

深層学習を用いたダム貯水量予測におけるデータの与え方

A study on How to Give Data in Dam

Water Resources Volume Prediction Using Deep Learning

20N3100046E 八木 将 (都市人間環境学専攻設計工学研究室)
Sho YAGI/ Design Lab.

Key Words : DeepLearning, Dam, WaterResources, YoshinoRiver

1. はじめに

日本の河川は「狭い・急勾配・短い」といった特徴があるため、降った雨が短期間で海へ流れてしまう。そのため、洪水が多いことはよく知られているが、それらの特徴は同時に渇水が起こりやすい原因ともなっている。

日本の年間降水量は長期的にみると減少傾向にある。さらに、近年の異常気象の増加から水資源の安定した配給はより難しくなっている。洪水に比べて渇水は重要視されていないが、水資源問題は家庭用水・都市活動用水・工業用水・農業用水と広範囲の人々の生活に影響を及ぼす。加えて、再生可能エネルギーである水力発電は世界的に注目を集めているため、今後は水資源を最大限に活用するためにダム貯水量を長期的に把握する必要がある。

2. 研究の目的

渇水対策となるダムは洪水・渇水どちらの災害の対策ともなる施設である。しかし、それぞれの現象の特徴とダムの機能から分かるように、この先どちらの現象が起こるかを予測できないと災害に対する正しい貯水量がわからず災害がおこってしまう。実際、2016年6月から計5河川で行われた利根川水系での取水制限は、5月からの過剰放流が原因であると非難を浴びたことがあった。

安定的な水資源の供給に加え、恒久的な発電に向けても期待されているダム貯水量の減少を長期間のリードタイムを取って予測することを目標に研究を行っていた。

降水が河川に通りダム貯水量となるリードタイムを超えた予測を行うため、深層学習を用いて年毎や、降雨イベントごとに発生するトレンドを捉えようと研究を行っていたが、リードタイムを超えた予測はうまくいかなかった。そこで、リードタイムの穴を埋めるために、それまでに使っていたデータに加えて、天気予報などの現在や近い将来予測可能になると考えられる未来に関するデータを入力データに加え、それらの与え方による結果の変化を調査する。

表-1 本研究で利用した早明浦ダムの基本諸元

貯水量に関する早明浦ダムの基本諸元		
有効貯水容量 (千m ³)		289000
利水容量 (千m ³)	洪水期又は設定なし	173000
	非洪水期	173000
サーチャージ水位 (EL. m)		343
常時満水位 (EL. m)		331
最低水位 (EL. m)		275
制限水位		水位
1次制限水位 (EL. m)		329.5
設定期間		
7月1日	～	10月10日
確保貯留量 (千m ³)		基準日
1	104000	2月20日
2	147000	4月5日
3	173000	5月5日
4	173000	5月20日
5	147000	6月5日
6	147000	6月30日
7	156000	7月5日
8	164000	7月10日
9	173000	7月15日
10	173000	8月15日
11	147000	8月31日
12	147000	9月30日
13	158000	10月10日
14	158000	11月10日
15	104000	12月31日

3. 対象とする地域

本研究で対象とする地域は、国内で渇水の被害を最も受けている四国地方の吉野川に位置する四国地方最大の多目的ダムである早明浦ダムとした。

四国地方は、深刻な取水制限が長期間続くような渇水が高頻度で起きており、国内では数少ない慢性的な水不足地域である。長年渇水に苦しめられておりダムの数も多いため利用可能なデータは豊富である点と、積雪が少なく他地方に比べて水資源量の推定が容易である点が機械学習の対象として適していると考えた。

また、吉野川は水流が四国4県に及ぶ水系であり四国内で流域面積が最も広く流域内人口も多い。加えて、香川県民の水源とも呼べる香川用水の水源である早明浦ダム・池田ダムも流域内にあるため、四国で一番の重要河川といえる。その吉野川は渇水リスクを頻度と災害規模で考えると日本で最も渇水リスクが高い河川だといえる。

4. 研究手法

ダムの流入量や貯水量を長期的に予測する研究は、タンクモデルやNNで行われているが、洪水を意識した研究が多く、NNで2日を超えた予測を行っている研究は見受けられなかった。

