河川技術論文集，第26巻, 2020年6月

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

DEVELOPMENT OF A FLOOD PREDICTION MODEL BY LSTM WITH A TRANSFER LEARNING APPROACH

木村延明[1]・吉永育生[2]・関島建志[2]・安瀬地一作[3]・馬場大地[4]

Nobuaki KIMURA, Ikuo YOSHINAGA, Kenji SEKIJIMA, Issaku AZECHI and Daichi BABA

[1]正会員　PhD　農研機構　農村工学研究部門（〒305-8609 茨城県つくば市観音台 2-1-6）

[2]非会員　農博　農研機構　農村工学研究部門（〒305-8609 茨城県つくば市観音台 2-1-6）

[3]正会員　農博　農研機構　農村工学研究部門（〒305-8609 茨城県つくば市観音台 2-1-6）

[4]非会員 ㈱アーク情報システム アドバンストテクノロジー部（〒102-0076東京都千代田区五番町4-2）

A real time forecasting system for flood events requires a robust, accurate and quick response model that is able to predict water levels. We focus on a deep neural network model embedding a long, short-term memory architecture (LSTM), which utilizes only observed data without any hydrological-based setups. However, lots of data are necessary to develop an accurate LSTM model. To achieve this requirement even in insufficient data, an alternative method is proposed in this study using transfer learning (TL) coupled with the LSTM model. For TL, the model is pre-trained in a source watershed with lots of flood-event data, then being directly applied to a target watershed with small-size data. The results for the top three largest flood events in the target watershed show that predicted water levels with TL were improved by 30% of those without TL. Additionally, the TL-based model with 10-time re-trainings predicted more accurately by 2% if compared with the model without re-training. The computational cast was reduced by up to 1/10.

Key Words : *LSTM, transfer learning, flood forecast, time-series data, insufficient dataset*

1．背景と目的

リアルタイムの洪水予測には, 頑健且つ高精度で, さらに短時間応答のモデル開発が求められている[1]. 予測時間が非常に短く，また，流出解析モデルのように物理パラメータなどの調整をする必要のないデータ駆動型のニューラルネットワークモデルが，洪水予測モデルとして開発されてきた[2]. 近年，深層ニューラルネットワークが急速に発達してきたお陰で, さらに予測精度の高いニューラルネットワークモデルの構築が可能になっている[3]. 特に，時系列データの予測を行う場合に，過去のトレンドのような特徴量を効率的に抽出する機能に優れ，さらにそのトレンドを記憶できる機能を備えた長・短期記憶（Long Short-Term Memory, LSTM）アーキテクチャ[4]を実装した深層ニューラルネットワークのモデル（以下，「LSTMモデル」という）は，ここ数年来，国内外で広く洪水予測に利用されている[5),6)]．しかし，予測精度の向上のために大量なデータによるモデル内部のパラメータなどの調整（以下，「学習」という）が必要なことに加え，学習のために計算コストがかかることが主な課題である. これらを解決するために，不十分なデータを有する対象流域（ターゲット流域）とは別の流域（ソース流域）で大量なデータで事前学習させたモデルを再利用する手法である転移学習[7)] を用いる（**図-1**のデータフローを参照）.

転移学習の導入は，画像分類に優れた畳み込みニューラネットワーク（Neural Network Convolution, CNN）[8)]とのカップリングで実施された画像解析に多くの事例が見られるものの，時系列データの予測に利用されるLSTMモデルとの適用事例は未だに少なく，規則的な変動を伴う信号解析などに見られるのみである[9)]. この理由について, 時系列データでは, 現時点の値から次の時間ステップの値を予測する場合に，時間変化のパターンが無数に存在し，ソースとターゲットのデータ間の類似性を有する特徴を見出すことが困難だからである.

