

リカレント型ニューラルネットワークによる渇水期日流入量系列の推定精度の研究
Forecast of Daily Inflow Series in Dry Season by Recurrent Neural Network Model

鈴木正人*・長尾正志**

By Masato SUZUKI and Masashi NAGAO

This study presents an application of neural network model to forecast of daily inflow series in a dry season with hydrological data at Makio dam basin area. The daily inflow and daily precipitation of a period before the day of forecast are used for the input to neural network. The lead time are one day and five days. In the case of one day, the average of relative error to data is about 20%. And in the case of five days, it is shown that the study of recurrent neural network with feed back is effective to progress of precision of forecast.

Keywords : neural network, forecast, daily inflow, dry season

1. はじめに

入力と出力の関係が物理的に解明されていない場合でも、入力と出力が与えられるとその対応関係を満たすように学習し、自らシステムを構築するニューラルネットワークは数多くの分野で成果を上げている。本研究は、いくつかの条件のもとで、ニューラルネットワークにより渇水期日流入量を推定し、その推定精度を比較、検討することにより渇水期日流入量系列の推定へのニューラルネットワークの適用の有効性を示すことを目的としている。

具体的には、木曾川水系牧尾ダム流域の12月日流入量を予測対象とした。まず最初に1日先の日流入量の推定を行った。この際の入力とは推定日より前の日流入量、日降水量であるが、前何日分の入力が適当かを検討している。これらの結果を踏まえて、さらに長期間にわたる予測を意図して5日先までの日流入量系列をまとめて推定し、その可能性を検討した。

2. 1日先流入量の推定

2.1 計算に用いたネットワーク

基本的な構成である、入力・中間・出力の3層からなる階層型のネットワークを計算に、学習アルゴリズムには教師付き学習では最も一般的なバックプロパゲーション法¹⁾を用いた。入力層のユニット数は、入力として与える日流入量、日降水量の数に一致させ、最も少ない場合で、1日前の流入量・降水量を入力

* 正会員 工博 岐阜工業高等専門学校講師 環境都市工学科

(〒501-04 岐阜県本巣郡真正町)

** 正会員 工博 名古屋工業大学教授 工学部社会開発工学科

(〒466 愛知県名古屋市昭和区御器所町)

としたときの2個、最も多い場合で前5日間の流入量と降水量を入力としたときの10個である。中間層のユニット数は入力層のユニット数と等しくし、出力層のユニット数は1日先流入量に対応して1個である。中間層のユニット数に代表されるネットワーク構造の適切な選定²⁾は、推定の安定性や学習効率の面から考えて重要な課題の一つと考えられるが、本研究では実際に計算を行っていく際に構造の選定をそれほど考慮しなくても、ある程度の推定精度が得られることが分かったので上述のような基本的なネットワーク構造を用いた。

2.2 計算対象データ

木曽川水系牧尾ダムの1969～1991年の12月の日流入量と日降水量を用いた。なお、計算には、流入量、降水量のそれぞれについて最大値を用い0～1の値になるように基準化した。推定する流入量は、入力として前5日分の日流入量を用いるので各年当たり12月6～31日の26日分、全体では598日分になる。

2.3 計算方法

推定精度の検討といった観点からは、推定値と実際のデータを比較することが必要になる。そこで本研究では、全23年分のデータのうち各1年分を除いたもの(572組)を教師信号としてネットワークに与え学習させ、学習済みのネットワークを用いて教師信号から除外した年の流入量(26日分)を推定し、得られた推定値と実際のデータを比較する事にした。学習の程度は、学習段階における出力値と教師信号(=実データ)の相対誤差の和により判断した。相対誤差は次式で表されるが、この和が小さいほど教師信号と出力値が良く合っている、すなわち学習の回数が多いことになる。

