

令和 6 年 6 月 4 日現在

機関番号：82111

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K05838

研究課題名（和文）データが不十分な低平地排水管理における深層学習と物理モデルとの融合技術の開発

研究課題名（英文）Development of an integrated method between deep learning and physical models in lowland drainage management with insufficient observed data

研究代表者

木村 延明（Kimura, Nobuaki）

国立研究開発法人農業・食品産業技術総合研究機構・農村工学研究部門・上級研究員

研究者番号：40706842

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：本研究では、河川や農業水利施設（例：排水機場）の大規模洪水を対象に、水位予測を行うための深層学習によるAI予測モデルを構築することが目的である。大規模洪水のようなデータサンプル数が少ない事象の場合に、AI予測モデルはデータ駆動型モデルのために十分な能力が発揮できない。そこで、まず、物理モデル（例：流出解析モデル）等で大規模洪水に匹敵する仮想データを人工的に多く生成し、それらをAI予測モデルの学習データとして用いて、事前学習モデルを構築した。次に、実測値の特徴を取り込むために、少ないサンプル数の実測値を用いて事前学習モデルの一部を再学習させること（転移学習）で、高精度な予測手法を確立した。

研究成果の学術的意義や社会的意義

本研究の学術的意義について、一般に、深層学習モデルは、データサンプル数が少ない場合には、予測精度が劣るものの、その欠点を補うために物理モデルからの疑似データを割増し、さらに、転移学習で疑似データの特徴をサンプル数が少ない対象に転移することで、予測精度の向上が可能な手法（物理ガイド深層学習モデル）を開発した。

社会的意義について、現地の観測データのサンプルが少ない場合でも、物理モデルで疑似生成された大量のデータを併用して学習する、物理ガイド深層学習モデルは、現地への適用を通して、実用的に有用であることを明らかにし、さらに、データ保有に関して同様な条件の他の地区へ普及させる可能性を見出せた。

研究成果の概要（英文）：This study aimed to create a robust AI model that predicts water levels for severe flood events in rivers and agricultural water facilities (e.g., drainage pumping stations) using deep learning. For a small number of data samples, such as the number of large-flood occurrences, the AI models have poor performance in accurate prediction due to data-driven models that require large amounts of data. To improve this problem, first, we artificially generated a large amount of virtual data comparable to large floods using a physical model (e.g., a runoff analysis model), and used the data as training data for the AI model to construct a pre-learned model. Next, to incorporate with the observed data features, we established a highly accurate prediction method by retraining part of the pre-trained model (i.e., transfer learning) using few observed data samples.

研究分野：水工学

キーワード：少ないデータ量 深層学習 水位予測手法 物理モデル 転移学習

1. 研究開始当初の背景

低平地の持続可能な排水管理を行うために、土地利用の変遷に伴う排水状況の変化、気候変動による豪雨対策、管理者の高齢化による若手管理者の育成などの様々な課題を解消する必要がある。その一助として、効率かつ迅速なリアルタイム排水管理を行う水位・流量予測システムの構築が求められる。排水機場調整池の水位を予測するために、これまで排水管理に関する物理モデルの構築が試みられ、高精度の予測ができるように物理モデルの精緻化が行われてきた。しかし、物理モデルの欠点は、複雑な現象を扱う場合に要素還元主義の慣例からマイナーな要素を省略したり、人間行動などの不確実性を含む現象に対して、数式モデルを構築できなかつたりと、不確実性なまま・複雑なまま、現象を予測するのが困難なことである。また、物理モデルの精緻化を追求するにつれて、計算コストの増加を伴い、迅速な予測が困難となる。これらの物理モデルに関する課題に対して、最先端の技術であり、人工知能 (AI) の一種である深層学習を用いて改善をはかる試みがある。深層学習の利点は、観測データからパターンを自動抽出し、不確実かつ複雑な現象についても予測が可能で、一般に計算コストも安い。申請者らは、この深層学習を組み込んだニューラルネットワークモデル (以下、「深層学習モデル」という) を開発し、低平地の排水管理に適用した (木村ら 2019a)。他方、深層学習モデルの欠点は、観測データが不十分な場合に、水位・流量の予測精度は低下する傾向を示す。これを改善するために、物理モデルから生成された多数の疑似データと深層学習モデルを用いれば、数時間先までの水位・流量予測は、最大 10%誤差を含むものの良好な予測結果が得られた (木村ら 2019b)。また、木村ら (2020) は、他の低平地で蓄積された観測データを多く学習することで、観測データが不十分な流域の洪水予測でも、深層学習モデルの予測精度が向上する手法を提案した。この予測手法では、基データ (領域 A) で内部パラメータが最適化された深層学習モデルを、対象データ (領域 B) で内部パラメータを部分的に調整して転用する「転移学習」 (Pratt 1992) の考え方が用いられた。しかし、この手法を現地に適用した事例は見られなかった。