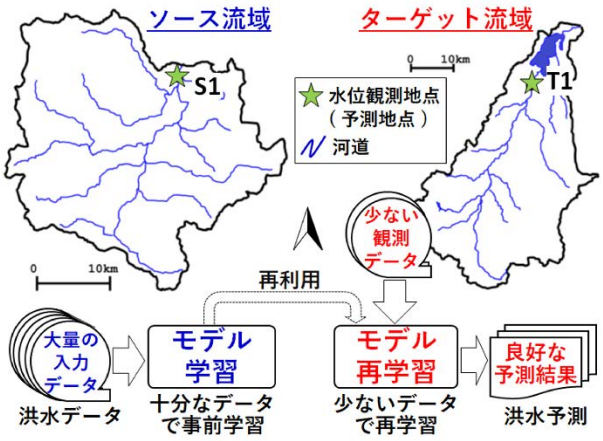

図-1 対象流域のマップと転移学習のデータフロー

なお，画像分類において，転移学習を導入したCNNが有用な理由は，類似性のある画像タイプ，例えば猫と犬のような動物画像の間で転移学習を行う場合には，共通の特徴量（例：目が2つ，足が4本など）を抽出しやすいからと考えられる．

時間変化を伴う洪水予測を行うために構築されたLSTMモデルでは，2つの流域間で転移学習を適用した事例は過去に見られない．そこで，本研究では，洪水予測を行うLSTMモデルに，転移学習を導入してデータ量が少ない流域でも良好な予測ができ，またモデルを再利用することで計算コストも削減できる洪水予測手法を開発する．

2．方法

(1) 流域とデータ

人為的な操作で河川水位が変化するような治水ダムなどの大規模構造物を有しない2つの流域（ソース流域とターゲット流域）をそれぞれ選定する（**図-1**）．ソース流域は，大淀川水系の中・上流域を含み，861 km² の集水面積をもち，標高範囲は118–1,350 mである．幹川流路の距離は，水位計が設置された地点（観測地点，**図-1**）まで約52 kmである．ターゲット流域は，網走川水系で，1,380 km² の集水面積をもち，標高範囲は0–978 mである．幹川流路の距離は，予測地点まで約100 kmである．なお，観測地点は，本研究で用いたモデルの予測地点でもある．入力データとして使用する洪水イベントについて，河川水位のみを変数とし，ソース流域では過去30年間のうち4月～10月までに発生した洪水イベントのデータ（43イベント）と，ターゲット流域では過去20年間のうち比較的データ量の少ない（17イベント）を利用する．これらのイベントの抽出は，ソース流域で，データ量を増やすためにピーク水位が4.5m以上とし，期間最大のピーク水位の45%程度の大きさに対応する．

一方，ターゲット流域では，氾濫注意水位（3.2 m）以上のピーク水位が起こった場合（期間最大ピークの約55%に対応）を洪水イベントとして抽出する．なお，水位のみを入力データとして用いた理由は，山田ら[5]の既往研究を参考にした．1イベントあたりのデータについて，一言ら[3]と同様に，1時間毎の観測値で，洪水ピークの前72時間からピーク後の48時間までの合計，約5日間分の連続観測データを有する．データは，国土交通省が管理する水文・水質データベースより取得する．データの詳細は，Kimura et al.[10] の先行研究を参照されたい．

(2) LSTMモデル

本モデルは，連続性を有する時系列データなどの予測に特化したアーキテクチャを内蔵し，長期的なトレンドを記憶しながら学習を行うことができるものである．このモデルの内部構造は，現在時刻tの入力変数（x_t）に1ステップ前の時刻$t-1$の出力変数（h_{t-1}）を追加した後で，2つの活性化関数（σ：シグモイド関数，tanh：双曲線正接関数）と過去情報を記憶・忘却する機能を利用し，特徴的な情報を抽出して出力変数（h_t）を算出する（**図-2**）．LSTMモデルの内部構造は，以下の複数の式のように3つのゲートをもつ．なお，式中の$w[h_{t-1}, x_t]$ は，1ステップ前の出力変数 h_{t-1}と現在の入力変数 x_tに関する行列演算，b バイアスを意味する．忘却ゲートはシグモイド関数で制御され，次式で与えられる．

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right). \tag{1}$$

入力ゲートは，次式のように２つの活性化関数を用いて，入力データを制御する．なお，双曲線正接関数はシグモイド関数と比べ勾配の消失をより弱める働きをもつために導入される．

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i), \tag{2}$$

$$z_t = \tanh(w_z[h_{t-1}, x_t] + b_z). \tag{3}$$

忘却ゲートを含めた上記 3 つの入力データはブレンドされ，次式のように与えられる．但し，時刻 $t-1$で計算された記憶情報 C について一時保存されたメモリーセルから，過去の情報を取り入れるためにC_{t-1}が導入されている．

$$C_t = f_t \otimes C_{t-1} \oplus i_t \otimes z_t, \tag{4}$$

ここで，⊗= 積と⊕= 和の演算子である，f_tとi_t は，過去情報のC_{t-1} と入力情報の z_tに対して，それぞれの情報をどの程度の割合で取り入れるかを調整する．時刻 $t-1$の出力変数 h_{t-1}を，現在ステップの入力に対して，フィードバックする出力ゲートは次式で与えられる．