$$\text{相対誤差} = |\text{実データ} - \text{出力値}| \div \text{実データ} \quad (1)$$

ただし、ここで注意を要するのが、“学習回数を多くすること”が必ずしも“良い学習を行うこと”とは一致しないことである。ニューラルネットワークには、過剰に学習させ過ぎる(過学習する)と教師信号とそれほど変わらない入出力関係を持つ場合には、入力値に対して精度の高い出力値が得られるが、教師信号とは異なる入出力関係を持つ場合にはその差異がわずかでも出力値の精度は極端に悪くなる性質がある³⁾。そこで、本研究では学習段階における出力値と教師信号の相対誤差の和をerrorと表記し、14,12,10,8(それぞれデータ1個当たりの相対誤差で約40,35,30,25%に相当)の4段階設定し、学習の程度と推定精度との関係を検討した。

2.4 推定精度の表現

本研究では推定精度を推定値と実データの相対誤差により表現した。一般的には、推定値に対する相対誤差の平均が小さいことは推定精度が良いことを意味すると思われる。しかし、先述のように過学習が起きると、数は少ないものの極端に大きな相対誤差を持つ推定値が全体の平均を引っ張ることも考えられるので、平均のみならず相対誤差の最大や頻度分布も併せて検討した。

2.5 適用計算結果

1972年の12月日降水量の推定結果を図-1に示す。凡例のerror14,error8は先述のように、学習段階における教師信号と出力値の相対誤差の和が14,8の場合に対応している。図を見ると実際のデータでは12日に流入量の山があるのに対して推定値では13日に山があるというように1日のずれはあるものの、変化傾向そのものは比較的良く合致している。流入量のピーク位置がずれたのは、流入量、降水量ともに日を単位としているので、降雨が2日間にまたがったり、また逆に短時間の降雨が1日のどの時間帯にあったかがデータに現れず、今回対象とした流域が、比較的流出時間が短いこともあり、1日未満の流出時間のずれが推定結果に反映されないためであろう。学習の程度による差は全体的にみてerror14の場合の推定値が全体的に実データより少な目になるのに対して、error8の方では実データの変動傾向がほぼ再現されている。

つぎに、入力として流入量5日、降水量5日を与えた場合の推定値と実データの相対誤差の頻度を図-2に示す。図によるとerrorが小さくなるほど、つまり学習の回数を多くするほど出現頻度のピークが量の少ない方に近づいていて、全体的にみると推定精度も上がっているといえる。error8の場合は、全体の約

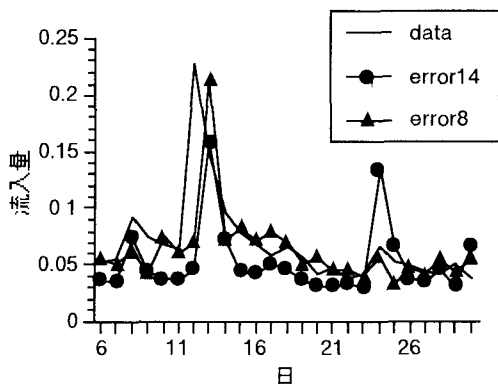


図-1 1972年日流入量推定値
(入力：流量5日 降水量5日)

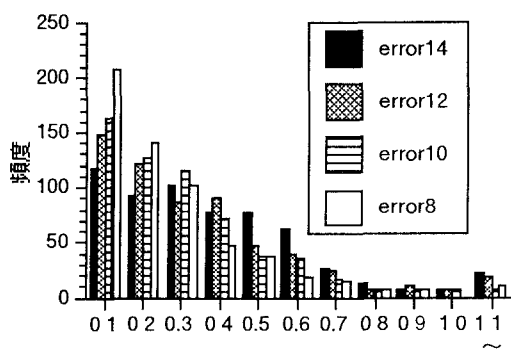


図-2 推定値相対誤差頻度
(入力：流量5日 降水量5日)

表-1 推定値の相対誤差の平均(%)

