

転移学習を用いた河川洪水イベントの事前学習モデルの構築と検証 Assessment of an AI pretrained model created for flood prediction using transfer learning

○木村延明*・皆川裕樹*・福重雄大*・馬場大地**
○Nobuaki KIMURA・Hiroki MINAGAWA・Yudai FUKUSHIGE・Daichi BABA

1. はじめに

近年、地球温暖化の影響に起因する大規模洪水の発生リスクが懸念されている。河川氾濫等の逐次対策や早めの避難誘導のために、迅速かつ正確なリアルタイム予測の構築が求められている。迅速性に優れたニューラルネットワーク(ANN)は、リアルタイム洪水予測に最適であるものの、洪水イベントのような学習のためのデータ量が非常に少ない極端事象や、そもそも学習データに含まれない未経験事象の場合には予測精度が低下する。この課題を解決するために、転移学習を用いることで予測精度の向上を図る手法を提案する。なお、転移学習とは、基ドメインの特徴を対象ドメインに転移させる学習方法である。ANNによる洪水予測に転移学習を実装した既往研究¹⁾では、地形条件などが類似した2つの流域を対象に、大量のデータを有する基ドメインで事前学習されたANNを、対象ドメインで再学習することで、少ないデータでも期間最大の洪水波形の予測精度の向上を図った。しかし、類似する2つの流域を選定することは困難であり、また、洪水イベントのデータが大量にある流域も稀なので、水文情報(気象や地形など)が類似するのであれば、複数の流域の観測データを可能な限りかき集めて基ドメインで大量データを生成することは可能である。本研究では、基ドメインで収集されるデータに基づく網羅的な河川洪水の水位変化のパターン(特微量)を学習した事前学習モデルを構築し、その事前学習モデルを、対象ドメインの特定の洪水イベントに適用して水位予測精度の改善を図る。

2. 方法

水文情報が比較的類似すると考えられる九州地方の主要河川である11水系を対象に、水文水質データベース(国交省)から収集される洪水イベントの観測値(最長期間:1979年~2021年)を利用する。本研究では、観測水位が氾濫危険水位を超えた期間を含む最大ピークの前後120時間を洪水イベントとする。なお、ダムによる洪水調整などの影響を排除するために、上流にダムがない水位観測地点24箇所を選定した(全洪水イベント数=113個)。降雨量は近隣あるいは上流の雨量観測地点のデータである。本手法について、次の3つのステップで実施される(図1)。STEP1:対象ドメインを除く全ての観測地点のデータをかき集めて、基ドメインとする。STEP2:基ドメインのデータをANNの学習データにして、事前学習済みのANNを構築する(以下、「AI事前学習モデル」という)。STEP3:対象ドメインの期間最大の洪水イベント(TOP1)を除く観測データで、AI事前学習モデル

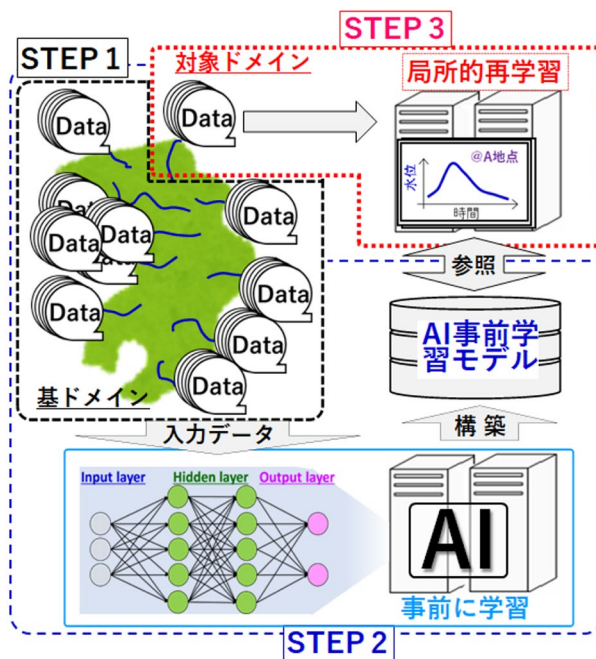


図1 新規手法の手順(STEP1~3)

Procedures of a proposed method: Step 1 to 3.

対象ドメインの期間最大の洪水イベント(TOP1)を除く観測データで、AI事前学習モデル

* 農研機構-農村工学研究部門 The Institute for Rural Engineering, NARO

**株式会社アーク情報システム ARK Information Systems, INC.

キーワード 河川洪水予測, ANN, 事前学習モデル, 少ないデータ

を再学習させて、TOP1 を予測する。なお、STEP 2, 3 が転移学習である。

本研究で提案された AI 事前学習モデルを利用する ANN (以下、「新 ANN」という) は、時系列データの予測に有用な長・短期予測 (LSTM) と全結合層を組み合わせたものである。全結合層のみを再学習 (0~50 回) することで、対象ドメインの観測データに最適化する。なお、十分なデータ (洪水イベント 10 個以上) が収集された観測地点で最適なハイパーパラメータ (エポック数=100, バッチサイズ=100, ノード数=77, 学習率=0.001, オプティマイザー=確率的勾配降下法) を調整した。入力データは、雨量・水位データの過去 6 h~現在 (1 h 間隔) までの 7 個を使い、出力は 1 h 間隔で現時点から 1 h~6 h 先の水位予測値である。leave one out 交差検証法を用いて予測精度の評価を行い、二乗平均平方根誤差 (RMSE) と Nash-Sutcliffe 係数 (NSE) で定量評価を行う。なお、比較用として、AI 事前学習モデルを利用しない ANN (以下、「旧 ANN」という) を用いる。

