTM，Neural Computation，12，pp.2451-2471 (2000)
4) 松原篤：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，山口大学総合技術部技術報告集 No.2 (2022) https://www.yamaguchi-u.ac.jp/tech/wp-content/uploads/2023/04/22-2-01.pdf
5) F. A. Gers，N. N. Schraudolph and J. Schmidhuber：Learning Precise Timing with LSTM Recurrent Networks，Journal of Machine Learning Research，3，pp.115-143 (2002)
6) C.M. ビショップ，(監訳) 元田，栗田他，パターン認識と機械学習　上，丸善，pp. 100– 102(2012)
7) 竹内淳：高校数学でもわかる統計学(BLUE BACKS B-1757)，講談社，pp. 149-202(2012)
8) 統計検定のまとめブログ：ウェルチの t 検定，https://data-science.gr.jp/theory/tst_welch_t_test.html

附録 A　t 検定 [6)~8)]

t 検定は，本報告の趣旨に沿えば，2 つの方法論それぞれにより，最終的に得られた評価値(相関係数Co or 二乗誤差Er)の平均値に差が生じた場合，その差に統計学的な意味があるか否かを調査する手段となる．いま，各方法論から得られた評価値の集合をベクトル表記として$\boldsymbol{a} = [a_0 \quad \cdots \quad a_{N-1}]^T \in \mathbb{R}^N$，$\boldsymbol{b} = [b_0 \quad \cdots \quad b_{N-1}]^T \in \mathbb{R}^N$とするとき，$N$個のデータとしてのそれぞれの平均値$\bar{a}$，$\bar{b}$および不偏分散$v_a$，$v_b$は，

$$\bar{a} = \frac{1}{N}\sum_{i=0}^{N-1} a_i, \quad \bar{b} = \frac{1}{N}\sum_{i=0}^{N-1} b_i, \quad v_a = \frac{1}{N-1}\sum_{i=0}^{N-1}(a_i - \bar{a})^2, \quad v_b = \frac{1}{N-1}\sum_{i=0}^{N-1}\left(b_i - \bar{b}\right)^2 \tag{16}$$

いま，集合データ$\boldsymbol{a}, \boldsymbol{b}$を共に標本扱いとし，それらの母集団が不確定のため，双方の標本に対する母分散が等しいと仮定できないとした場合，検定手段として Welch の方法が採用可能となる．このとき，統計量t_wおよび自由度f_wは，以下のような式に基づいて算出される．

$$t_w = \frac{\bar{a} - \bar{b}}{\sqrt{\frac{v_a}{N} + \frac{v_b}{N}}}, \quad f_w = \frac{\left(\frac{v_a}{N} + \frac{v_b}{N}\right)^2}{\frac{\left(\frac{v_a}{N}\right)^2}{N-1} + \frac{\left(\frac{v_b}{N}\right)^2}{N-1}} \tag{17}$$

有意水準αに対する両側検定の場合，有意水準$\alpha/2$とした片側検定と同等となるため，t 分布表を参考に，自由度f_wに対し有意水準$\alpha/2$以下の確率となるときの統計量t_{rp}を活用することにより，上記統計量t_wの評価を行うのが一般的であるが，実際には自由度f_wが実数になる現状を踏まえ(分布表は主に整数の場合で提供されているケースが多い)，本報告では，解析プログラムを作成することにより，次に示す t 分布の面積の視点から統計量t_{rp}を算出することにした．

t 分布にしたがう確率変数の確率密度関数$p(x_k)$は，

$$p(x_k) = \frac{\Gamma\left(\frac{f_w + 1}{2}\right)}{\sqrt{\pi f_w}\,\Gamma\left(\frac{f_w}{2}\right)}\left[1 + \frac{(x_k)^2}{f_w}\right]^{-\frac{f_w+1}{2}} \tag{18}$$

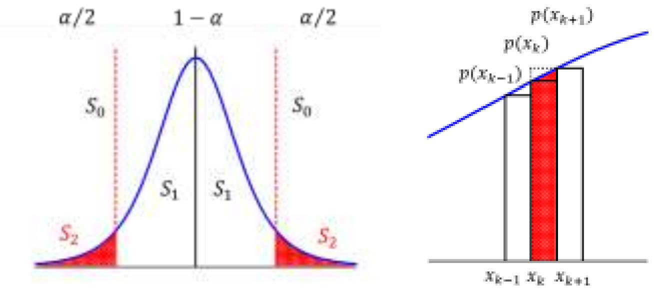
$\Gamma(\cdot)$はガンマ関数である(プログラムへのガンマ関数の実装では，tgamma 関数を利用)．