入力データ		学習段階の相対誤差の和(error)			
流量	降水量	14	12	10	8
5日	5日	35.8	31.4	28.3	27.8
5日	3日	35.3	29.3	25.6	21.3
5日	1日	35.5	34.3	25.1	28.3
3日	5日	35.3	31.1	28.7	28.6
3日	3日	36.2	30.1	26.2	22.1
3日	1日	38.0	31.1	25.1	-
1日	5日	39.1	32.4	28.5	29.3
1日	3日	40.3	32.0	24.9	-
1日	1日	45.5	37.5	27.9	-

表-2 推定値と実データの相関係数

入力データ		学習段階の相対誤差の和(error)			
流量	降水量	14	12	10	8
5日	5日	0.536	0.585	0.547	0.591
5日	3日	0.524	0.613	0.538	0.636
5日	1日	0.575	0.645	0.645	0.288
3日	5日	0.585	0.597	0.582	0.613
3日	3日	0.585	0.605	0.627	0.658
3日	1日	0.594	0.635	0.672	-
1日	5日	0.606	0.602	0.623	0.597
1日	3日	0.565	0.599	0.605	-
1日	1日	0.587	0.620	0.681	-

60%に相当する359日分の推定値において相対誤差が20%未満になっている。error14の場合、相対誤差が20%未満になるのは209日分なので、学習の回数を多くするほど精度が高い推定値が得られやすいといえる。しかし、error8の場合でも相対誤差が110%以上になる推定値が11日分ある。

そこで、入力に用いる流入量と降水量の日数をそれぞれ1, 3, 5日として、それらを組み合わせて入力とした場合の相対誤差の平均を表-1に示す。error8の列で値が記入されていないのは、学習を30万回行っても所要の誤差が得られなかったので計算を打ち切ったことによる。図-1に対応する流入量5日、降水量5日の場合、errorが小さくなるほど推定値の相対誤差の平均は小さくなるものの、error10とerror8の比較においては、図-1でみられたほどの差はない。これは、error10の場合の実データと推定値の相対誤差の最大値が323%であるのに比べてerror8の場合の最大値が2434%と、異常に大きい。したがってこの最大値の存在が全体の平均を引っ張り、error8まで学習させると過学習に相当する現象が起きていると思われる。入力データの組み合わせによる差はそれほどみられないが、流入量5日、降水量3日の組み合わせが推定精度が良さそうである。

また、推定値と実データとの相関係数が大きいことも推定精度が良いことを意味すると思われるので、表-1と同様に、入力に用いる流入量と降水量の日数の組み合わせにおける、推定値と実データとの間の相関係数の平均を表-2に示す。全体的には表-1と同じ傾向を示し、errorが小さくなるほど相関係数も大きくなり推定精度も良くなっているが、細かく比較すると若干傾向が異なる場合がある。例えば入力として流入量、降水量ともに5日の場合、表-1での相対誤差の平均ではerror10の方がerror12に比べて相対誤差の平均が小さくなっているが、表-2の相関係数ではerror10の値の方が小さく推定精度が悪くなる、というように表-1と表-2で精度の関係が逆転している。特に、入力が流入量5日、降水量1日でerror8の場合の相関係数は、他に比べて極端に小さくなっている。このように、相対誤差と相関係数で精度

という観点からは逆の傾向が現れることがあるのは、大きな値の実データに対して推定値が小さい場合、相対誤差では式(1)で示したように、実データと推定値の差を実データで割るため、その差はそれほど大きくは現れてはこないためであろう。

3. 推定期間を5日間とした場合

3.1 計算に用いたネットワークと計算方法

1日先の日流入量はある程度の精度で推定できることが示されたが、有効な水利用といった観点からはより長い期間の流入量の予測が望まれる。そこで推定期間長を5日間に設定し、5日分の流入量を一度に推定することを試みる。つまり、ニューラルネットワークにおいて出力層のユニット数を5個として、各ユニットからの出力を各日の流入量に対応させた。また、入力には1日先流入量の推定結果を踏まえて、推定期間より前5日分の日降水量と、前5日分の日流入量とを用いた。

