

論文の内容の要旨

論文題目 ニューラルネットワークの学習と
 パラメータのゆらぎに関する研究

氏 名 小野田 崇

1986年にRumelhartらによって提案された階層型ニューラルネットワークの誤差逆伝播学習法(Back-Propagation)は、ニューラルネットワークの潜在能力の高さを一般に知らしめるとともに、様々な分野での応用を促した。現在でも、ニューラルネットワークの研究においては、様々な応用を目的にネットワークアーキテクチャと学習法が提案されているが、ニューラルネットワークのパラメータの分析はある意味で近年までおざなりにされてきた。しかし、現実問題にある種のニューラルネットワークを適用する場合、ネットワークの中で自由に更新できるパラメータの数をどのくらいにすれば、実環境でどのくらいの能力を発揮できるのかということが重要な問題となる。

我々は、このニューラルネットワークのパラメータの数とその能力との関係がニューラルネットワークの学習によって推定されるパラメータの特徴を数理的に分析することで明らかにできると考え、観測例題数が無限に得られ、ニューラルネットワークによる学習が理想的に行なわれた場合のパラメータを最適パラメータとして、実適用下で推定されるパラメータの最適パラメータからのずれを次の2つの「ゆらぎ」に分けて分析した。

- 現実には観測される例題数は有限である。そのため、有限個の偏った例題から最適パラメータに近い値を推定することになる。有限個の例題に対してニューラルネットワークの学習が理想的に実行されて得られるパラメータである準最適パラメータは最適パラメータのまわりにゆらぐことになる。このゆらぎを「統計的ゆらぎ」と呼ぶことにし、このゆらぎを分析する。
- 現実にはニューラルネットワークが準最適パラメータを求められるのは稀である。つまり、ニューラルネットワークの学習によって推定されるパラメータは準最適パラメータの周辺にゆらぐことになる。このゆらぎを「学習にともなうゆらぎ」と呼ぶことにし、このゆらぎを分析する。

我々はニューラルネットワークモデル $M_i = \{p_i(y|x; w^i)\}$ を $M_1 \subset M_2 \subset M_3 \subset \dots$ のような階層モデルであるとして、推定パラメータの有する上述2つのゆらぎの数理的に分析し、ニューラルネットワークが実環境下でどれくらいの能力を発揮できるかを評価する次式で表される規準 (Neural Network Information Criterion:NNIC) を導出した。

$$\text{NNIC}(\tilde{w}_n^i) = D(q_n^*, \tilde{w}_n^i) + \tilde{t}^i (D(q_n^*, \tilde{w}_n^i) - \langle D(q_n^*, \tilde{w}_m^i) \rangle) + \frac{1}{n} \text{tr} \tilde{V}_n^i \tilde{H}_n^{i-1}. \quad (1)$$

ここで、 \tilde{w}_n^i はデータ数 n の例題をニューラルネットワークモデル M_i で学習して得られる推定パラメータであり、 $D(q_n^*, \tilde{w}_n^i)$ は次式で与えられる。

$$D(q_n^*, \tilde{w}_n^i) = \frac{1}{n} \sum_{j=1}^n |y_j - f(x_j, \tilde{w}_n^i)|^2 = \frac{1}{n} \sum_{j=1}^n d(x_j, y_j; \tilde{w}_n^i).$$

また、 $\tilde{V}_n^i, \tilde{H}_n^i, \tilde{t}^i$ は次のように表現される。

$$\begin{aligned} \tilde{V}_n^i &= \frac{1}{n} \sum_{k=1}^n \nabla d(x_k, y_k; \tilde{w}_n^i)^\top \nabla d(x_k, y_k; \tilde{w}_n^i), \\ \tilde{H}_n^i &= \frac{1}{n} \sum_{k=1}^n \nabla \nabla d(x_k, y_k; \tilde{w}_n^i), \\ \tilde{t}^i &= \frac{\langle D(q_n^*, \tilde{w}_m^i) \rangle - \langle D(q_m^*, \tilde{w}_m^i) \rangle - \frac{1}{m} \text{tr} \tilde{V}_n^i \tilde{H}_n^{i-1}}{\langle D(q_m^*, \tilde{w}_l^i) \rangle - \langle D(q_l^*, \tilde{w}_l^i) \rangle - \frac{1}{l} \text{tr} \{\tilde{V}_m^i \tilde{H}_m^{i-1}\}}. \end{aligned}$$

$\langle \cdot \rangle$ は、例題 ζ_n から抽出することによって得られるいくつかの例題 ζ_m, ζ_l から求められる $\tilde{w}_m^i, \tilde{w}_l^i$ の分布での平均を表す。

この規準を「翌日最大電力需要予測」と「マニピュレータの正しい姿勢位置の学習」を行なうニューラルネットワークの構成決定に適用し、その有用性を確認した。図1に各々の問題に適用したニューラルネットワークを示す。