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o). \tag{5}$$

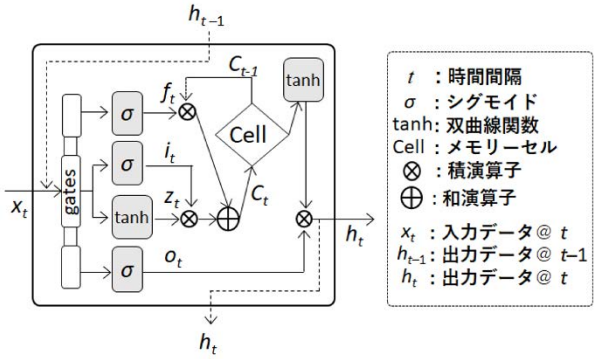

図-2 LSTM アーキテクチャの構造

表-1 ハイパーパラメータの条件

項目	値
バッチサイズ	100
エポック数	100
損失関数	E（ 式(9) ）
オプティマイザー	確率的勾配降下法
学習率	0.01
LSTM 層（ノード数）	1 層（20）

最終的にメモリーセルからの出力値と出力ゲートからの値との積によって，アップデイトされる出力変数は次式で与えられる．

$$h_t = o_t \otimes \tanh(C_t). \tag{6}$$

本モデルは，Python[11]を用いてプログラミングされ，また，機械学習のためのライブラリなどは，Keras[12]を利用する．最適な学習を行うために，LSTM モデルのハイパーパラメータなどを調整し，最終的に採用されたハイパーパラメータは**表-1**に示される．なお，本モデルの構築には，巣籠[13]のテキストを参考にした．

(3) 転移学習

転移学習とは，一般的に多数のデータ量を有するドメインで事前に学習させたモデルを，少ないデータ量を有するドメインで再利用する方法のことである．まず，転移学習無しの場合の一般的な機械学習について説明する（**図-3**）．例えば，2つの独立したドメインAとBを用意する．データ量を多く有するドメインAで，モデルAを用いて大量のデータで学習すれば，高精度な予測結果が得られる．一方，データ量が少ないドメインBで学習されたモデルBは，データのもつ特徴量を十分に抽出できないので精度の低い予測結果となる（**図-3a**）．ここで，ドメインAで構築された高精度な予測ができるモデルを，ドメインBでも再利用し，本来低精度な予測となるモデルBの予測精度を改善することを転移学習と言う．このとき，モデルA（ソースモデル）で抽出された特徴量（つまり，モデル内部の最適化されたパラメータの