ところで、1日先の流入量を推定する場合に比べて、5日先までの日流入量をまとめて推定する場合には、出力値が時系列としての性質を持っており、入力に用いる推定日以前の日流入量の系列と、出力される推定日流入量の系列の間には系列内に時間的な相関、つまり自己相関性を有していると思われる。このような時系列データを対象としてニューラルネットワークを用いる場合は、学習段階においてネットワークからの出力を入力にフィードバックさせるリカレント型のネットワークを用いることにより自己相関性をうまく捉えることができ、推定精度が良くなることが知られている⁴⁾。そこで、本研究では、フィードバックさせた場合とさせない場合の二つの学習方法を用い、両者の推定精度を比較検討した。なお、具体的に学習段階においてフィードバックさせる方法は、

Q_j ($j=t+1, t+2, \dots, t+5$): 推定期間における教師信号 (すなわち実データ)

$Q_j(L)$ ($j=t+1, t+2, \dots, t+5$): 学習段階における出力値

とする。そして、学習段階における出力層の各ユニットからの出力値と教師信号との差の絶対値である、

$$|Q_j - Q_j(L)| \quad (j=t+1, t+2, \dots, t+5) \quad (2)$$

を入力層に再入力させて入力情報として与える。学習が終了した段階では、 $Q_j \div Q_j(L)$ となるので、(2)式の値はほぼ0となる。したがって、学習終了後のネットワークを用いた推定では、学習段階において(2)式の値を与えた入力層の入力情報としては0を与えればよく、推定の段階ではフィードバックは見かけ上消滅している。

入力層のユニット数は、フィードバックさせない場合が10個 (前5日分の日流入量と日降水量)、させた場合がフィードバック分の5個を加えた15個である。中間層のユニット数は、入力層と出力層のユニット数の中間的な数として、フィードバックさせない場合が8個、させた場合は10個とした。学習アルゴリズムはバックプロパゲーション法を用い、1日先日流入量の推定と同様に出力値と相対誤差の和をerrorで表し、errorを4段階設定することで学習の程度と推定精度との関係を検討した。

なお、用いたデータも1日先流入量の推定と同じ12月の牧尾ダムのもので、12月1～5日の日流入量・日降水量を入力し、12月6～10日の日流入量を推定、12月6～10日の日流入量・日降水量を入力し、12月11～15日の日流入量を推定・・・、というように1年当たり5組の推定を行った。なお、簡単に、例えば12月6～10日の日流入量を推定した場合、12月6日を1日先、12月7日を2日先、・・・、12月10日を5日先と表記している。

3.2 適用計算

1972年の日流入量推定結果を図-3、4に示す。図-3はerror14の場合、図-4はerror8の場合であり、図-4の方が学習回数が多くなっている。図-3をみると、フィードバックの有無による差はほとんどない。また実データでみられる流入量のピークが再現されていない。また、4日先の推定値が小さくなる傾向があり、ちょうど5日ごとに極端に小さい推定値が出現している。これに対して、図-4の場合は、やはり流入量のピークは再現されていないものの、フィードバックの有無による差があり、フィードバックさせたものの方が実データと適合している。また、図-3でみられた4日先の推定値が小さくなる傾向は

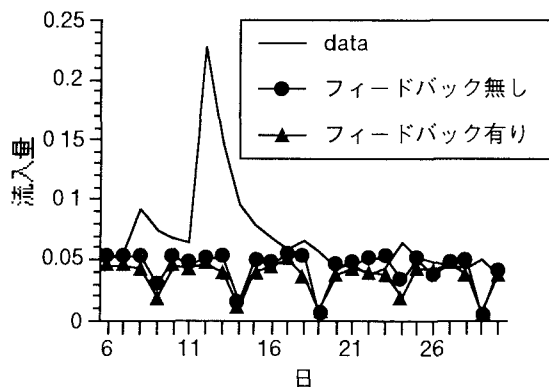


図-3 1972年日流入量推定値(error14)

