

LH-005

雑音を含むデータに対する TDNN のロバスト学習

Robust Learning of TDNN for Noisy Data

西山 清*, ○山田 知明

Kiyoshi NISHIYAMA* and ○Tomoaki YAMADA

1 はじめに

時系列の学習・予測には Time-Delay Neural Network (TDNN) が広く用いられている。しかし、TDNN は出力が入力へフィードバックされるため出力の誤差が次の時刻の入力の誤差となり、さらに出力の誤差が増大すると云った誤差の悪循環が生じる。本論文では、教師データの雑音だけでなく入力雑音も考慮した拡張カルマンフィルタ (extended Kalman filter : EKF) に基づく新たな学習を提案し、これが TDNN のもつ誤差の悪循環の解決に有効であることを示す。

2 TDNN

図1に雑音を考慮した1出力の3層 TDNN (Time-Delay Neural Network) を示す。ここで、第1、第2、第3層内のニューロン数はそれぞれ N_1 、 N_2 、1 であり (N_1 - N_2 -1 ネットワーク)、 $w_{i,j}^n$ は n 層の i 番目のニューロンから ($n+1$) 層の j 番目のニューロンへとつながる結合重みであり、 θ_i^n は ($n+1$) 層の i 番目のニューロンのしきい値である。

また、入力 $u_{k,i}$ には雑音 $v_{k,i}$ が、出力層のニューロンの出力 z_k には雑音 v_k が加わるので、最終的に出力の観測値は次のように表される。

$$y_k = z_k + v_k = f \left(\sum_{j=1}^{N_2} w_{j,1}^2 f \left(\sum_{i=1}^{N_1} w_{i,j}^1 (u_{k,i} + v_{k,i}) + \theta_j^1 \right) + \theta_1^2 \right) + v_k$$

ここで、 $y_k = z_k + v_k$ 、 $u_{k,i} = z_{k-i}$ 、 $v_{k,i} = v_{k-i}$ であり、 $f(\cdot)$ は次のシグモイド関数である。

$$f(x) = \frac{1 - \exp(-\alpha x)}{1 + \exp(-\alpha x)}$$

ただし、 α はシグモイド関数の傾きを表すパラメータである。

*岩手大学 工学部 情報システム工学科, 〒020-8551 盛岡市 上田 4-3-5, e-mail: nishiyama@cis.iwate-u.ac.jp

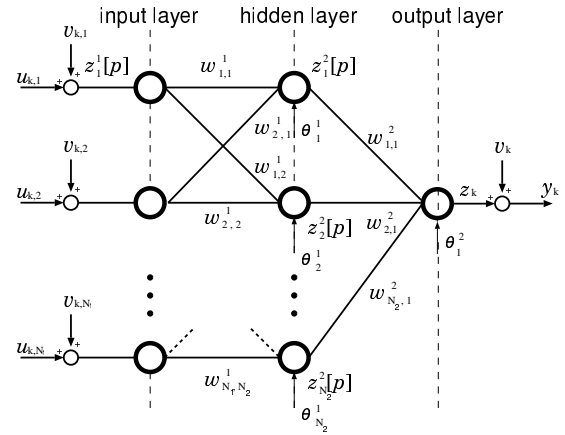


図1: TDNN と雑音

3 ロバスト EKF 学習

3.1 非線形モデル

入力と教師データの雑音 v_k を考慮した TDNN は次の状態空間モデルで表すことができる。

$$\mathbf{w}_{k+1} = \mathbf{w}_k \quad (\text{状態方程式}) \quad (1)$$

$$y_k = h_k(\mathbf{w}_k, \mathbf{v}_k) + v_k \quad (\text{観測方程式}) \quad (2)$$

ただし、

$$\begin{aligned} h_k(\mathbf{w}_k, \mathbf{v}_k) &= f \left(\sum_{j=1}^{N_2} w_{j,1}^2 f \left(\sum_{i=1}^{N_1} w_{i,j}^1 (u_{k,i} + v_{k,i}) + \theta_j^1 \right) + \theta_1^2 \right) \\ \mathbf{v}_k &= [v_{k,1}, v_{k,2}, \dots, v_{k,N_1}]^T, \quad v_{k,i} = v_{k-i} \\ \mathbf{w}_k &= [w_{1,1}^1, w_{2,1}^1, \dots, w_{N_1,1}^1, \theta_1^1, w_{1,2}^1, w_{2,2}^1, \dots, \\ & \quad w_{N_1,2}^1, \theta_2^1, \dots, w_{1,N_2}^1, w_{2,N_2}^1, \dots, w_{N_1,N_2}^1, \\ & \quad \theta_{N_2}^1, w_{1,1}^2, w_{2,1}^2, \dots, w_{N_2,1}^2, \theta_{N_2}^2]^T \in \mathcal{R}^{N_w \times 1} \\ N_w &= (N_1 + 1)N_2 + (N_2 + 1) \end{aligned} \quad (3)$$

3.2 線形化モデル

新たに考案した状態空間モデルでは観測方程式が非線形なので $h_k(\mathbf{w}_k, \mathbf{v}_k)$ を $(\hat{\mathbf{w}}_{k-1}, \mathbf{0})$ の近傍でテ

イラー展開 (Taylor expansion) し、

$$h_k(\mathbf{w}_k, \mathbf{v}_k) = h_k(\hat{\mathbf{w}}_{k-1}, \mathbf{o}) + \mathbf{H}_k(\mathbf{w}_k - \hat{\mathbf{w}}_{k-1}) + \mathbf{D}_k \mathbf{v}_k + \dots \quad (4)$$

