

第II部門

ニューラルネットワークを用いた積雪流域の河川流量予測

京都大学工学部

学生会員 ○滝口修司

京都大学大学院工学研究科 正会員 Kim Sunmin

京都大学大学院工学研究科 正会員

立川康人

京都大学大学院工学研究科 正会員 市川温

京都大学大学院工学研究科 正会員

萬和明

1 序論 水文学の分野において自然現象を解析または予測するときはその現象の物理的な過程を解明し、モデル化することが一般的である。しかし、これらモデルを実用可能なレベルで構築するには高精度のデータや現象の詳細な解明が求められ、モデル化が困難な場合が多い。このような現状を踏まえ、本研究では近年目覚ましい進化を遂げているニューラルネットワークを用いた事象の予測に注目し、積雪流域の河川流量をニューラルネットワークによって予測した。また、積雪流域の河川流量を予測するとともに、ニューラルネットワークの学習効率化とニューラルネットワークによる入力データの抽出についても考察を行った。

2 ニューラルネットワークを用いた積雪流域の河川流量予測

2.1 ニューラルネットワークについて ニューラルネットワークは、生物の脳内における学習メカニズムにヒントを得てそれをコンピュータ上で再現したアルゴリズムである。ニューラルネットワークには様々な種類や名前があるが、本研究では中間層が1つで事前学習を行わないものを Traditional Neural Network(TNN)、中間層が2つ以上あり、AutoEncoder によって事前学習を行うものを Stacked AutoEncoder(SAE) と呼び、これらを用いて積雪流域の河川流量を予測した。

2.2 ニューラルネットワークの構築及びデータ成形 本研究では群馬県にある奈良俣ダムへの流入量を周辺地域の平均気温、日降水量合計、最深積雪深、日降雪量合計、平均日照時間、平均風速のそれぞれ当日、1日前、2日前分の6項目18種類の入力データを基に予測することにした。なお、入力データについてはデータ成形として一般的な正規化を行った。また、予測を行うにあたってニューラルネットワークのパラメータを表1に記すようにそれぞれ設定した。

2.3 予測結果 予測を行うにあたっては、1993年～2002年までのデータを基にニューラルネットワークを学習させつつ、2003年度におけるテスト結果を基準と

表1 ニューラルネットワークの構築

パラメータ	設定値	パラメータの種類
入力データ	6項目18種類のデータ	
隠れ層の数 ニューロンの数 活性化関数	1,2,3,5 9,18,36,54,90 シグモイド関数	外形を司るパラメータ
Learning Rate Learningrate Scale モメンタム ドロップアウト率	0.2,0.4,0.6,0.8,1.0 1.0 0.5 0	学習を司るパラメータ
Batchsize Number of Epochs	15,30,90,180,365,730,1095 5,25,100,250,500,1000	学習回数を司るパラメータ

して最適なニューラルネットワークの構造を選択した。そして、選択された構造のニューラルネットワークによって2004年～2007年までの奈良俣ダム流入量を予測したところ、図1と図2の結果を得た。赤線が予測値、黒線が観測値をプロットしたものであり、TNNとSAE共に特徴を捉えた予測を行えていることがわかる。TNNとSAEの予測精度に関してはどちらもNash指標では0.7程度、1日当たりのRMSEは $3\text{m}^3/\text{s}$ 程度であり、中間層の数の違いや事前学習の有無による予測精度の差はあまり見られなかった。