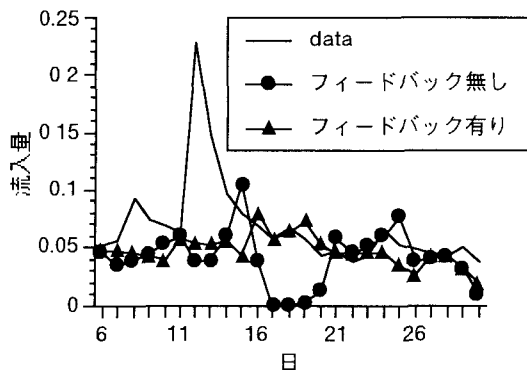


図-4 1972年日流入量推定値(error8)

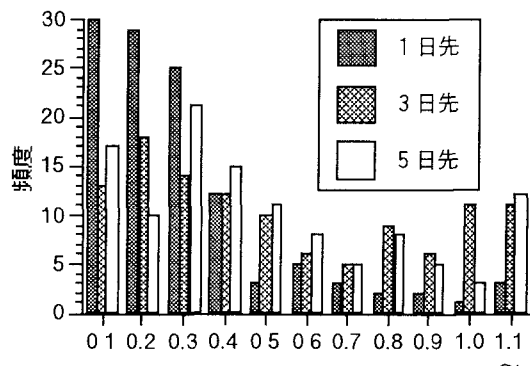


図-5 推定値相対誤差頻度（フィードバック無し）

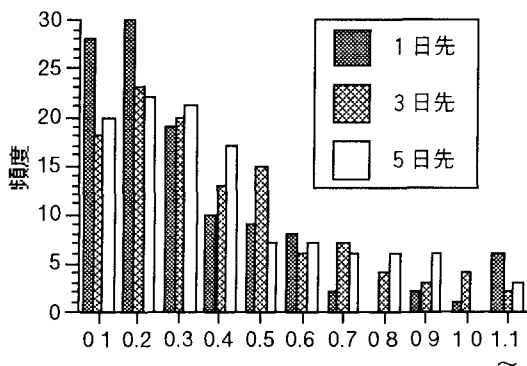


図-6 推定値相対誤差頻度（フィードバック有り）

フィードバックの有無によらず出現していない。

つぎに、error10まで学習させた場合の実データと推定値の相対誤差の頻度について、フィードバックさせなかった場合を図-5に、フィードバックさせた場合を図-6に示す。どちらの場合も、1日先の推定値については頻度のピークが相対誤差が小さい領域にあり、ある程度の推定精度が期待できる。ただし、3日先、5日先の推定誤差はフィードバックの有無により差がでていいる。フィードバック無しの場合、誤差はかなりばらつき、相対誤差が1.1(110%)以上になる場合も10個以上みられる。一方フィードバックさせた方は、3日先、5日先のいずれにおいても1日先に比べればばらついているものの、相対誤差が0.1～0.2(10～20%)の辺りにピークを持ち、1.1以上の相対誤差になる場合もそれほど多くない。したがって、この図から判断するに1日先より長い期間の推定精度において、フィードバックの効果がでていると思われる。

つぎに推定値と実データの相対誤差の平均について、フィードバック無しの場合を表-3に、フィードバックさせた場合を表-4に示す。両者を比較すると、1日先の相対誤差の平均はあまり差がなく、学習の程度によってはむしろフィードバックさせないものの方が平均が小さくなっている場合もある。しかし、2～5日先については明らかにフィードバックさせた方が相対誤差の平均が小さくなっている。表-3で特徴的なのは、2～5日先においてerrorが小さくなくても、つまり学習回数を多くしても、相対誤差の平均の減少に連ならず学習の程度と推定精度にはほとんど関係がないことである。一方、表-4では程度の差こそあれ、1～5日先の全てにおいて学習回数を多くしたerror8の場合の相対誤差の平均が小さく、学習効果が現れている。また、同じように相対誤差の最大値を表-5、6に示すが、フィードバックさせない場合は、error10,8において以上に大きな値が出ているが、フィードバックさせた場合は4日先の推定値