2. 研究の目的

深層学習モデルと物理モデルのそれぞれの欠点を解決するために、また、観測データが少ない低平地においても、物理モデルによる疑似データ (領域 A に相当) の生成と、転移学習によって、疑似データの特徴を転移されたデータ (領域 B に相当) を用いて、迅速かつ高精度な予測手法を構築し、その予測精度の向上を検証する。

本予測手法は、物理モデルによってガイドされる手法 (以下、「物理ガイド深層学習モデル」という) である (図-1)。なお、本研究のテーマにつながった、申請者らが担当してきた深層学習モデルの開発の経緯について表-1 にまとめる。

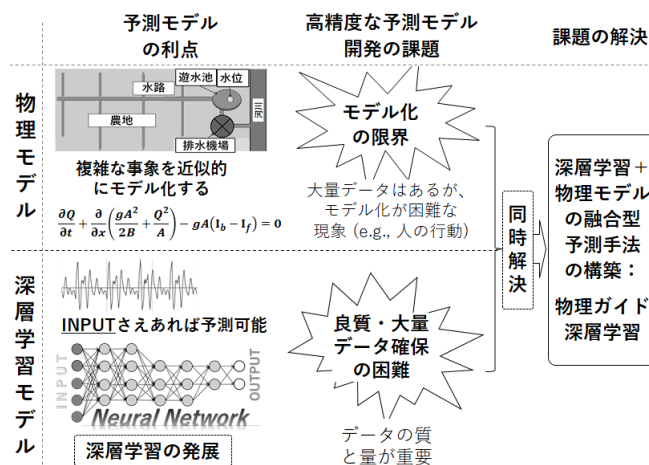


図-1 本研究の着想に至る背景の概略図

表-1 深層学習モデルの開発状況 (○は良好な予測精度が得られた結果を示す)

観測されたデータ量	深層学習モデル	深層学習モデル + 物理モデル	深層学習モデル + 転移学習*	深層学習モデル + 物理モデル + 転移学習
十分**	○	○(疑似データ代用)	—	—
不十分	—	—	○(河川洪水データ)	本研究目標

*) 国交省から提供を受けた十分な河川洪水データ (排水機場の水位・流量データとは異なる) を用いて、事前学習された深層学習モデルを転移学習によって、データが不十分な流域に適用した。

**) 十分な観測データの基準について、水工学分野では、データ数が約 10 万点以上と想定した。

3. 研究の方法

物理ガイド深層学習モデルは、事前学習と転移学習を組み合わせている。事前学習では、物理法則に基づく現象のエッセンス (特徴量) を取り出し、次の転移学習で観測値の特徴量を考慮して、より現実的な予測を行うことができる。物理ガイド深層学習モデルは図-2 に示すように、以下のステップで実施する。

- i) アメダスに基づく模擬雨量を入力データとし、物理モデルによって大量の疑似データを生成する。

- ii) 上記 i)で生成された疑似データを深層学習モデルの入力データとして事前学習を行う。
- iii) 不十分な観測データを、ii)で事前学習された深層学習モデルの入力データとし、深層学習モデルの内部パラメータの一部を再調整するための転移学習を用いて高精度の予測を行う。

