



岐阜大学機関リポジトリ

Gifu University Institutional Repository

Investigation of long short-term memory model  
applied for imputation of missing discharge data

メタデータ	言語: eng 出版者: 公開日: 2022-10-07 キーワード (Ja): キーワード (En): 作成者: WEILISI メールアドレス: 所属:
URL	<a href="http://hdl.handle.net/20.500.12099/88932">http://hdl.handle.net/20.500.12099/88932</a>

氏名（本籍）	WEILISI（中華人民共和国）
学位の種類	博士（工学）
学位授与番号	甲第630号
学位授与日付	令和4年6月30日
専攻	生産開発システム工学専攻
学位論文題目	Investigation of long short-term memory model applied for imputation of missing discharge data (流量欠測データの補完への LSTM モデル適用に関する研究)
学位論文審査委員	(主査) 教授 篠田 成郎 (副査) 教授 玉川 一郎 准教授 児島 利治

### 論文内容の要旨

水文分野においてデータの欠測は、水文解析の前処理において避けることが困難な重要な問題である。観測機器の更新により観測ミスによるデータ欠測を減少させることができるが、様々な理由によりデータ欠測が発生する可能性は残る。欠測部分の補完は問題解決方法の一つである。水文データの欠測補完方法には様々なものがあり実用化されている。短期間の欠測は、欠測期間前後の観測データを用いて補完される場合が多い。一方、長期間の欠測に対しては、近傍の観測点のデータを用いた補完、タンクモデル等の流出モデルを用いた補完等がある。しかし、このような手法には、洪水等の極端な水文条件下での精度が低いこと、データ変換量が多い、時間を多く消費するなどの問題点がある。ディープラーニングはここ数年で急速に発展しており、流出予測や洪水予測に多くの応用例がある。しかし、データ欠測の補完に関しては、欠測前後のデータを用いた補完や、上下流の近傍観測点のデータを用いた補完が比較的有効であるため、ディープラーニング技術の適用例はあまり無い。

本論文では、流量欠測データの補完のために、時系列データの解析に適した Long Short-term Memory Model (LSTM) を適用し、その精度評価や様々なハイパーパラメータの設定についての検討を行っている。岐阜県大八賀川流域の塩屋橋観測点と三福寺観測点を対象として、塩屋橋地点の流量の推定を試みている。隣接する三福寺地点の流量と、欠測期間を考慮して推定対象時間より 24 時間、168 時間過去の塩屋橋地点の流量、および気温、降水量を入力データとした検討を行っている。比較対照とする従来手法として、隣接する三福寺地点と推定対象の塩屋橋地点の観測データより求めた線形回帰式による補完を行い、ハイドログラフの評価に良く用いられる Nash-Sutcliffe 係数(NSE)によって精度の比較を行った。ディープラーニングの最適化アルゴリズムには一般に確率的勾配降下法(SGD)が用いられる。SGD は高速に良好な学習結果を得ることができるが、ランダム化アルゴリズムにより学習ごとに結果が異なる。1000 個の学習結果の中には NSE で 0.7 を下回る精度の結果も散見されたが、複数の学習結果をアンサンブル平均することによって、定常的に良好な結果が得られた。具体的には 1000 個の中からランダムにサンプリングした 20 個以上の学習結果のアンサンブル平均により、下側 5%点での NSE が 0.904 と従来手法の線形回帰式より良い精度を示すことができた。次に、LSTM の入力データやハイパーパラメータについて検討を行った。降水量、気温は欠測値の補完という目的に対してはノイズとなり精度低下の原因となることが分かった。推定対象の過去の流量と隣接地点の流量を入力データとした場合が最も精度が良かった。推定対象の過去の流量のみを用いた補完では、24 時間前のケースでは NSE で 0.7 程度を示したが、168

時間前では非常に精度が低かった。隠れ層のユニット数はユニット数が多いほど精度が良い傾向があったが、100 ユニットと 200 ユニットでほとんど差が無く、100 ユニット程度で十分であることが示唆された。一方、精度向上にあまり効果が無い中間層ユニットと出力層ユニットの連結を削除して精度を向上させるパラメータであるドロップアウト率は、0.00 の場合が最も精度が良かった。LSTM は画像解析に良く用いられる CNN に対してユニット数が非常に少ないため、ドロップアウトの効果があまりなかったためと思われる。以上のように本論文の検討により、流量欠測データ補完へ LSTM を適用する際の様々な知見を得ることができた。

### 論文審査結果の要旨

本論文では、近年発展してきたディープラーニング技術を、水文解析の前処理には欠かせない欠測データの処理に適用し、1) ランダム性を含んだ推定結果の不確実性に着目し、どのような処理を加えることにより、安定的な精度を提供できるようになるか検討し、2) 入力データの種類や LSTM のハイパーパラメータの設定について検討を行っている。ランダム性を含んだ SGD により結果は毎回変化するものの、20 個以上の学習結果をアンサンブル平均することにより 95%以上は従来手法より高精度の結果が得られることを示すことができた。入力データについては、流出モデルを介する補完方法とは異なり、気象データを混入することにより逆に精度が低下することが示された。隣接地点の観測データを用いることが最も良いが、当該地点のデータのみを用いる場合 1 日程度の欠測期間であれば、ある程度の精度で補完することができることが示唆された。また、ディープラーニングを適用する際に常に問題となるハイパーパラメータの設定値についても、隠れ層のユニット数は 100 程度、ドロップアウト率は 0.00 から 0.01 と低く設定しても構わないことが示されている。

これら 2 つの研究を通じて、水文解析の前処理に欠かせない欠測データの補完について有用な知見を示すことができている。本論文は、下記の発表論文 1)および 2)の内容を取り纏めたものであり、学術的にもその価値は認められている。

学位論文審査委員会では、以上の学位申請論文および発表論文（原著 2 編）を慎重に検討した結果、提出された論文は上述のように工学的価値が高く、かつ、学術的に完成された内容を有しているものと認め、論文審査に合格と判定した。

### 最終試験結果の要旨

学位論文審査委員会では、申請者に対し、学位論文公聴会等において、学位審査論文や発表論文の内容、および、それに関連した解析方法や評価方法の事項についての質疑を行うとともに、別途関連の研究状況とそれらの成果を実際に適用する場合の問題点の解決への展望など、細部にわたる口頭試問を行った結果、適切な解答等を得たので、最終試験に合格と判定した。

---

発表論文（論文名、著者、掲載誌名、巻号、ページ）

- 1) 深層学習による流量欠損値の補完方法の検討 (2020) 児島利治, Weilisi, 大橋慶介, 河川技術論文集, Vol. 26, 137-142.
- 2) Investigation of Hyperparameter Setting of a Long Short-Term Memory Model Applied for Imputation of Missing Discharge Data of the Daihachiga River (2022) WEILISI, Toshiharu KOJIMA, Water, Vol.14, No. 2, 213.