表-3 推定値の相対誤差の平均(%)
(フィードバック無し)

	1日先	2日先	3日先	4日先	5日先	平均
error14	38.9	48.4	53.3	69.1	57.9	53.5
error12	32.7	42.9	49.8	63.5	49.6	47.7
error10	26.0	80.5	80.6	67.8	54.6	61.9
error8	29.4	78.6	83.8	75.0	68.5	67.1
平均	31.8	62.6	66.9	68.8	57.6	57.5

表-4 推定値の相対誤差の平均(%)
(フィードバック有り)

	1日先	2日先	3日先	4日先	5日先	平均
error14	37.6	34.8	37.2	65.6	46.0	44.2
error12	34.7	40.9	46.6	72.8	44.1	47.8
error10	30.2	32.7	36.0	58.7	37.3	39.0
error8	24.2	31.3	34.2	49.4	31.0	34.0
平均	31.7	34.9	38.5	61.6	39.6	41.3

表-5 推定値の相対誤差の最大(%)
(フィードバック無し)

	1日先	2日先	3日先	4日先	5日先	平均
error14	231	227	221	930	440	410
error12	210	374	249	1684	366	577
error10	141	2463	2259	1581	499	1389
error8	551	2462	2240	1797	720	1554
平均	283	1381	1242	1498	507	982

表-6 推定値の相対誤差の最大(%)
(フィードバック有り)

	1日先	2日先	3日先	4日先	5日先	平均
error14	223	203	116	680	200	284
error12	142	122	140	1699	151	451
error10	192	342	196	1809	446	597
error8	238	260	251	1800	159	542
平均	199	232	176	1497	239	468

に関して大きな値がみられるものの、その他ではそれほど大きな値にはなっていない。なお、フィードバックさせた場合に4日先の推定誤差が大きくなるのは、図-3で示した推定結果において4日先の推定値が小さくなる傾向を反映したものと思われる。

以上の結果を総合的に判断して、5日間の日流入量をまとめて推定する場合には、フィードバックさせることにより推定精度が上がるとと思われる。むしろフィードバックさせないと1日先より長い期間の推定はほとんど意味をなさないのでないだろうか。

4. まとめ

本研究により、ニューラルネットワークによって渇水期における1日先の日流入量の推定が、相対誤差で20%程度の精度で推定可能ことが示された。また、5日先までの推定については、1日先の推定値に比べれば精度は劣るものの、学習段階で出力値をフィードバックさせるリカレント型のネットワークを用いることにより推定精度が向上することが分かった。今回の計算では、学習段階において教師信号を吟味することなく、存在するデータを全て教師信号として与えたが、推定期間内における降雨の有無などにより教師信号を分類して学習させることにより、より精度の高い予測が可能になるものと考えている。

5. 参考文献

- 1) ニューロンネットグループ・桐谷滋：入門と実習ニューロコンピュータ，技術評論社，pp.28-84,1989.
- 2) 長尾正志・鈴木祐一・佐野正嗣：ニューラル・ネットワークによる渇水期日流量の予測（1）-階層型ネットワークの構造選定-，土木学会中部支部平成6年度研究発表会講演概要集，pp.269-270,1995.
- 3) 例えば，計測自動制御学会編：ニューロ・ファジィ・AIハンドブック，オーム社，pp.1160-1167,1994.
- 4) 松葉育雄：バックプロパゲーションによる特長抽出，数理科学，No.338,pp.31-38,1991.