対象地区は、ある低平地で常時及び洪水時の排水管理が行われているポンプ場を有する地区である。予測対象は、ポンプ場近くに位置する遊水池の水位であり、この水位変化に基づいてポンプ操作が行われる。水位データは、遊水池で毎時観測された過去8年間分のデータである。特に、その期間に発生した期間最大の洪水イベントに着目する。期間最大の洪水イベントを予測するためには、比較的規模が大きい洪水イベントを抽出し、それらを学習データにする必要がある。しかし、観測データを確認すれば、洪水イベントの回数は少ないものであった。本研究では、これを不十分なデータと定義する。他方、深層学習モデルの学習データの量を補完するために、確率プロセスに基づいた模擬降雨を入力データにして、対象地区で構築された物理モデル

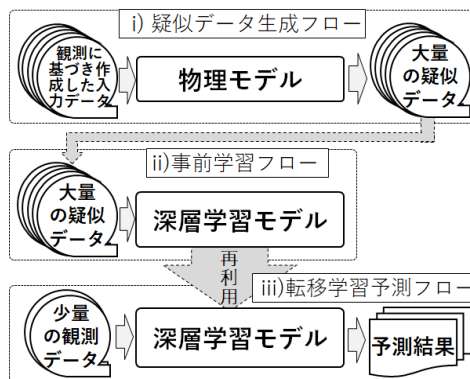


図-2 物理ガイド深層学習モデルの処理手順

によって、大量の疑似データ（水位）を生成した。3つの模擬豪雨の組（100, 300, 500 mm/72 hours）に対応する洪水イベントの水位データを、模擬豪雨群毎に対応する各1000個の疑似学習データをそれぞれ作成した。これらの3つの組の疑似水位データを用いて、学習した事前学習モデルを、事前学習-100mm, 事前学習-300mm, 事前学習-500mmと定義する。なお、事前学習-300mmがTOP1の豪雨に匹敵する事前学習モデルである。

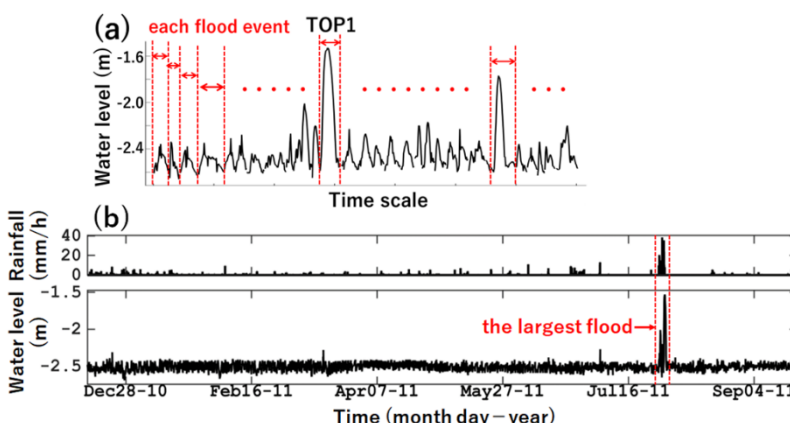


図-3 対象データ：(a)各洪水事象を集合したデータ、(b)連続データ(常時排水含む)

ここで、2つの計算ケースを準備した。Case1では、洪水イベントのみを抽出したデータセットを作成（図-3a）し、期間最大の洪水イベント（TOP1）のみを予測対象にする。Case2では、利用できる全観測区間（非洪水イベント期間、つまり常時排水時の区間も含む）を利用データにして、TOP1を含む期間（図-3b）を予測対象にし、その他の区間を学習データにするものである。なお、図-3bの区間は、全観測区間の1/10の区間である。

