流域雨量を入力データとした機械学習モデルによる洪水時の水位予測

高知高専　特別会員　○今井 裕也
高知高専　正会員　岡田 将治
高知高専　非会員　木村 竜士

1. 研究の背景

近年、防災の観点から豪雨時の被害予想が必要とされている。高知でも記録的な豪雨となった平成 26 年 8 月豪雨は高知県 高岡郡 四万十町でも大きな被害となった。よって災害時の早急な対応には上流での雨量の値を用いた予測モデルの構築が必要と考える。

2. 研究の目的

本研究では、2015 年から 2019 年現在において上流で観測される雨量と下流である津野川での水位の相関を LSTM(Long short-term memory)を利用した機械学習モデルに学習させ、機械学習モデルにとって未知のケースとなる 2014 年の豪雨に対して予測が行えているかを検証及び評価する。ここで学習された機械学習モデルは最終的に国土交通省 四国地方整備局の Web サイト上で公開され異常な豪雨時に下流での被害をいち早く推定するツールとして活用されることを目標としている。

3. 学習で利用するデータ構造

本研究では時系列データである雨量及び水位を取り扱うことに際し、機械学習で用いるモデルに LSTM を利用することとした。LSTM モデルは再帰型ニューラルネットワークである RNN(Recurrent Neural Network)の一種である。RNN では時系列で入力されるデータに対して隠れ層での出力を次の時間での隠れ層に対しても出力するという特徴を持っている。RNN では、セル出力は tanh 関数によって-1～1 の値に計算され次のセルに与えられる。RNN モデルによって入力データを時系列に扱うことは可能であるがセル内部の構造から長期的な要素の記憶が困難ということが言われており数日前までの雨量が下流の水位に影響することが予想される水位予測では精度が悪い可能性がある。よって本研究では RNN モデルのセル内をより複雑化したもので前セルからの入力の強度を保持するセルステートを持つことによって情報の再利用、保持、忘却を行う LSTM モデルを利用する。LSTM モデルの利用によって数時間前の雨量がのちの流での水位に影響する相関を表現できることを期待する。

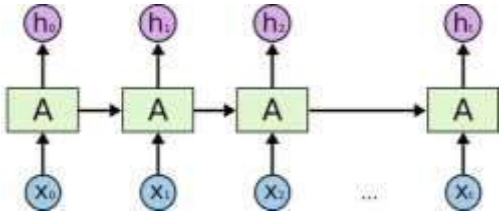

図 1 RNN モデルの構造

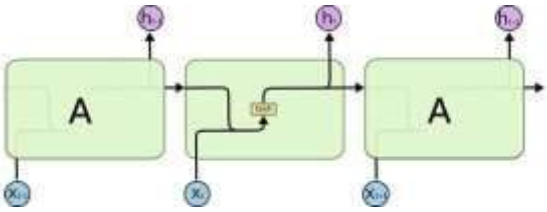

図 2 RNN セル内の構造

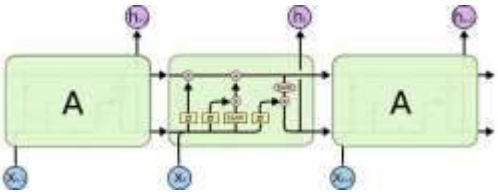

図 3 LSTM モデルの構造

4. 学習で利用するデータ構造

本研究ではサイト上からダウンロードした上流の目黒、近永、大正、窪川、大野見、船戸での雨量と下流となる津野川での水位の 2014 年から 2019 年までの 1 時間ごとのデータを用いる。モデルの評価に際して 2014 年 8 月 10 日に発生した豪雨による下流での水位上昇を予測したいため 2015 年から 2019 年までを学

習のパラメータの最適化に用いる訓練データ,2014 年のデータをモデル精度の評価のみに利用するテストデータとする。

5. データの整形及び正規化

本研究で LSTM モデルに与える予測に用いるデータ構造を[データセット数、要素数、ルックバック数]として表 1 に示す。ここでは時間 t からルックバック数までさかのぼったデータを用いて t+1 における水位を予測する。また、その際 t+1 における水位の予測にそれまでの水位をルックバック数分用いることでそれまでの水位の推移も考慮する。また、入力値及び出力値は最小値を 0、最大値を 1 する MinMax normalization にて正規化を行った。