横軸をx，縦軸をpとすれば，t 分布は自由度f_wに応じて，例を挙げると，その形状は図(a)のようになるが，横軸と分布曲線の全体で囲まれる面積Sについては，同図(b)の区分求積法に基づく離散的表記を導入すると，

$$S=\int_{-\infty}^{\infty}p(x_k)dx=1$$

$$\equiv \lim_{n\to\infty}\left[\frac{1}{2}[p(x_0)+p(x_{n-1})]\Delta x+\sum_{k=1}^{n-2}p(x_k)\Delta x\right]$$

ただし，$\Delta x=x_k-x_{k-1}$であり，nは $n\to\infty$ に極限をとれば$S\simeq 1$となる(有限な)整数値である．



(a) 確率密度関数　　(b) 区分求積

図 A.t 分布曲線

いま，図 A(a)のように，分布の対称性に鑑み，面積として，$S=2S_0$とし，また，両側検定の場合，有意水準$\alpha/2$以上，およびそれ以下の確率となるときの分布領域の面積をそれぞれ$2S_1$，$2S_2$とする場合，$S=2S_0=2S_1+2S_2$，$2S_1=(1-\alpha)S$，$2S_2=\alpha S$により，$S_1=(1-\alpha)S_0$，$S_2=\alpha S_0$と表すことができる．

実装上では，図(b)を例に挙げると，座標軸第 1 象限に位置する分布領域において，区分求積法により各区分の面積を求め順次足し合わせたものが面積S_1と等しくなるポイント$x_k^{\#}$(有意水準αとの境界)を探索する手続きをとる．畢竟するに，当該ポイント$x_k^{\#}$が目的の統計量t_{rp}となる．

位相θ_j'	有意差(あり/なし)	統計量t_w	統計量t_{rp}	平均(RNN)	平均($LSTM$)	不偏分散v_{RNN}	不偏分散v_{LSTM}	自由度f_w
0	あり	-2.05033	1.972	0.99881	0.99899	5.2366×10^{-7}	3.0221×10^{-7}	184.71
π/4	あり	-19.0856	1.976	0.99461	0.99815	2.8713×10^{-6}	5.6672×10^{-7}	136.61
π/2	あり	-5.08436	1.973	0.99854	0.99901	6.1116×10^{-7}	2.5071×10^{-7}	168.52
3π/4	あり	-15.4678	1.978	0.99360	0.99761	5.8320×10^{-6}	8.7881×10^{-7}	128.17
π	あり	-3.24981	1.973	0.99856	0.99892	8.6290×10^{-7}	3.7091×10^{-7}	170.83

(a) 相関係数Co

位相θ_j'	有意差(あり/なし)	統計量t_w	統計量t_{rp}	平均(RNN)	平均($LSTM$)	不偏分散v_{RNN}	不偏分散v_{LSTM}	自由度f_w
0	あり	4.3252	1.975	0.15228	0.099009	0.012173	0.0029967	144.95
π/4	あり	18.100	1.978	0.48780	0.16911	0.027015	0.0039858	127.59
π/2	あり	7.0609	1.973	0.15373	0.094035	0.004958	0.0021889	172.14
3π/4	あり	12.662	1.981	0.61777	0.20472	0.10052	0.0058848	110.55
π	あり	4.8571	1.980	0.19537	0.09742	0.037981	0.0026928	112.96

(b) 二乗誤差Er

表 A. 3 種$(0,\ \pi/2\ ,\pi)$同時推定に対する t 検定

以上により，帰無仮説の棄却条件 $t_w<-t_{rp}$ or $t_{rp}<t_w$ を満足する場合，2 種の方法論(RNN および LSTM)による評価値に有意差が認められることになる．