4. 研究成果

(1) Case1の予測結果について、約30個の洪水イベントのうち、TOP1を対象に1時間先（リードタイム1h）の予測を行った。転移学習で再学習を5回刻みで5~50回行った予測値とそれに対応する観測値を図-4に示す。転移学習なしの予測値と比較すれば、事前学習-300mmを用いた場合（図-4b）に、ピーク水位を良好に再現することが示された。この結果は、TOP1の降雨量が、事前学習-300mmを生成した模擬降雨に匹敵する量であったことと整合がとれる。また、事前学習-500mmもピーク水位を再現できるものの、図-4cに見られるように波形のピークから下降曲線部分で強いオーバーシュートやアンダーシュートが発生し不自然な形状になるので、再学習で過剰な最適化が行われたことが示唆される。

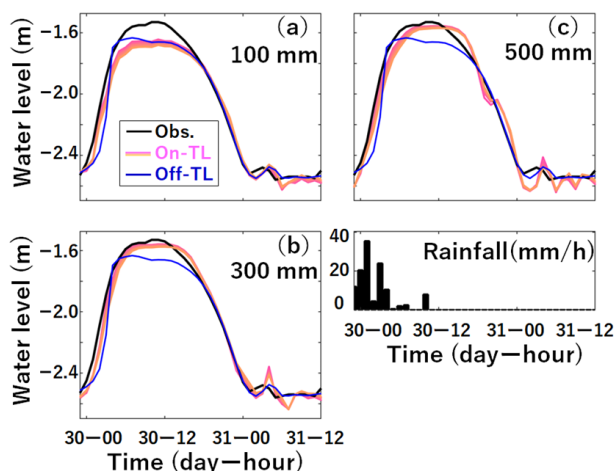


図-4 Case1の予測結果：72時間の降雨量 (a)100 mm, (b)300 mm, (c)500 mmの疑似データを利用、On-TL=転移学習あり、Off-TL=転移学習なし。

(2) Case2の結果では、TOP1を含み図-3bに示される観測データの区間（以下、「区間3b」という）を予測対象にし、3つの事前学習モデル（事前学習-100mm, 自前学習-300mm, 自前学習-500mm）を適用した。転移学習を用いて、5回刻みで5~50回、区間3b以外の観測データで再学習をし、区間3bの水位予測を行った。リードタイム1hの水位予測精度の評価について、各

事前学習共に 50 回までの再学習で、予測値の変化が安定した (図-5)。図-5 において、対象区間 3b の観測値に対して、再学習 50 回の転移学習ありの予測結果と転移学習なしの予測結果を比較した場合、自前学習-300mm と自前学習-500 mm の転移学習ありの予測結果、概ね同程度の誤差を示した。一方、自前学習-100mm は他の事前学習モデルと比べて、予測精度が低下した。TOP1 の短い期間を対象にした水位の予測精度について (図-5a,b,c の右側小パネル)、自前学習-300mm は、特に水位ピークを含む波形の再現性がより良好であった。他方、典型的な常時排水時の予測結果は、再学習回数が少ない時には、観測値からアンダーシュートする分布が見られるものの、概ね再現性が良いことがわかる (図-5a,b,c の左側小パネル)。

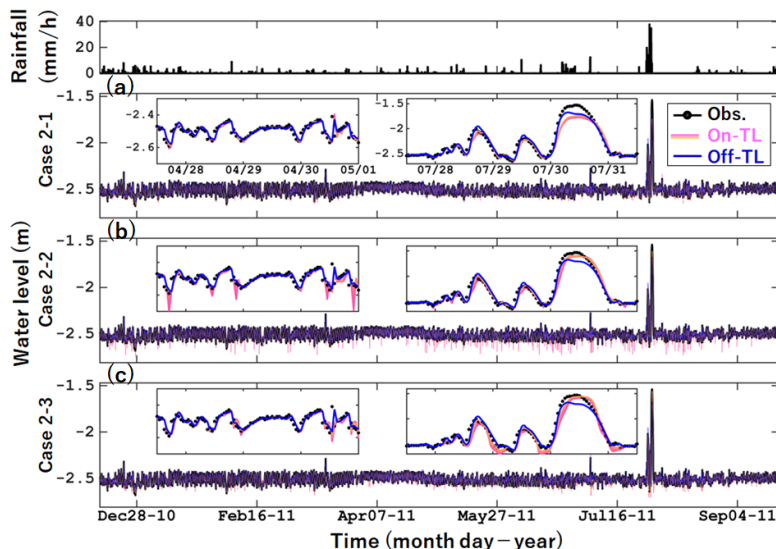


図-5 Case 2の予測結果：72時間の降雨量 (a)100 mm, (b)300 mm, (c)500 mmの疑似データを利用。Obs. = 観測値, On-TL = 転移学習あり, Off-TL = 転移学習なし。

