

II - 119

ニューラルネットワークによる日流量予測での未学習データへの適合方法の研究

名古屋工業大学大学院 学生員 ○佐野正嗣

名古屋工業大学大学院 肖 麗亜

名古屋工業大学 正 員 長尾正志

1. はじめに

(1) 目的 入出力関係をブラックボックス的に表現できるニューラルネットワークにおいて、学習データに基づいてネットを構築していくが、その際未学習データに対してもできるだけ内挿的になるように配慮しておけば、精度の高い予測が可能になるものと考えられる。そこで本研究ではそのために有効な学習方法を提案し、重線形回帰モデルとの比較を通して、未学習データに対する適合精度を検討する。この際、過学習による適合精度の低下がないように、未学習データに対して適合精度を調べ、結合係数を決定する。この結合係数による未学習データに対する適合精度と、重線形回帰法による適合精度を比較・検討する。

(2) 対象資料 牧尾ダムの渇水期において、前日までの日流量、日雨量を説明変数とし、翌日の日流量を目的変数とする。これまでの日流量時系列の研究より、日流量については2日前から前日までを用いる。計算には、1969年から1990年まで22年間の12月14日から翌年の2月11日までの60日間を使用した。この22年分のデータの各年の流量の最大値を求め、それを小さいものから順に並べかえる。結果を表1に示す。この順番のうち12番から22番目の年と、流量の最小値を含む年を学習データの年とし、残りの年を検定データとする。学習データの流量(教師信号)を小さいもの順に並べ、1番目から3つおきに170個取る。170番目の流量の値は12.7m³/sであり、22年分のデータの累積分布で0.97の値であった。この170個を学習データに使い、残りを未学習データとする。

表1 各年の最大流量の順位 (流量の単位：m³/s)

西 暦	1973	1980	1975	1969	1990	1981*	1982	1983	1987	1974	1970
最大流量	2.19	3.97	4.70	4.74	4.88	5.53	6.91	7.15	8.31	9.91	13.12
西 暦	1976	1979	1971	1985	1972	1984	1978	1986	1977	1989	1988
最大流量	18.13	19.19	22.66	30.54	30.76	35.49	51.49	69.97	83.86	155.89	230.01

*：流量の最小値を含む年

2. 予測モデル

ニューラルネットワークは入力層、中間層、出力層はそれぞれ単層とし、出力層ユニット数は1とする。評価基準AICを用いてネットワークの構造選定を行った結果より、入力層ユニット数は5、中間層ユニット数は3のモデルが最適となった。なお、入力層ユニット数5とは、3日前からの日雨量ユニット3と、2日前からの日流量ユニット2の和を意味する。

3. 計算方法

(1) 学習データ170個を使い、階層型ニューラルネットワークにより学習させる。この際、学習回数100回ごとの結合係数を5000回まで求める。

(2) 各学習回数の結合係数で各年の未学習データについて観測値と推定値の相関係数を求める。このとき、未学習データは学習データの各年の中から学習データに使ったものを除いたもので、入力値の関係から境界値(27.1m³/s)以下のものをいう。

(3) 各学習回数ごとの相関係数の平均値を求める。この平均値が最大になった回数の結合係数を最適な結合係数とする。

(4) (3)の結合係数を使い、検定データについて観測値と推定値の相関係数を求め、予測精度を調べる。

(5) 同じ 170 個の学習データを使い、重線形回帰法により回帰係数を求め、検定データに対して推定値

を求め、その適合精度を検討する。

4. 結果と検討

(1) 各学習回数ごとの相関係数

学習データの観測値と推定値の相関係数と、未学習データの相関係数の平均値のグラフが図1である。この図より学習回数が増えると学習データに対しては相関係数が増加し安定してくる。しかし未学習データに対しては学習回数が増えるにつれて相関係数は低下していく。これより学習の早い段階から過学習が起きていることがわかる。そこで未学習データの相関係数の平均値が最大になった学習回数 400 回の結合係数を最適な結合係数とする。

(2) 検定データに対する適合精度の比較

最適な結合係数での相関係数と重線形回帰法による相関係数を求めると表2になる。相関係数は平均約12%大きくなっているのが、重線形回帰法と比べ適合精度は高い。なお、1973年はニューラルネットワーク、重線形回帰法ともに相関係数が小さい値になっているが、この年は降雨が少なく渇水となった年であり、観測所において小さな流量が正確に計測されなかったことが原因ではないかと考えている。

表2 検定データにおける相関係数

西 暦	1969	1970	1973	1974	1975	1980	1982	1983	1987	1990	平均
本方法	0.950	0.941	0.365	0.750	0.928	0.757	0.593	0.945	0.906	0.655	0.779
重回帰法	0.886	0.907	0.175	0.682	0.849	0.605	0.554	0.913	0.893	0.495	0.696

最適な結合係数での推定の一例として1987年を経過日数順に示すと図2となる。これより全体的に観測値にかかなり近い推定を行うが、観測値が突発的に大きくなる時には、小さめの推定をしている。この傾向はこの年だけではなく、検定データのすべての年に対してほぼ共通している特徴である。表2で相関係数が小さい値になっている年は、急激な変化に対しては推定がうまくいかないが、それ以外は観測値に近い推定を行っている。最後に観測値と推定値とを比較し、ばらつきを示すと図3となる。図中の斜線上では観測値と推定値が等しいことを意味する。図3よりも先述の特徴が読みとれるようである。

5. 結論

① できるだけ内挿的になるような代表データで学習させたので、かなり高精度な推定ができることを示した。

② 学習の早い段階から過学習が起きており、比較的少ない学習回数で予測した方がよいことを示した。

③ 本方法では重線形回帰法よりもほぼ1割以上相関係数が大きく、良好な予測結果が得られた。

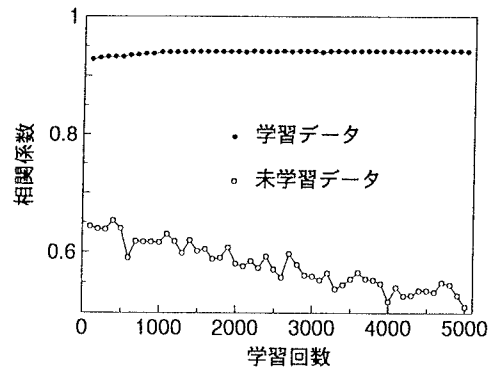


図1 学習回数と相関係数の関係

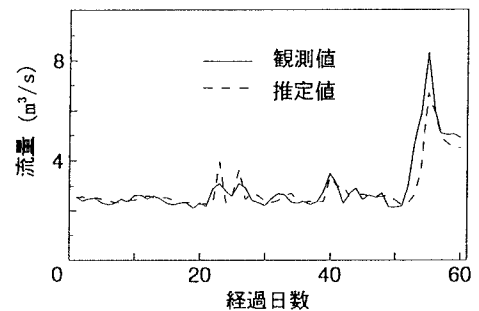


図2 経過日数順の観測値と推定値の対比

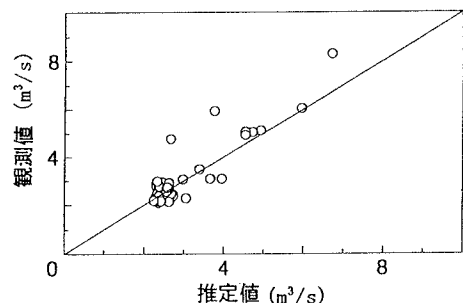


図3 観測値とその推定値の比較