

深層学習モデルの損失関数と学習データの与え方による水位予測精度向上の検討

An Examination for accuracy improvement of water level prediction focusing on Setting of Error function and training data in Deep Learning

○福重 雄大*・皆川 裕樹*・木村 延明*

○FUKUSHIGE Yudai・MINAKAWA Hiroki・KIMURA Nobuaki

1. 背景と目的

比較的大きな農業地区の下流低平部では、用水確保と内水（雨水）排除という、相反する目的を満たすよう排水システムが整備され、緻密な運用がなされている場合がある。このような地域においても、排水対策が進み都市化すると、農業水利用に加えて洪水安全度を満たす施設操作が要求され、管理者の負担は増加すると考えられる。その負担軽減にむけて、筆者らは、深層学習モデルを活用した水位予測システムの開発に取り組んできた（皆川ら、2020）。ここでは、系列データの予測を得意とする LSTM（Long Short-Term Memory）モデルを適用している。このシステムを後述の対象地区に適用すると、全体的には十分な精度で水位を予測できることが示された。しかし、その結果を詳細に見ると、大きな雨量が発生した際の水位のピーク立ち上がりの部分で、予測の時間遅れが確認された（Fig.1）。本システムの目的は、管理者の意思決定を早期化し操作にかかる時間的余裕を生む点にあるため、予測時間の遅れはその目的に反する課題である。この要因として、深層学習モデルの構造や学習データの精度、時間間隔等、様々なものが考えられる。そこで本研究では、これらの要因の中からまず学習時の損失関数及び学習データの与え方に注目し、それらが予測結果に与える影響を検討したので報告する。

2. モデル概要と検討方法

(1)対象地区 本研究では、新潟県新潟市亀田郷地区を対象とした。本地区では降雨時の流出水は最下流部にある鳥屋野瀉に集められ、そこに設置された親松排水機場（最大排水能力 60 m³/s）によって信濃川に機械排水される。そのため水位予測の対象は親松排水機場地点とした。

(2)モデル概要 本研究で用いたモデルは TensorFlow で実装しており、6 時間前までのデータを用いて 1 時間先の水位を予測する。通常、深層学習モデルのパラメータ最適化では大量の学習データが必要となるが、豪雨時を対象とする場合は現地において十分な観測データを用意できない。そこで本研究では、降雨パターンの異なる模擬豪雨（皆川ら、2014）と、それを排水解析モデル（吉川ら、2011）に入力した計算水位を疑似の観測データとして 1000 セット作成し、学習に用いた。実水位へ適用した場合の LSTM モデルの検証には、2010-2018 年に対象地区で観測された 263 個の降雨イベントと観測水位を用いた。観測イベント中の最大降雨イベント時の予測結果である Fig.1 を見ると、全体としてよく水位を予測できているが、前述のように立ち上がりに遅れがみられた。以下に示す

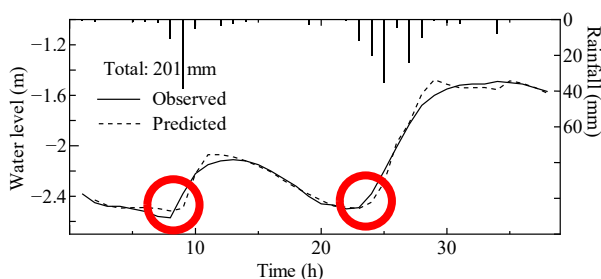


図 1 LSTM を用いた予測結果¹、赤丸が遅れ部
Fig. 1. Prediction results using LSTM

* (国研) 農研機構 農村工学研究部門 Institute for Rural Engineering, NARO
キーワード：深層学習，水位予測

検討は、この遅れの解消を目指したものである。

(3) 検討項目 表 1 に示すように、まず損失関数による結果への影響を確認した (Case 2) 後に、学習データとなる降雨量の与え方を工夫した (Case 3, 4)。損失関数は、TensorFlow 内に実装されており、且つ系列データに用いられるものを対象とした。Case 3 で与えた累積雨量は、地域の湿潤状況等が降雨流出へ与える影響が考慮されることを期待して加えた。Case 4 で与えた 1 時間先降雨量は予報値をイメージし、事前に降雨を察知することで水位上昇の刺激をモデルに与えることを期待した。学習の度に結果に僅かな違いが生じるため、各ケースにつき学習を 10 回実施しその平均値を比較に用いた。結果の評価対象とするイベントは Fig.1 と同じとした。