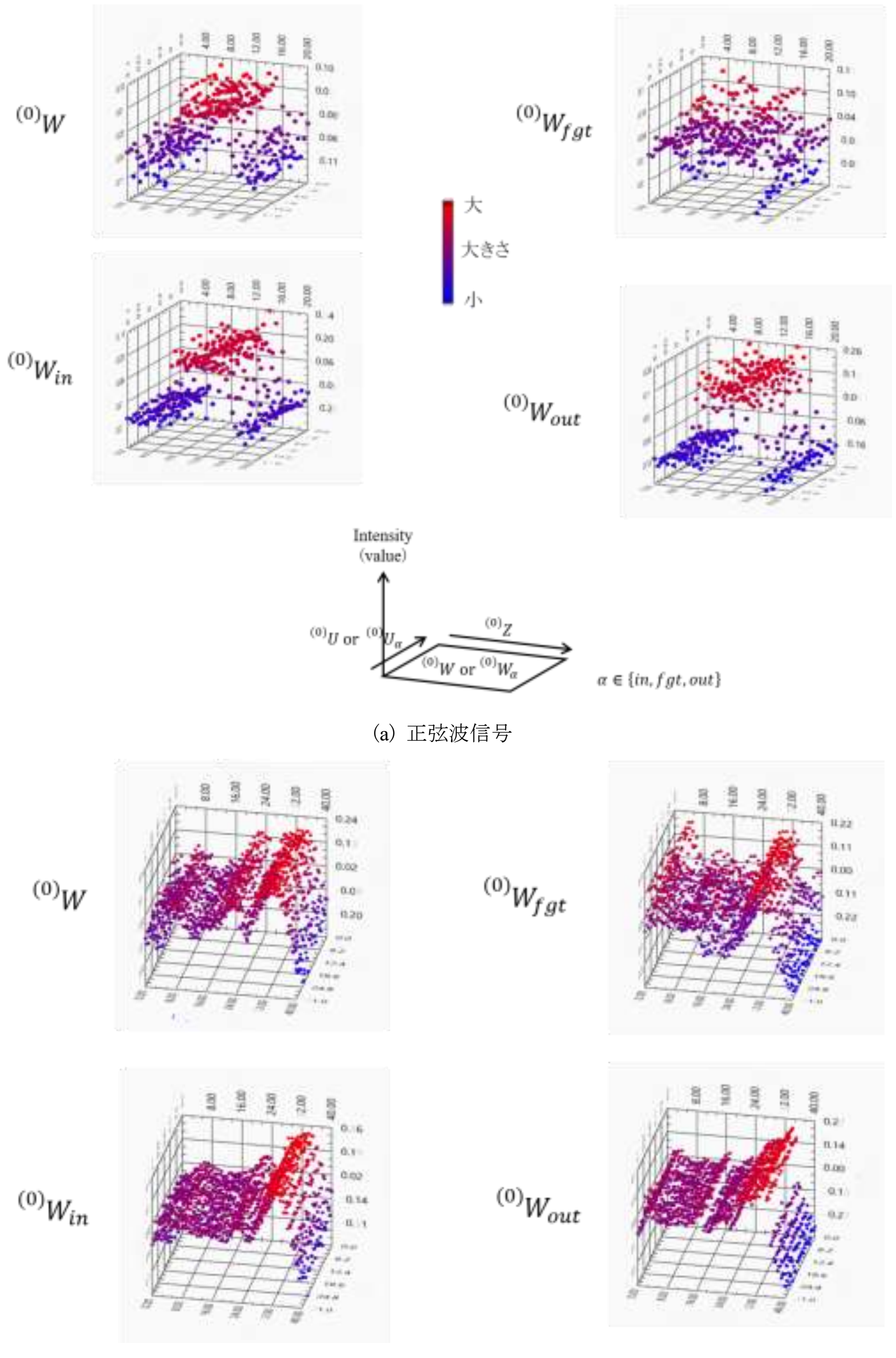
なお，前述の図 8 は，上記統計的処理を経由し算出された表 A に示す内容を基に，描画したものである．

附録 B ネットワークパラメータ～重み係数

訓練データによる学習器の生成は，重み係数などのネットワークパラメータの決定に帰着する．いま，LSTM において，正弦波信号(位相:θ_1 =0)に対し，2 層ネットワーク構成による単体推定を行った際，学習器として確定した各種重み係数の状況を図 B(a)に，また，脳波信号に対し，同レベルのネットワーク構造において，単体推定（測定電極:O2)を行ったときの同種パラメータの状況を同図(b)に，それぞれ示す．なお，入力層ユニット数N_τおよびバッチ数Qの組み合わせ$(N_\tau,\ Q)$は，前者が(21,20)[信号の 1 周期分相当]，後者が(41,40)[α波の 2～3 周期分]である．

正弦波信号の場合は，図(a)より，忘却ゲートの重み係数${}^{(0)}W_{fgt}$については複雑な分布を呈しているものの，メモリセルの前段に位置するユニットに作用する重み係数${}^{(0)}W$や，入出力ゲートに寄与する係数${}^{(0)}W_\alpha$ $\alpha\in\{in, out\}$については，訓練データである正弦波信号に由来した山型の分布を形成していることがわかる．また，脳波信号では，図(b)より，4 種すべての重み係数において，信号 2～3 周期分相当とした信号の振舞い(振幅変化)をうまくとらえた分布形状をなしている様子がわかる．

ネットワーク内では，実際，信号は然るべきタイミングにおいてこれらのネットワークパラメータの作用を受けるだけでなく，活性化関数などの様々な処理も経由しながら，上位層に伝達され，最終的には，推定値として出力層からの発出に至るところではあるが，図に示したような重み係数の可視化に鑑みた場合，信号由来の形状を成していることから，換言すれば，データの転写にも似た効果をもたらした結果から，重み係数は鋳物，シリコン，ゼリー状などの材質の型と同類とも言うべき（訓練データに応じた）“造形の型(モールド：mold)”としての役割を担うものとも推察される．



(a) 正弦波信号

(b) 脳波信号

図 B. 各種重み係数の可視化