河川・ダムの水位から学習する吉野川流域の渇水予測 AI の開発

中央大学　学生会員　○八木　将　中央大学　正会員　佐藤　尚次

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，降った雨が一気に海へ流れてしまう．そのため洪水が多いことはよく知られているが，それらの日本の河川の特徴は同時に渇水が起こりやすい原因となっている．

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ため軽視してよい問題ではない．

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，ダムがあるにもかかわらず災害を起こしてしまうことになる．実際，去年の 7 月初めに起こった愛媛県での洪水の際には降雨量が多くダムが耐え切れず放流し事故が発生した．さらに，2016 年の 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もある．

そこで，ダムのより良い利用のために NN(Neural Networks)を構築し渇水予想を行う．

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流域とした．

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，国内で数少ない慢性的に水不足である地域である．

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．

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．

表-1　構築したプログラムの概要

Train:Test		6500:2000
ニューロン数	入力層	8
	隠れ層	300
	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		adagrad
予測一回分のデータ数		60(日分)
エポック数		2000
損失関数		binary_crossentropy
metrics		acc,loss,recall,precision

表-2　metrics 計算用の変数

予測＼正解	○	×
○	TruePositive(TP)[回]	FalsePositive(FP)[回]
×	FalseNegative(FN)[回]	TrueNegative(TN)[回]

4. 研究手法

本研究では日毎の吉野川の河口部・下流域・中流域にある 3 地点の水位と本流の早明浦ダム・池田ダム、支流である銅山川の宮川内ダム・柳瀬ダム・新宮ダムの貯水量、計 8 変数を入力データとし 30 日後の取水制限の有無を予測する．構築する NN モデルは，時系列データを扱うことに長けている RNN(Recurrent Neural Networks)，そのなかでも利用例が多く比較的新しいネットワークモデルである LSTM(Long Short-Term Memory)を選んだ．NN モデルは Tensorflow をバックエンドとした Keras を利用し構築した．RNN の特性上，学習データは 3 次元データ(分割データ数, 予測 1 回分のデータ数，変数の数)に変換してインプットする．

本研究は，渇水という頻度が低い現象の予測が目的である．その為正解データは，取水制限有りの割合が 13%と偏りのあるデータとなっている．モデルの評価指標となる metrics として一般的に使われている Accuracy は，正解データに偏りのあるデータの予測に対して適切に評価できないため Recall・Precision も加えて出力した．

キーワード　深層学習　渇水　再起型ニューラルネットワーク　吉野川　TensorFlow
連絡先　〒112-8551　東京都文京区春日 1-13-27 Tel 03-3817-1816　Fax 03-3817-1803
E-mail: a16.dkrw@g.chuo-u.ac.jp

© Japan Society of Civil Engineers

Recall, Precision, Accuracy は，次のように定義した．

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

ここで，TP，FP，TN，FN は予測値と真値の関係を**表-2** のように分類した際に，学習中の出力がそれぞれどこに何回分類されたかを表すものである．

Precision は取水制限があると予測した際の精度，Recall は渇水が実際に起こった際の予測の精度をそれぞれ表すものになっている．

5. 結果の分析

約 24 年分の日毎のデータ 8500 個のうち，Train データ 6500，Test and Validation データ 2000 でそれぞれの metrics を epoch 毎に計算し，その epoch 毎の推移を**図-1** のように出力した．Recall，Precision のグラフには TensorFlow の定義上 Test の曲線は出力されていない．

Loss，Accuracy の傾向を見る限り，過学習の傾向は無い．Accuracy に目立った変動がないことからも Precision，Recall を導入した妥当性がわかる．

しかし，Recall の推移は上昇傾向にあるものの精度自体は 5 割程度とよくないので改善が必要である．Precision に関しても精度が 8 割程度で滞っているため更なる改善が必要である．

6. おわりに

今後は，新たな入力変数の追加，入力データの範囲拡張，ネットワークモデルの最適化を行い精度の向上を行う．

現在は 30 日後の予測を行うにとどまっているので精度を保持したまま予測期間の延長も行う．

さらに，各水道局が行う取水制限は 7 段階のレベルに分けて行っているため，それに合わせての現在の 2 分類問題から分類分けを詳細化し 8 分類問題の予測を行う．

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/ 2020/03/29
3) 国土交通省水門水質データベース http://www1.river.go.jp/ 2020/03/29
4) 国土交通省ダム諸量データベース http://mudam.nilim.go.jp/home 2020/03/29
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) 巣籠悠輔，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) Tensorflow https://www.tensorflow.org/ 2020/03/29

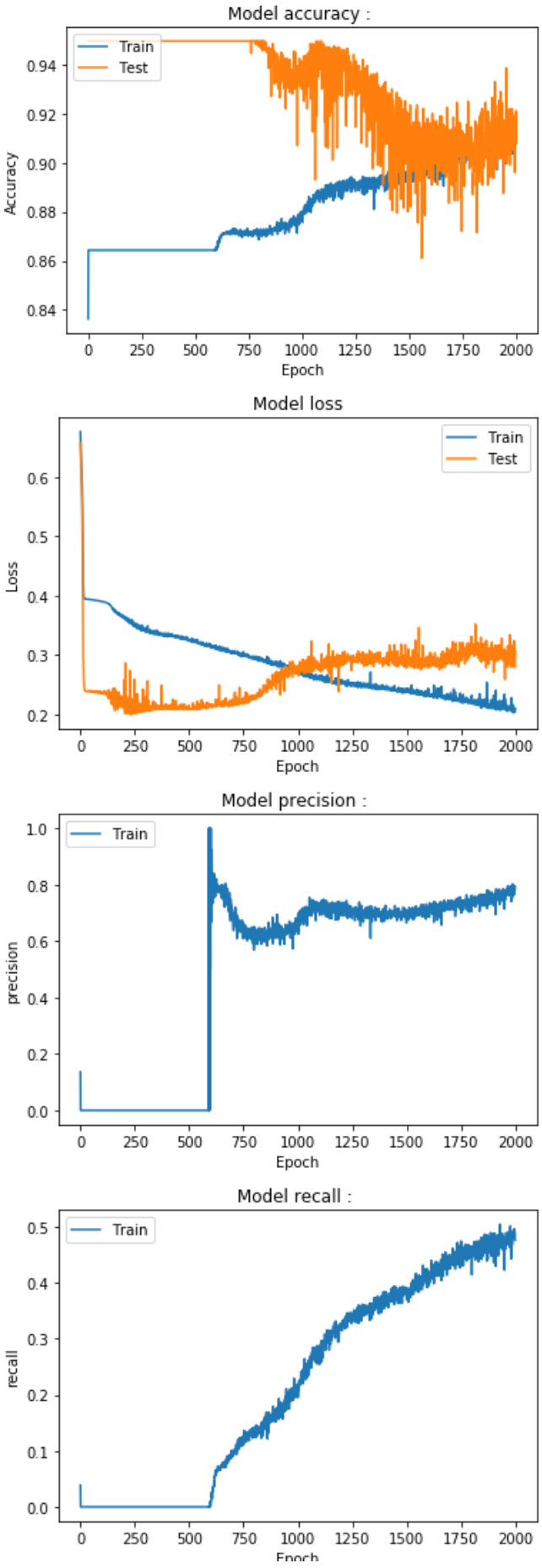

図-1 Accuracy, Loss, Recall, Precision の推移

© Japan Society of Civil Engineers