(1) 入力変数

本研究の予測NNは18種類の基本変数に将来データを加えたものを入力変数として予測を行う。a)の基本データは予測時点の3日前まで、b)の将来データは未来の値として予測時点当日までのデータを与える。

NNは活性化関数にシグモイド関数やtanh関数を使うため、入力変数を0~1までの数値に整形しなおす必要があり、この操作を行わない場合エラー等で異常を発することはないがただ精度が著しく落ちる。そのため、各入力変数は各値からそのラベルの値の最小値を除算したのちに、処理した各ラベルの最大値で割ることで0~1の値に整形している。

a) 基本データ

基本変数は日付、早明浦ダム・池田ダム・富郷ダム・柳瀬ダム・新宮ダムの貯水量、早明浦ダム・池田ダムの放流量・流入量、早明浦ダムの集水域に位置する5地点での雨量、そのうち2地点での平均気温、早明浦ダム下流部に位置する1地点での吉野川本流の水位、表-1に示した早明浦ダムの日ごとの確保流量を含む利水容量、及び常時満水位、計20種類の日毎のデータである。それぞれの値の地理的位置は図-1に示すとおりで、他の支流の河川水位や、早明浦ダム以外の集水域にある観測所の降雨データなども試しに入力していたが、どれ

表-2 降水量の分類

降水量(mm/d)	0	~3	~5	~10	~20	~30	~50	~80	80~
レベル	0	1	2	3	4	5	6	7	8

表-3 天気の種類

晴	快晴	曇	雨	大雨	雪	その他
0	0	0	1	1.5	1	0
基準		後		後がない場合		
時々・一時がある		時々・一時がある		全体を×2		
ある	ない	ある	ない	時々	一時	
×0.35	×0.5	×0.35	×0.5	×0.15	×0.15	
例. 曇時々雨後雨→曇(0×0.35)時々雨(1×0.15)後雨(1×0.5)=0.65						



図-1 各観測所の位置

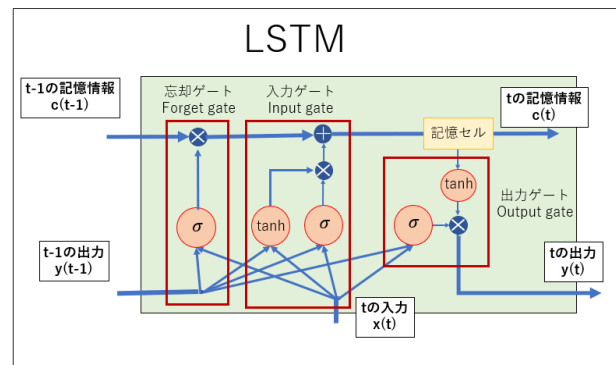


図-2 LSTMの概要

も結果に与える影響は非常に小さく、計算時間が増えていくだけだったので基本変数からは外した。また、入力データとしての有効度が高そうに見えた観測値はほかにもいくつかあったが欠測が多くデータとして利用できないものもあった。これらのデータは将来データを与えない状態での予測結果を指標として最適なものを選んだ。

早明浦ダムだけでなく池田ダムの放流量・流入量デ

ータも入った理由としては、池田ダムは貯水量が十分でない場合に早明浦ダムへ放流を要請することがあるためだと考えられる。

b) 将来データ

将来データの与え方は計3種類、5地点での降水の有無、表-2のように分類した5地点での降水レベル、高松・徳島で観測された天気を表-3のように数値化したものをそれぞれ与えて、結果を比較する。

(2) 予測ニューラルネットワーク

本研究では、時系列データを扱うことに長けている再起型ネットワーク(Recurrent Neural Networks), そのなかでもRNNにおける勾配消失という問題を解消したネットワークモデルである長短期記憶(Long Short-Term Memory)を利用する。RNNでは長い入力データを与えるときバックプロパゲーションを行った時点で、過去のデータをどの程度参考にするべきかという情報を保存している勾配が消失してしまう問題が起こりやすく問題になっていた。しかし、LSTMでは使わなくなったデータをメモリから消す役割の忘却ゲートと過去の情報を保管する記憶セルを設けることで勾配消失を解消している。