3. 結果と考察

実測値全体に対する各ケースの平均 RMSE を表 2 に示す。対照実験である Case 1 (Fig.1 と同じ結果) の RMSE が最も小さくなっており、今回検討した損失関数は、モデルの全体的な精度向上には寄与しない結果となった。二乗和誤差は、誤差が大きいほど過大に評価しパラメータを大きく更新するので、より良いパラメータを獲得した可能性がある。また、Case1 の入力データに累積雨量や 1 時間先降雨量を加えても、結果は改善しなかった (Case 3, 4)。さらに、本研究で注目した水位の立ち上がり部を見ても (Fig.2)、全ケースで遅れ現象は改善されなかった。全検討ケースの予測の挙動が同じことから、この遅れ現象はモデル又はデータセットに特有の問題である可能性がある。例えば、降雨量と比べて水位の方がモデルの出力に与える影響が大きい可能性が考えられる。今後は、学習に用いたモデルとデータの精査と共に、実運用を考慮しながら、課題の解決に向けて例えば数時間先の雨量を与える、または 1 時間よりもさらに短い時間間隔 (10 分毎など) のデータを学習/予測に用いる試験や、そもそもの深層学習モデル構造の再検討が必要と思われる。

4. まとめ

深層学習モデルを用いた水位予測における後追い問題の解消のため、損失関数や入力データについて検討した。本検討では改善が見られず、本問題の解決には別のアプローチが必要と考えられた。農村地域では限られたマンパワーで施設管理を担っている現状からも、本システムのような支援を目的とした技術開発と、現場への適応を進めることが重要と考えられる。

参考文献: 1) 皆川裕樹ら. 土木学会論文集 B1 (水工学), 65, pp.1_349-354, 2020. 2) 皆川裕樹ら. 農業農村工学会論文集, 291, pp.15-24, 2014. 3) 吉川夏樹ら. 土木学会論文集 B1 (水工学), 67(4), pp.1_991-1_996, 2011.

表 1 検討項目の一覧

Table 1. Items to be considered

| ケース名 | 検討項目 | 損失関数 | 入力変数 (入力特徴量) |
|---------|-----------|---------------------------|--------------|
| Case1 | 対照実験 | 二乗和誤差 | 水位+降雨量 |
| Case2.1 | 損失関数 | 平均二乗誤差(mse) | 同上 |
| Case2.2 | 同上 | 平均絶対誤差(mae) | 同上 |
| Case2.3 | 同上 | 平均絶対パーセント誤差(mape) | 同上 |
| Case2.4 | 同上 | 対数平均二乗誤差(msle) | 同上 |
| Case2.5 | 同上 | コサイン類似度(cosine_proximity) | 同上 |
| Case3 | 累積雨量追加 | 二乗和誤差 | 水位+降雨量+累積雨量 |
| Case4 | 1時刻先降雨量追加 | 同上 | 水位+1時刻先降雨量 |

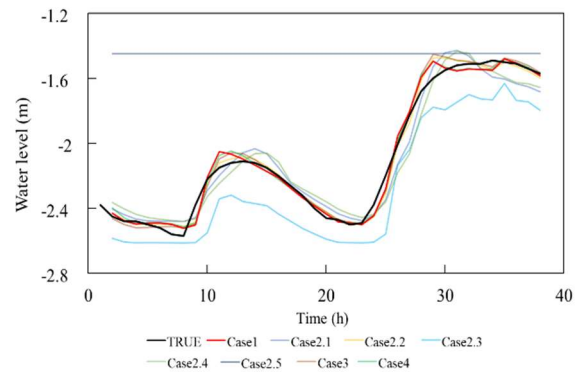


図 2 各ケースでの予測結果 (10 回の平均)

Fig.2. Prediction results for each case

表 2 各ケースの平均 RMSE
Table 2. Average RMSE for each case

| 試行 | 10回平均 RMSE |
|---------|------------|
| Case1 | 0.037 |
| Case2.1 | 0.088 |
| Case2.2 | 0.047 |
| Case2.3 | 0.342 |
| Case2.4 | 0.093 |
| Case2.5 | 1.065 |
| Case3 | 0.045 |
| Case4 | 0.043 |