その2次以降の項を無視(1次の項までを採用)すれば、次の線形化システムが得られる。

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k \\ m_k &= \mathbf{H}_k \mathbf{w}_k + \mathbf{D}_k \mathbf{v}_k + v_k \\ &= \mathbf{H}_k \mathbf{w}_k + \begin{bmatrix} 1 & \mathbf{D}_k \end{bmatrix} \begin{bmatrix} v_k \\ \mathbf{v}_k \end{bmatrix} \end{aligned} \quad (5)$$

ただし、

$$\begin{aligned} m_k &= y_k - h_k(\hat{\mathbf{w}}_{k-1}, \mathbf{o}) + \mathbf{H}_k \hat{\mathbf{w}}_{k-1} \\ \mathbf{H}_k &= \left. \frac{\partial h_k(\mathbf{w}_k, \mathbf{o})}{\partial \mathbf{w}_k^T} \right|_{\mathbf{w}_k = \hat{\mathbf{w}}_{k-1}} \\ \mathbf{D}_k &= \left. \frac{\partial h_k(\hat{\mathbf{w}}_{k-1}, \mathbf{v}_k)}{\partial \mathbf{v}_k^T} \right|_{\mathbf{v}_k = \mathbf{0}} \\ \mathbf{v}_k &= [v_{k,1}, v_{k,2}, \dots, v_{k,N_1}]^T, \quad v_{k,i} = v_{k-i} \\ v_k &= y_k - h_k(\mathbf{w}_k, \mathbf{v}_k) \end{aligned} \quad (6)$$

ここで、 v_k は平均0、分散 $\sigma_v^2 > 0$ の白色雑音と仮定する。

3.3 学習アルゴリズム

この線形化システムにカルマンフィルタを適用すれば、拡張カルマンフィルタ (extended Kalman filter: EKF) に基づく TDNN のロバスト学習アルゴリズムが次のように得られる。

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \mathbf{K}_k (y_k - h_k(\hat{\mathbf{w}}_{k-1}, \mathbf{o})) \quad (7)$$

$$\mathbf{K}_k = \hat{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T \cdot (\mathbf{H}_k \hat{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T + \mathbf{D}_k \mathbf{D}_k^T + 1)^{-1} \quad (8)$$

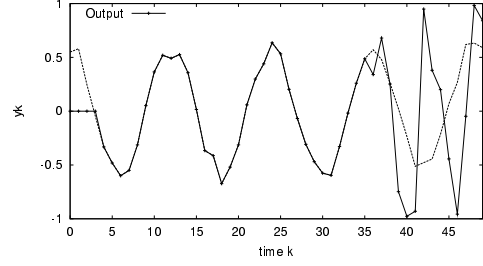
$$\hat{\mathbf{P}}_{k+1|k} = \hat{\mathbf{P}}_{k|k-1} - \mathbf{K}_k \mathbf{H}_k \hat{\mathbf{P}}_{k|k-1} \quad (9)$$

ただし、 $\hat{\mathbf{P}}_{k|k-1} = \hat{\Sigma}_{k|k-1} / \sigma_v^2$ とした。また、 $\frac{E\{m_k m_k^T | y_0, \dots, y_{k-1}\}}{\sigma_v^2} = \mathbf{H}_k \hat{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T + \mathbf{D}_k \mathbf{D}_k^T + 1$ となることに注意されたい。

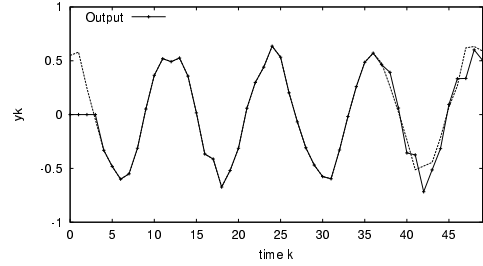
通常の EKF 学習では式 (8) の分母に $\mathbf{D}_k \mathbf{D}_k^T$ が現れない。

4 シミュレーション

ロバスト EKF 学習アルゴリズムの有効性を検証するため、平均 $\bar{v}_k = 0$ 、分散 $\sigma_v^2 = 0.05$ の白色雑音



(a) g -EKF; 学習回数: 65 回



(b) ロバスト g -EKF; 学習回数: 46 回

図 2: ロバスト g -EKF の予測結果;

4-6-1 ネットワーク、 $\bar{v}_k = 0$ 、 $\sigma_v^2 = 0.05$ 、 $\alpha = 1.5$
打ち切り誤差: 1.0×10^{-3} 、学習パターン数: 32
学習区間: $k = 0 \sim 35$ 、非学習区間: $k = 36 \sim 49$ 、

を加えた正弦波を用いて、従来の EKF 学習アルゴリズム ($\mathbf{D}_k \mathbf{D}_k^T = 0$) である g -EKF[2] と比較した。4-6-1 ネットワークのシグモイド関数の傾き $\alpha = 1.5$ 、重みベクトル $\hat{\mathbf{w}}_0$ の初期化の乱数の種: 751、 $\mu = 10$ 、打ち切り誤差 1.0×10^{-3} 、 $\sigma_v^2 = 0.05$ で学習を行い、その後予測を行った結果を図 2 に示す。ただし、学習の際、雑音の分散は未知とした。

5 結論

入力と教師データの雑音を考慮したロバスト EKF 学習を新たに提案し、これが TDNN の誤差の悪循環による予測性能の低下を低減する効果があることを示した。今後は RBF(radial basis function) ネットワークに対してもその有効性を検証したい。

参考文献

- [1] 西山 清: 最適フィルタリング, 培風館, 2001.
- [2] K.Nishiyama and K.Suzuki: "H ∞ -Learning of Layered Neural Networks", IEEE Transactions on Neural Networks, Vol.12, No.6, pp.1265-1277, 2001.