NNモデルはTensorflowをバックエンドとしたKerasを利用して構築し、ハイパーパラメータの最適化はoptunaを利用して、ベイズ最適化の一種であるTPE(Tree-structured Parzen Estimator)を利用して表-3のように決めた。TPEはパラメータ同士の相関関係を考慮せず、独立を仮定し確率密度関数の推定をモデル化している点が従来のベイズ最適化と異なる、本研究のモデルではどちらの手法も同じ結果を出せることができるがTPEの方が少ない探索数で結果を出していたため、計算時間を考慮した結果こちらを採用した。

検証(validation)は学習データを等分しそれぞれを検証用データとして計算し検証データによる結果の偏りを減らす交差検証(cross-validation)は有名であるが、本研究では時系列データを扱うため交差検証(cross-validation)は行わず、学習データの内500日分のデータを利用して行い、繰り返し計算ごとに精度の推移を確認するためのパラメータとして利用する。

5. 結果の分析

各条件での貯水位予測の結果を図-1に示す。どの条件でも長期間貯水量が減り続けるような変化に関しては、大きな誤差なく予測できている。しかし、図-4のように貯水量が大きく変化しない期間の予測結果が安定していない地点が見受けられる。本研究の目標でもあった水位減少時の判定は過去3日間の貯水量が減少しているという条件で図-3のように行い、その地点での誤差と予

表-3 構築したNNモデルの概要

Train:Test		5000:3000
ニューロン数	入力層	基本変数18
	隠れ層	1000
	出力層	1
LSTM層の活性化関数		tanh
再起計算用活性化関数		tanh
出力層の活性化関数		sigmoid
最適化アルゴリズム		adam
予測一回分のデータ数		30
epoch(繰り返し計算数)		1000
ミニバッチサイズ		30
損失関数	Mean Abs Error	
metrics	MeanAbsError, MeanSquareError	

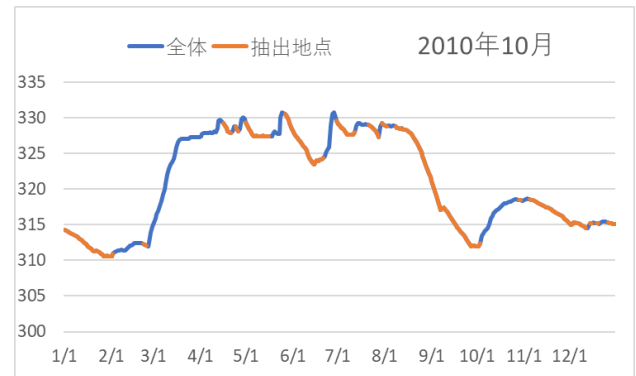


図-3 水位減少時と判定した地点

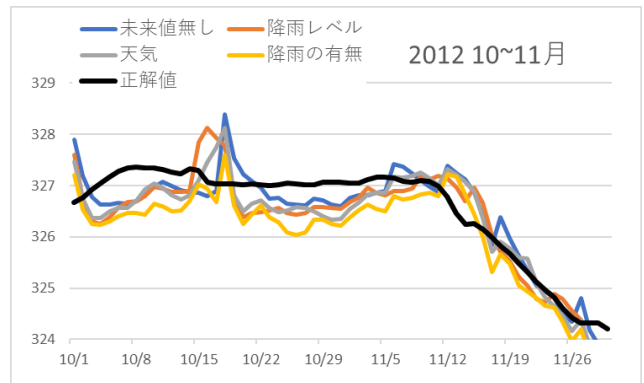


図-4 2012年10~11月の各値

測期間全体誤差は表-4のようになった。平均貯水量を基準とした貯水量誤差は有効貯水量との相対誤差で1.5%程度となっている。現在でも入力できる可能性の高い天気データのデータについては、観測地点が早明浦ダムから遠い点や数値化するにあたりファジィ的处理を挟んだせいか将来データを与えた中では最も低い精度となっている。しかし、降水の有無は早明浦ダムの集水域にある5地点のデータを与えているのに対して、天気は気象条件についての情報は多いものの観測地点が離れている点や高知・徳島の2地点しか入力していない状態であるにもかかわらず、どちらも近い結果を出したのは私の予想とは異なる結果となった。