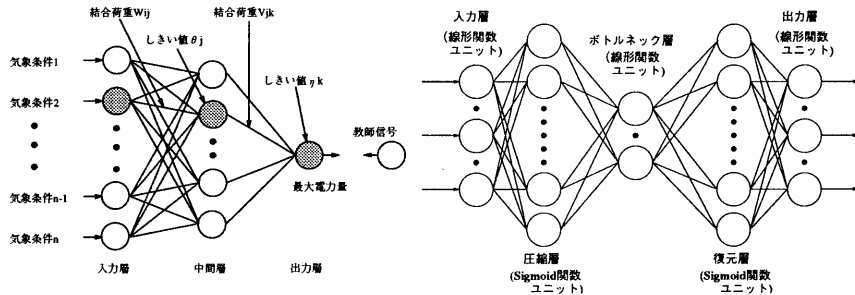


図 1: 最大需要予測のためのニューラルネットワーク (左図) とマニピュレータの姿勢の学習用ニューラルネットワーク (右図)

表 1: 1988, 1989 年度における各モデルの予測誤差

	評価誤差	重回帰モデル	中間層ユニット数固定化 N.N.	中間層ユニット数最適化 N.N.
1988 年度 夏季	RMSE(MW)	668.4	519.2	507.0
	MAPE(%)	3.27	2.50	2.33
1989 年度 夏季	RMSE(MW)	665.7	605.5	564.4
	MAPE(%)	2.90	2.56	2.41
1988 年度 冬季	RMSE(MW)	298.8	298.9	278.0
	MAPE(%)	1.58	1.59	1.48
1989 年度 冬季	RMSE(MW)	372.4	351.5	344.0
	MAPE(%)	1.90	1.80	1.72

翌日最大電力需要予測は、主に気象要因に基づいて行なわれる。夏季と冬季の最大電力需要と気象要因との関係には非線形性が強く、その予測モデルの構築には、非線形の関数近似能力の高いニューラルネットワークの適用の有用性が指摘されている。しかし、ニューラルネットワークの構成、つまりその中間層ユニット数についての議論は十分行なわれておらず、その数をここで提案する規準で決定すれば予測精度の向上が図れるのではないかと考えた。表 1 に重回帰分析によるモデル、経験的に中間層ユニット数を決定したニューラルネットワークモデル (中間層ユニット数固定化 N.N.)、中間層ユニット数を NNIC で決定したニューラルネットワークモデル (中間層ユニット数最適化 N.N.) の 1988, 1989 年度の夏季、冬季の予測結果を示す。表 1 から、1988, 1989 年度夏季および冬季について、いずれの場合も提案する規準を用いて構成を決めたニューラルネットワークモデルの Mean Absolute Percentage Error が小さいことから平均的な予測誤差が他のモデルより小さくなっていることがわかる。また、Root Mean Square Error が小さいことから同モデルが他

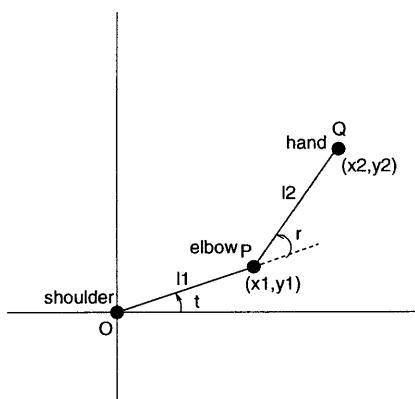


図 2: マニピュレータモデル

表 2: 学習誤差, NNIC の値, 未知データに対するテスト誤差

圧縮復元層ユニット数	8	9	10	11	12	13
学習誤差	1.834	1.762	1.626	1.610	1.594	1.585
NNIC の値	1.844	1.770	1.641	1.611	1.618	1.630
テスト誤差	4.161	4.059	3.975	3.629	3.739	3.848

のモデルより予測が大きく外れることが少ないこともわかる。

図 2 のような平面 2 関節マニピュレータを考える。マニピュレータの正しい姿勢位置の学習とは、図 2 の肘 (elbow) と手 (hand) の位置が観測雑音を加えられて獲得され、その観測データから雑音を除去したマニピュレータの正しい姿勢位置を推定する図 1 右図に示す砂時計型ニューラルネットワークモデルを構築する問題である。表 2 に圧縮、復元層ユニット数、観測データに対する学習誤差、NNIC の値、観測データとは別の未知データに対するテスト誤差を示す。ここで、学習誤差とは入力に雑音の加わった観測データを用い、教師信号にも同じデータを用いた場合のネットワークの出力と教師信号との平均自乗誤差であり、テスト誤差とは入力に観測データとは別の雑音の加わったデータを用いた際のネットワークの出力と雑音を除去したデータとの平均自乗誤差を表す。また、このシミュレーションではボトルネック層ユニット数を 2 としている。表 2 から、NNIC の値が最も小さくなる場合の砂時計型ニューラルネットワークのテスト誤差が最も小さくなっており、雑音除去を最も適切に行なうネットワーク構成が得られていることが確認できる。

本論文ではニューラルネットワークの実環境下での能力を評価する規準を提案し、その有効性をシミュレーションによって示した。