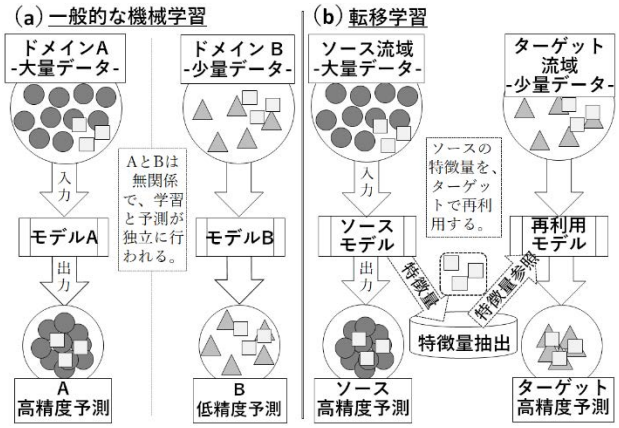

図-3 転移学習の概念図

こと）を取り出して，モデルB（再利用モデル）に組み込むことで転移学習が可能となる（**図-3b**）．

(4) 予測精度の検証と転移学習の導入

検証方法は，洪水イベントの水位データを用いて，学習・バリデーション・テストのプロセスを通して行われる[14]．モデル内のパラメータを調整するために，流域の 1～3 番目までの大きな洪水イベントと 2 つの中規模の洪水イベントを除いた残りのイベントに対して学習を行う．学習結果の妥当性を確認するために，過去 2, 3 番目に大きな洪水イベントと 2 つの中規模イベントについてバリデーションを実施する．モデルの汎用性の確認は過去最大の洪水イベントを用いてテストを行う．LSTM モデルに転移学習を導入した効果を評価するために，ソース流域で事前学習したモデルを，そのままターゲット流域に適用してテストを行う．これを，Case 1 とする．Case 2 は，過去 2 番目の洪水イベントについて，テストを行い，過去 1, 3 番目と中規模の 2 つの計 4 つのイベントをバリデーションに使い，残りのイベントを学習に利用する．Case 3 と Case 4 は，過去 3 番目と 1 つの中規模の洪水イベントについて，それぞれテストを行い，残りのイベント群はバリデーションと学習に振り分けられる．なお，中規模の洪水イベントの 1 つは，共通イベントとしてバリデーションのみに使われる（**表-2**，**表-3**）．

転移学習を行う場合に，ソース流域でテストされた 4 つのケースのモデルについて，それぞれを直接にターゲット流域の 4 つのケースに適用して，テストを行う（再学習無し）．さらにモデルの予測精度を向上させるために，モデルの一部を再学習してテストを行う（再学習有り）．この方法は，fine tuning[15]とも言われる．**図-4a** に示されるようにモデルのネットワーク構造は，LSTM 層と全結合層が 1 つずつである．ソース流域の学習と予測では，この構造のモデルを用いる．転移学習では，ターゲット流域のデータを用いて再学習無しの場合のネットワーク構造

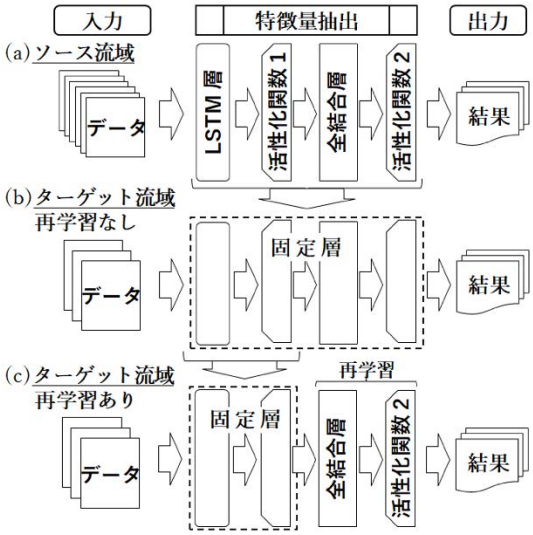

図-4 LSTM モデルのネットワーク構造(a) と転移学習の導入方法：再学習無し(b)，再学習有り(c).

表-2　テストのケース一覧（ソース流域）

ケース名	洪水イベントの規模	ピーク水位（m）	ピーク水位発生日時
Case 1	期間最大	10.65	2005/09/06 09:00
Case 2	2 番目	9.80	2004/08/30 11:00
Case 3	3 番目	9.50	1993/08/01 21:00
Case 4	任意の中規模	6.03	1999/08/17 16:00

表-3　テストのケース一覧（ターゲット流域）

ケース名	洪水イベントの規模	ピーク水位（m）	ピーク水位発生日時
Case 1	期間最大	5.03	2016/08/21 17:00
Case 2	2 番目	4.84	2001/09/12 07:00
Case 3	3 番目	4.77	2006/10/08 15:00
Case 4	任意の中規模	3.20	2011/09/22 13:00

（**図-4b**）と再学習有りのネットワーク構造（**図-4c**）のモデルをそれぞれ用意する．なお，再学習の回数は，エポック数に対応する．

予測結果の誤差評価には，二乗平均平方根誤差（RMSE）と Nash-Sutcliffe 係数（NS），機械学習の収束確認に損失係数（E）が，それぞれ適用される．それらは，次のように定義される．

$$\mathrm{RMSE}=\sqrt{\sum_{i=1}^{N}(V_{ci}-V_{oi})^2/N}, \tag{7}$$

$$\mathrm{NS}=1-\sum_{i=1}^{N}(V_{ci}-V_{oi})^2/\sum_{i=1}^{N}(V_{oi}-\langle V_{oi}\rangle)^2, \tag{8}$$