3. 予測結果

2 つの水位予測モデル新・旧 ANN の予測結果について、対象ドメインをデータ数が比較的多い地点、少ない地点を、それぞれ栗野橋観測所 (17 個の洪水イベント, 川内川中流域) 伊田観測所 (3 個の洪水イベント, 遠賀川中流域) とした。それぞれの対象地点では、それ以外の観測地点で収集された洪水イベントを全て用いて AI 事前学習モデルを構築した。栗野橋観測所では、17 個のイベントの内、TOP1 を除くイベントで局所的再学習を行い、その上でリードタイム 1, 3, 6 h の TOP1 を予測した。新 ANN では、再学習 10 回以上で予測結果は収束し、リードタイム 6 h の大小ピークの部分で若干の差異は見られるものの、NSE は高々 3% の改善のみで、新・旧 ANN の結果に顕著な違いはなかった (図 2 左)。一方、伊田観測所の新 ANN の予測結果では、旧 ANN に比べ RMSE は 10~70% の改善が見られた。特に、再学習回数 0 回でも大幅な改善が見られた (図 2 右)。以上から、少ない観測データの場合には、新 ANN による予測結果の方が旧 ANN の結果に比べ、相対的に良好な水位予測が可能なが示された。しかし、両観測所ともにリードタイム 6 h では、観測値の再現性が悪化し、今後の改善が求められる。

謝辞：本研究は、JSPS 科研費 JP 21H02310, JP 21K05838 と (公財) 三菱財団自然科学研究助成により実施された。

参考文献
1) 木村ら：転移学習とカップリングした LSTM による洪水予測モデルの開発, 河川技術論文集, Vol.26, pp.187-192, 2020.

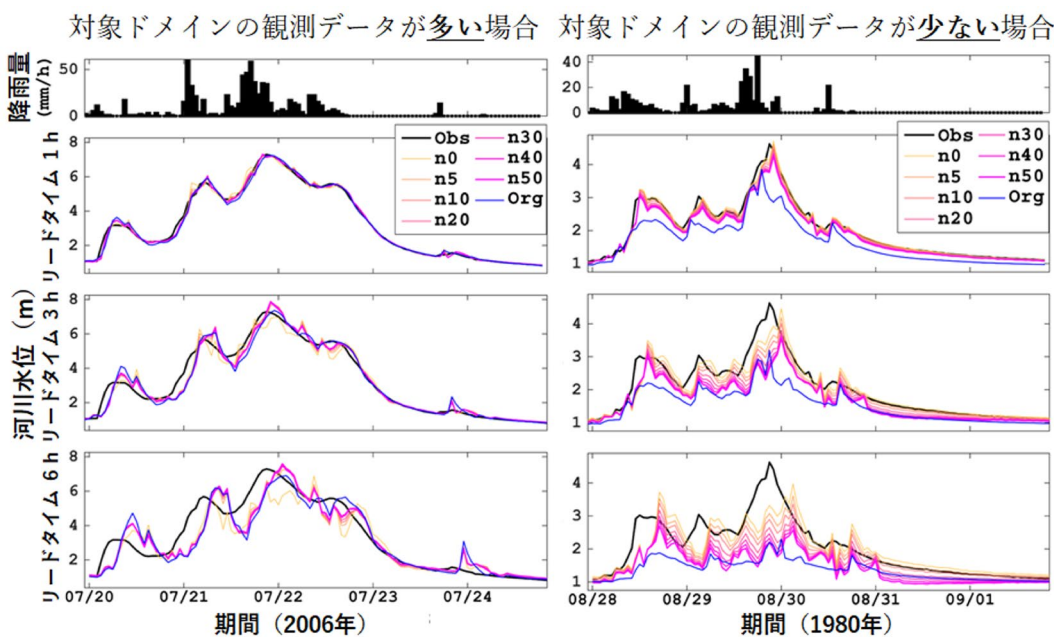


図 2 降雨量と新・旧 ANN のリードタイム 1 h, 3 h, 6 h の水位予測結果と観測値の比較, 黒線=観測値, 青線=旧 ANN, 黄色~ピンク=新 ANN, n=再学習回数。
Rainfall and old/new ANN-based predicted water levels in 1 to 6 h lead times vs. observations (left panel: lots of data implementation, right panel: few data being available).