6. LSTM モデルの構造

学習モデルである LSTM モデルは python で利用可能な機械学習のフレームワークである tensorflow を backend として実行する keras フレームワークを利用する。LSTM モデルの構造として要素数とルックバック数を指定した LSTM を 1 層、次に出力に際し 1 つのノードにした全結合ニューラルネットを介し予測される水位の値を出力する。モデルのハイパーパラメータについて表 2 に示す。

7. 結果

LSTM モデルの学習によるエポックごとの損失の推移を図 4 に示す。学習結果としてある時刻 t における予測値を Y 軸、と実測値を X 軸としたグラフを訓練ケースとテストケースで分け図 5 に示す。縦軸と横軸の長さ比を同じにしているため X=Y の線に近いほど良い精度が得られたといえる。また、予測値と実測値を時系列に表したグラフを図 6 に示す。テストデータにおける赤線枠の部分が今回の学習で予測したい 2014 年 8 月 10 日における水位である。また、モデルの精度の評価として平均二乗誤差（RMSE）を表 3 示す。

8. 考察

モデルのエポックと損失についての図 4 によると 100 エポックまでは順調に損失が降下したがそれ以降は損失がほとんど減っていないため今後の学習では 500 エポックまでの学習で概ね良いと言える。また訓練データとテストデータにおいて RMSE の差が大きかったため予測したい水位の大きなデータに対して予測が最適でないことがわかる。

9. まとめ

本研究では上流での 5 点での雨量と以前の水位を用いて 1 時間後の水位を推計するモデルの構築、学習を行った。現状では洪水時の精度が悪いためデータ構造、モデルのハイパーパラメータの変更が必要と考える。

10. 参考文献

1）中根 英昭、環境分野への深層学習応用研究の立ち上げについて、2018.05.07

2) Antonio Gulli, Sujit Pal、直感 Deep Learning Python×Keras、O'Reilly Japan, Inc、2018.8.

表 1 モデルに与えるデータ構造

	データセット数	要素数	ルックバック数
訓練	35458	7	72
テスト	8687	7	72

表 2 ハイパーパラメータ

誤差評価法	最適化関数	エポック数
平均二乗誤差	Adam	1000

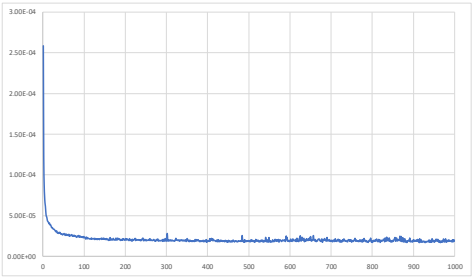

図 4 epoch 数-損失グラフ

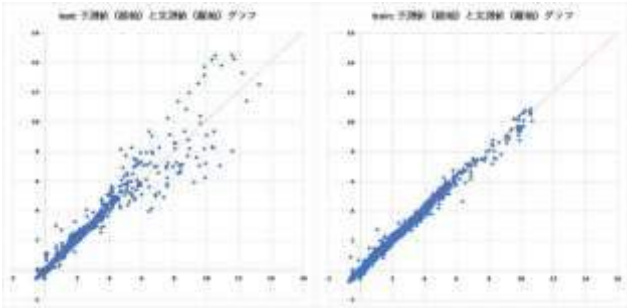

図 5 予測値-実測値グラフ

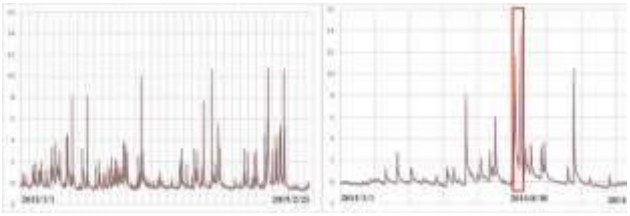

図 6 時間-水位グラフ

表 3 RMSE 値

	RMSE
訓練	0.0671
テスト	0.229