$$\mathrm{E}=0.5\times\sum_{i=1}^{N}(V_{ci}-V_{oi})^2. \tag{9}$$

ここで，N＝計算値/観測値の個数，V_o＝観測値，V_c＝計算値，$\langle\cdot\rangle$＝平均値である．

3．結果と考察

まず，ソース流域でのLSTMモデルの予測精度の検証を行った．4つのケースの学習後に，各ケースの4つの洪水イベントについて，バリデーションを行ったところ，RMSE = 0.30–0.43 m，NS = 0.97–0.98で学習は適切に行われたことがわかる．Case 1～Case 3のテストの予測結果は，RMSE = 0.38–0.66 m，NS = 0.97–0.98で良好な結果が得られた．しかし，**図-5**に示されるように，ピーク水位は軒並み過小評価されている．これは，学習に期間1～3番目に大きな洪水イベントを学習していない影響と考えられる．また，中規模の洪水イベントをテストしたCase 4の予測結果は，RMSE = 0.18 m，NS = 0.99となり，また，ピーク水位も良好に再現でき，より高精度な予測結果が得られた（**図-5d**）．中規模洪水イベントは学習のためのデータに多く含まれるので，予測精度が向上したと考えられる．他の研究と比較について，一言ら[3]の全結合型ネットワークをもつ深層学習（DNN）モデルを使った水位予測では，同じ流域で1時間後予測のRMSEは，本研究のものと比べて1/4–1/3となり，より高精度な予測結果が得られている．洪水予測の既往研究[5]では，LSTMモデルの方が，DNNモデルよりも予測精度が高いと報告されているものの，検証方法の違い，ハイパーパラメータの条件が異なることが，予測精度に影響しているものと考えられる．

次に，LSTMモデルへの転移学習の導入の有り無しについて比較を行う．ソース流域で事前学習された各ケースのモデルを，ターゲット流域の呼応するケースに適用する．つまり，ソース流域のCase 1 のモデルは，ターゲット流域のCase 1に直接適用され，再学習を行わずにテストのみを行った（**図-4b**）．

ここで，比較のために転移学習を行わない場合に，ターゲット流域の予測計算を実施した．この計算では，ターゲット流域のデータのみを使って，ソース流域で実施した方法と同様に，学習・バリデーション・テストを行った．テストの結果のみを，転移学習の有り無しの比較とする．**図-6**に示されるように，1時間後の予測水位の比較は，どのケースにおいても，転移学習（但し，再学習無し）を導入した予測結果の方が，より良好に観測値を再現できた．この理由は，事前学習を含めてより多くの洪水イベントの特徴を学習したためと考えられる．**表-4**で示されるように予測精度の定量評価によれば，転移学習有りの結果は，無しの結果に比べて，Case 1～Case 3は，30%前後の改善が見られたものの，ピーク水位値の再現は，ソース流域でのテスト結果と同様に過小評価であった（**図-6abc**）．また，Case 4は60%強の改善が