上述した Case 1 と Case 2 の結果から本研究で開発された物理モデルのデータを学習データに組み込み、さらに転移学習で観測値の特徴を取り込む予測手法は、従来のな深層学習のみを用いる予測手法 (つまり、転移学習なしの予測計算) と比べて、低平地の大規模な洪水イベントの予測に有用であることを示すことができた。

<引用文献>

- ① 木村延明, 中田達, 桐博英, 関島建志, 安瀬地一作, 吉永育生, 馬場大地: LSTM モデルを用いた低平地排水機場の水位予測、土木学会水工学論文集第 64 巻、2019a、I_139-I_144
- ② 木村延明, 中田達, 安瀬地一作, 関島建志, 桐博英, 馬場大地: 人工ニューラルネットワークモデルを利用した排水機場遊水池の水位予測に関する研究、農研機構研究報告 農村工学研究部門、3 巻、2019b、71-80
- ③ 木村延明, 吉永育生, 関島建志, 安瀬地一作: 不十分な時系列データを用いた ANN モデル予測の改善方法、水土の知: 農業農村工学会誌、88 巻 1 号、2020、35-38
- ④ Pratt, L.Y.: Discriminability-based transfer between neural networks. Proc. of the Adv. in NIPS 5, Denver, Colorado, USA, Nov. 30-Dec. 3, 1992.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 木村延明, 皆川裕樹, 福重雄大, 馬場大地	4. 巻 78
2. 論文標題 サポートベクター回帰を用いた水位予測モデルの低平地用排水管理における精度検証	5. 発行年 2022年
3. 雑誌名 土木学会論文集 B 1 (水工学)	6. 最初と最後の頁 I_139-I_144
掲載論文のDOI (デジタルオブジェクト識別子) 10.2208/jscejhe.78.2_I_139	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 木村延明, 皆川裕樹, 福重雄大, 木村匡臣, 馬場大地	4. 巻 77
2. 論文標題 深層学習モデルに物理モデルを融合させた排水機場水位予測への転移学習の適用	5. 発行年 2021年
3. 雑誌名 土木学会論文集 B 1 (水工学)	6. 最初と最後の頁 I_319-I_324
掲載論文のDOI (デジタルオブジェクト識別子) 10.2208/jscejhe.77.2_I_319	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 木村延明, 皆川裕樹, 福重雄大, 馬場大地	4. 巻 80
2. 論文標題 河川洪水予測用の事前学習モデルの構築と検証	5. 発行年 2024年
3. 雑誌名 土木学会論文集 特集号 (水工学)	6. 最初と最後の頁 n/a ~ n/a
掲載論文のDOI (デジタルオブジェクト識別子) 10.2208/jscejj.23-16147	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Nobuaki KIMURA & Hiroki MINAKAWA
2. 発表標題 An inexperienced flood prediction at a drainage pumping station in lowland using an advanced neural network model, coupled with a physical model and transfer learning.
3. 学会等名 PAWEES International Conference 2022 (国際学会)
4. 発表年 2022年

1. 発表者名 木村延明, 皆川裕樹, 福重雄大, 馬場大地
2. 発表標題 層学習に物理モデルを結合させた排水機場水位予測手法の構築
3. 学会等名 第70回農業農村工学会大会講演会
4. 発表年 2021年

1. 発表者名 木村延明, 皆川裕樹, 福重雄大, 馬場大地
2. 発表標題 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
3. 学会等名 第72回農業農村工学会大会講演会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関