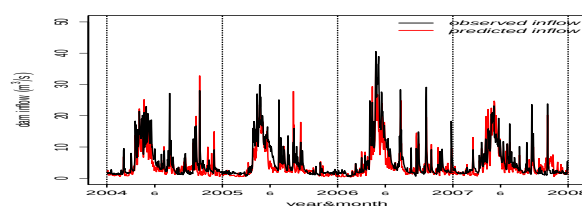


図1 TNNによる2004～2007予測結果例

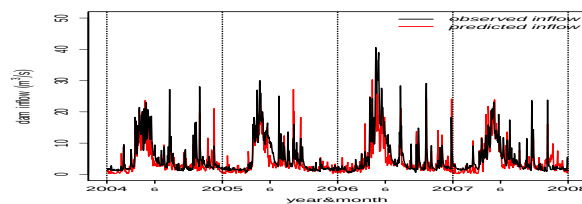


図2 SAEによる2004～2007予測結果例

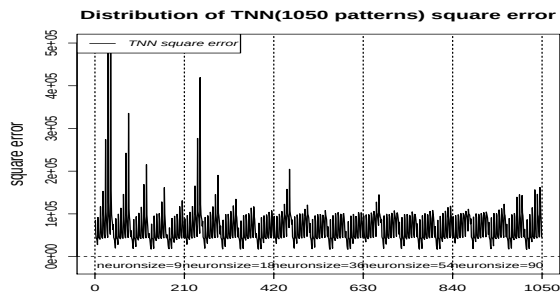


図3 TNNの二乗誤差合計値分布状況全体像

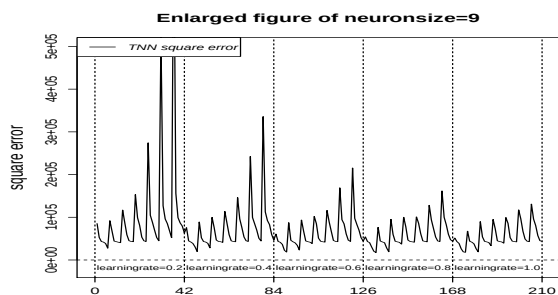


図4 neuronsize=9部分の拡大図

3 ニューラルネットワークのパラメータが与える学習精度への影響 本章では、ニューラルネットワークが学習を行う際に表1に示したパラメータの設定が学習精度にどのような影響を与えているのかについて考察を行う。考察を行うにあたってはニューラルネットワークの外形を司るパラメータ(隠れ層の数、ニューロンの数)、学習を司るパラメータ(Learning Rate)、学習回数を司るパラメータ(Batchsize、Number of Epochs)のオーダー順序で各パラメータを昇順で並び替え、ダム流入量の予測値と観測値の1993年～2002年までの二乗誤差の合計値をプロットした。TNNの場合では図3のようになり、図4は図3の一番左側neuronsize=9の部分拡大した図である。これらの図からTNNの学習については中間層のニューロンの数を54、Learning Rateは1.0、Batchsize(サンプリングする際に取り出す個数)は15、Number of Epochs(計算反復回数)は1000、またはこれに似た構造に設定すると効率よく行えることが分かった。SAEについても同様の方法をプロットして解析したところ、各中間層におけるニューロンの数を9、Learning Rateは0.6、Batchsizeは15、Number of Epochsは1000、またはこれに似た構造に設定すると効率よく行えることが分かった。

表2 TNNにおける入力データ抽出の例

	あるTNNの場合	学習への寄与度
1993～2002の二乗誤差合計値	16814	
平均気温に関する3種類の入力データを0にした場合	31180	○
日降水量合計に関する3種類の入力データを0にした場合	28563	○
最深積雪深に関する3種類の入力データを0にした場合	18506	△
日降雪量合計に関する3種類の入力データを0にした場合	16902	×
平均日照時間に関する3種類の入力データを0にした場合	18952	△
平均風速に関する3種類の入力データを0にした場合	45737	○
学習に不要な項目	日降雪量合計	
学習に必要な項目	平均気温、日降水量合計、平均風速	

4 ニューラルネットワークによる入力データの抽出 入力データの抽出方法としては、6項目18種類の入力データのうち1項目3種類の入力データ全てに0を代入して(例えば平均気温の当日、1日前、2日前分のデータ全てに0を代入する)、1993年から2002年までの二乗誤差の合計値が小さくなれば、その入力データは学習への寄与度が低いと判断し、逆に大きくなった場合は、その入力データは学習への寄与度が高いと判断した。ここで、1項目3種類の入力データ全てに0を代入している理由は、ある1項目のデータ全てに0を代入するとそのデータに関する計算過程が全て0となり、二乗誤差合計値の変化から学習におけるそのデータの寄与度を判断しやすいためである。表2にTNNにおける結果の例を示した。

先述の方法でTNNとSAEで実験を行った結果、ニューラルネットワークが奈良俣ダムの流入量を予測するために学習を行う際に、平均気温、日降水量合計、平均風速に関する9種類のデータは寄与度が高いが、日降雪量合計に関する3種類のデータは寄与度が低いことが分かった。

5 結論 本研究では主に3つの点に注目し、検証を行ってきた。まず初めにニューラルネットワークを用いた積雪流域の河川流量予測については、TNNとSAE共に比較的良好な予測結果を得ることが出来た。

次にニューラルネットワークのパラメータが与える学習精度への影響について調べたところ、Batchsize(サンプリングする際に取り出す個数)は小さめに設定し、Number of Epochs(計算反復回数)は大きめに設定すると学習を効率よく行えることが分かった。

最後にニューラルネットワークによる入力データの抽出について調べたところ、用意した入力データの中で学習を行う際に寄与度が高いものと低いものがあることが判明した。