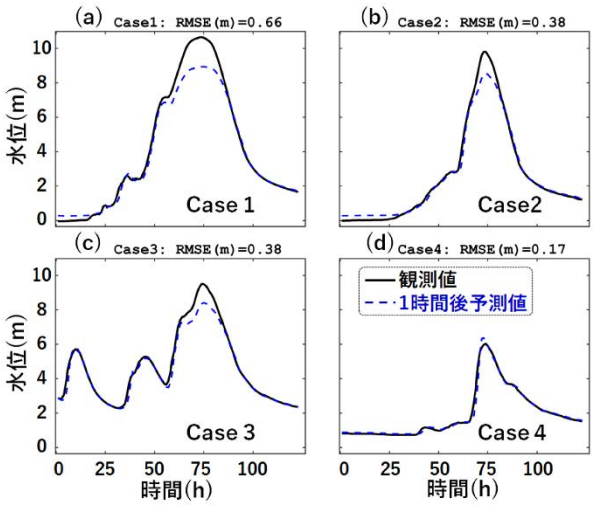

図-5 ソース流域のテスト結果

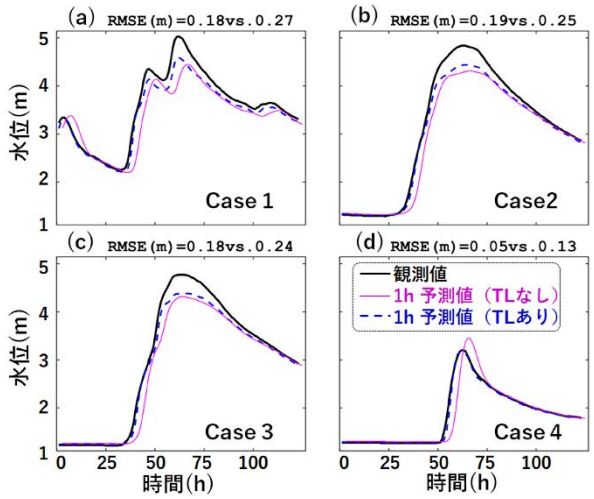

図-6 ターゲット流域のテスト結果
（TL＝ 転移学習，TL あり＝再学習無しの場合の TL）

表-4　予測精度の指標一覧（ターゲット流域）

ケース名	転移学習有り		転移学習無し	
	RMSE (m)	NS	RMSE (m)	NS
Case 1	0.181	0.943	0.271	0.871
Case 2	0.186	0.979	0.254	0.960
Case 3	0.176	0.982	0.243	0.967
Case 4	0.049	0.993	0.130	0.950

見られ，水位ピークもほぼ再現できた（**図-6d**）．ピーク水位の再現性に関して，Case 1～Case 3とCase 4の予測精度の結果が異なる理由は，ソース流域で事前学習されたモデルの予測精度に依存する，つまり，事前学習モデルで抽出された特徴量が再利用モデルに引き継がれていること示している．

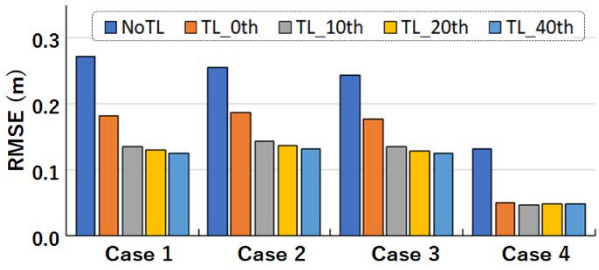

図-7 再学習の場合の予測精度の比較（NoTL= 転移学習無し，0th= 再学習無し，10th, 20th, 40th= 再学習回数）

表-5 ピーク水位の予測値（ターゲット流域）

ケース名	観測値 (m)	再学習後の値（m）			
		0 回	10 回	20 回	40 回
Case 1	5.03	4.59	4.66	4.67	4.68
Case 2	4.84	4.44	4.51	4.52	4.53
Case 3	4.77	4.38	4.46	4.47	4.48
Case 4	3.20	3.19	3.25	3.26	3.27

ターゲット流域において，転移学習を導入したLSTMモデルの予測精度をさらに向上させるために，ネットワークの一部分である全結合層（**図-4c**）を学習用の洪水イベントデータを使い10～40回と再学習を行った．**図-7**に示されるように，Case 1～Case 3は，再学習が10回で，誤差の値が収束していることが分かる．Case 4について，再学習0回のモデルでも，良好な予測が可能であった．

また，学習回数に対するピーク水位値の再現について定量的に評価すれば（**表-5**），Case 4を除き，再学習を40回実施することでで，再学習0回よりも約2%改善されることが分かった．10回まで行えば，ほぼ改善割合は収束する．一方，Case 4は，再学習による改善は見られなかった．再学習に関する検討から，ソース流域で事前学習されたモデルの精度が多少低くても，再学習によって精度が改善されることが明らかになった．

類似研究として，LSTMを用いて数多くの流域（約250個）で長期の経年変化を伴う流出解析を行った研究[16] が挙げられる．本研究と同様の手法で，複数の流域で事前学習を行ったモデルを，各々の流域で fine tuning を施した場合に，fine tuning 無しの結果と比較して，NS の差分値が 0.0–0.4の範囲で予測精度を改善し，複数の流域で共通にもつ水文学的な特徴を転移できたと結論付けている．この既往研究のように，本研究の短期間の洪水予測に転移学習を導入する場合にも，ソース流域の水文学的な特徴をターゲット流域に効果的に転移できたと考えられる．

本研究のもう1つの課題であった計算時間の削減

について，我々が開発したLSTMモデルは，比較的シンプルなネットワークなので，全行程の計算時間はエポック数の回数に要する計算時間にほぼ等しい．したがって，転移学習無しのケースに要する計算時間を1とすると，転移学習有りで且つ良好な結果が得られた再学習10回の計算時間は，およそ1/10に削減された．