また、本研究の主旨ではないが将来データを与えても図-5のように貯水量が大きく増える際の立ち上がりやその後の貯水量の増え方、貯水量のピークやそのタイミングを予測できているケースは少なく、洪水に関しては時間メッシュの細かいデータで行わねば将来データがあっても予測がうまくいかないことがわかる。

また、本研究で利用した、早明浦ダムの貯水量予測の結果を入力変数に加えた、池田ダムの貯水量予測も行った。結果は予測結果有りの貯水位誤差は0.046m、予測結果無しの貯水位誤差は0.054となった。下流側のダムの貯水量を予測する上でより上流側にあるダムの予測を入力データに加えることは有用であることが分かった。ただし、上流のダムが学習に使った期間は予測はすでに学習したデータから予測を行うため、精度の信用度が極めて低くなるため、予測用のデータから学習することとなる。今回でいうと、早明浦ダムの貯水量を予測するのに全8000日分のデータの内5000日分を学習に使っているため、池田ダムの貯水量予測モデルでは学習・予測を3000日分のデータで行う必要がある。

6. おわりに

本研究では、3日後の早明浦ダムを予測するNNを作成し、そのNNに将来のデータを入力することで、結果にどのような影響を与えるのかを調査し、天気という抽象的な観測値でも数値化することで精度向上が見込めることが分かった。

今後の課題としては、将来データの信用性を考慮する、将来データを現在の予報範囲・予報地点に則したものにすると、といったものがあげられる。

天気予報の結果はデータとして公開されていないが、予報結果や降水確率のデータを入手できれば信用度を考慮した結果を出せる。また、現在は将来データの降水レベルと降水の有無は観測所の値を基に作ってしまっているため、より天気予報の予測地点に則したデータにするべきである。

表-4 各条件での誤差

		水位(m)	貯水量(千m ³)
全体	未来値なし	0.89	4537.76
	天気	0.79	4028.68
	降水の有無	0.77	3925.24
	降水量レベル	0.60	3086.85
水位減少時	未来値なし	0.64	3249.86
	天気入力	0.58	2967.94
	降水の有無	0.58	2952.82
	降水量レベル	0.47	2424.45

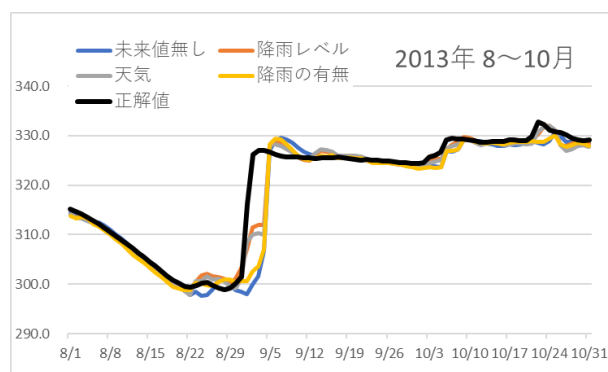


図-5 2013年8～10月の各値

謝辞

本研究では、国土交通省四国地方整備局 吉野川ダム統合管理事務所、水資源機構池田総合管理所 第一管理課、徳島県水管理政策課、より貴重なデータを使用させていただきました。ここに記して謝意を表します。

参考文献・出典

- 1) 土屋 大 藤村 和正:早明浦ダム流域における長期流出解析と渇水時の流入量予測, 第35回土木学会関東支部技術研究発表会, 2007
- 2) 国土交通省四国地方整備局
<http://www.skr.mlit.go.jp/yoshino/> 2022/02/18
- 3) 国土交通省水門水質データベース
<http://www.river.go.jp/> 2022/02/18
- 4) 国土交通省ダム諸量データベース
<http://mudam.nilim.go.jp/home> 2022/02/18
- 5) 我妻 幸長, はじめてのディープラーニング, SBクリエイティブ株式会社, 2018
- 6) 巢籠悠輔, 詳解ディープラーニング
TensorFlow/Keras・Pytorchによる時系列データ処理, 株式会社マイナビ出版, 2019
- 7) James Bergstra, Remi Bardenet, Yoshua Bengio and Balazs Kégl: Algorithms for Hyper-Parameter Optimization, NIPS 2011
- 8) Sepp Hochreiter and Jürgen Schmidhuber: Long short-term memory, Neural Computation 9 (8) 1997