４．まとめ

高精度な洪水予測のために開発されたLSTMモデルを，不十分な観測データを有する流域でも汎用的に適用できるようにするために，十分に事前学習されたモデルを再利用するための手法である転移学習を導入した．ソース流域の水位予測において，良好な再現結果が得られたモデルに転移学習を導入し，ターゲット流域へ適用後に，1時間後の水位予測を実施した結果は，以下のようにまとめられる．

- 転移学習有り（但し，再学習無し）の予測結果は，転移学習無しの結果と比較して，データ量が限られる期間最大クラスの洪水イベントの場合で，30%程度改善され，データ量が比較的多い中規模クラスの洪水イベントの場合には60%強の改善が見られた．
- 転移学習有りで，モデルの一部を再学習した場合に，再学習をしなかった場合に比べて約2%の改善が見られた．
- 計算時間の削減について，転移学習有り（但し，再学習有り）の計算時間は，転移学習無しと比べて1/10となった．

今後の改良点について，本モデルはまだ開発途上のため，1時間後までの予測しかできないので，数時間後の予測ができるように拡張する必要がある．

謝辞：本研究は農研機構重点化経費の支援を受けた．プログラム開発には，㈱アーク情報システムの木野由也氏の助言を頂いた．研究資金の提供，及び関係者のご協力などに謝意を表明する．

参考文献

1) Mosavi, A., Ozturk, P. and Chau, K.-W.: Flood prediction using machine learning models: literature review, *Water*, Vol. 10, 1536, 2018.
2) Hsu, M.-H., Lin, S.-H., Fu, J.-C., Chung, S.-F. and Chen, A.-S: Longitudinal stage profiles forecasting in rivers for flash floods, *J. Hydrol*., Vol. 388(3-4), pp.426-437, 2010.
3) 一言正之，櫻庭雅明，清雄一：深層学習を用いた河川水位予測手法の開発，土木学会論文集B1（水工学），72巻4号, pp. I_187-I_192, 2016.
4) Hochreiter, P. and Schmidhuber, J.: Long short-term memory, *Neural Comput*., Vol. 9(8), pp.1735-1780, 1997.
5) 山田恒輝，小林洋介，中津川誠，岸上順一：リカレントニューラルネットワークを用いた2016年の常呂川洪水事例の水位予測，土木学会論文集B1(水工学), 74巻5号, pp.I_1369-I_1374, 2018.
6) Hu, C., Wu, Q., Li, H., Jian, S., Li, N. Lou, Z.: Deep learning with a long short-term memory networks approach for rainfall-runoff simulation, *Water*, Vol. 10, 1543, 2018.
7) Pratt, L.Y.: Discriminability-based transfer between neural networks. In Proc. of the Advances in NIPS 5, Denver, Colorado, USA, Nov.30-Dec.3, 1992, pp.154-196.
8) LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition. In proc. of IEEE, Vol. 86, pp. 2278-2324, 1998.
9) Laptev, N, Yu, J., and Rajagopal, R.: Reconstruction and regression loss for time-series transfer learning. In proc. of the Special Interest Group on Knowledge Discovery and Data mining (SIGKDD) and the 4th workshop on the MIning and LEarning from Time Series (MiLeTS), London, U.K., Aug.20, 2018.
10) Kimura, N., Yoshinaga, I., Sekijima, K., Azechi, I. and Baba, D.: Convolutional neural network coupled with a transfer-learning approach for time-series flood predictions, *Water*, Vol. 12(1), 96, 2020.
11) 巣籠悠輔：詳解 ディープラーニング―TensorFlow・Keras による時系列データ処理，マイナビ出版，328p, 2018.
12) Python TM : http://www.python.org（閲覧日：2020年5月12日）
13) Keras : Pythonの深層学習ライブラリ, https://keras.io/ja/（閲覧日：2020年5月12日）
14) James, G., Witten, D., Hastie, T. and Tibshirani, R.: An introduction to statistical learning: with applications in R, 1st ed., p.176, Springer, 2013.
15) Yosinski, J., Clune, J., Bengio, Y. and Lipson, H.: How transferable are features in deep neural networks?, *Adv. Neur. In.*, Vol. 27, pp. 3320-3328, 2014.
16) Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrol. Earth Syst. Sci*., Vol. 22, pp. 6005–6022, 2018.

（2020.4.2受付）